

Developing Digital Libraries: Need For Proper Strategies

R. Raman Nair

Attention of digital library is placed on metadata collection and management. Metadata describes content, structure, context and other aspects of a record in digital library and critical to resource indigital libraries. Reviews standards provided by the UNESCO and growth of digital libraries. Research and development of this aspect occurs in library environment like that of on-line catalogue of library of congress (OCLC).

INTRODUCTION

Ultimately libraries of the present day as known to everyone, are likely to disappear. Facilities will still exist to preserve the print-on-paper record of the past, of course, but they will be more like archives, or even museums, providing little in the way of public service. As for the electronic sources, libraries may have an interim role to play. In the longer term, it seems certain that the libraries will be bypassed. That is, people will have very little reason to visit libraries in order to gain access to information sources observed Lancaster two decades ago.

Lancaster's prophecy has become real. We are witnessing revolutionary changes in all spheres of human activities due to the impact of information and communication technologies. The present changes are not a process of gentle evolution, nor the result of careful planning. Revolutions involve massive change, the adoption of totally new paradigms, the loss of much that has been taken for granted and the introduction of innovations that were never thought possible. Instant, pervasive access to the world's information from the home, from the workplace and even while on the move raises questions about our traditional structures for organising and managing information.

What is presently happening is not simply the effect of a single new technology but the convergence of a variety of technologies - which together may be termed 'digital' - that provide the opportunity to develop new forms of relationship between people and organizations, largely independent of time and place.

The growing synergy of networked communications is, with the exception of language itself, the communication medium par excellence. In the mega-convergence of hypertext, multimedia, virtual reality, neural networks, digital agents, and even artificial life, each medium is changing different parts of our lives -our modes of communication, entertainment and work- but the Net changes all of that and more, all at once.

Libraries are falling victim to these pressures, as their role changes beyond recognition. If online information underpins the future, is there any place for old ways of recording and organizing information or knowledge. Some years ago, most of the libraries used to occupy several large rooms and employ huge quantum of people. It has been now replaced at organizations in all the advanced countries

by digital libraries with less staff and less space but more services. A small room without any books and journals but with three or four computers connected to the net can provide access to world wide information and a hundred times more better services than a physically present library can provide. Users can read books, download books to their disks or take the print out of the book or journal article they require and take them away to their table. Digital library has now become an essential part of all organizations dependent on information.

The technology is complex and most of our librarians have not developed the skills to understand it, exploit it or create it. Those few for whom the technology is viable finds scope for them only in private sector not for developing digital libraries but for applying the technology elsewhere. But in India, the developments related to the digital library is not with correct concept about the system. Here we consider digital Library and IT based systems as something belonging to computer scientist's domain. If our libraries are to be digitised and our knowledge heritage is to be conserved for generations to come it should have a clear concept about information systems, the technology used for developing that and librarians role in the process.

Since the primitive man drew the first pictures in the stone walls of his cave which is to be considered the first information system, which tried to store knowledge that has been gathered, recorded, organized and stored for the use of the community libraries or information systems metamorphoses with the changing technologies for recording knowledge. A fully developed system that we knew is our traditional library. But it also has limitations associated with storage and access to information, because most of the knowledge collected by it has been recorded on physical media like in the primitive cave wall but in a more convenient surface that enables easy use and transfer. But now digital media has come into dominance and in recent years the number and kind of digital information resources have proliferated. Developments in Information Technology have resulted in a remarkable expansion in abilities to generate and disseminate information in digital formats on electronic media. These developments have led to speculation on new concepts of the nature, role and use of data archives and libraries. New forms of knowledge repositories have become feasible, both technologically and economically.

Now for the last ten years or more everywhere everyone is telling about Information Infrastructures at the national level, state level, regional level, and institutional level. Crores and crores of rupees has been invested which would have been sufficient to extend access to every bit of information our nation hold to each and every citizen of India living even in the remote areas of the country. But information infrastructures ended as computers, networks, cables and satellites their physical infrastructures and nothing beyond that. Even in that, the technology adopted always lagged behind by five to ten years and again and again cables were digged out to lay new ones, systems were replaced to achieve minimum currentness. But still the information Kiosks at the village wherever they are materialised provide to the farmer price of wheat in an American farm or a village of Japan, and gives information about the rules and regulations of local authorities of many foreign countries. The Panchayat Act of his state he needs, a Government Order he requires, a book published in India he want to scan through for some reference or information on some services available in his locality is not there when he goes to the Kiosk if he ever cares to.

INFORMATION COMPONENT

So we have not considered the information component of Information systems till now. But now the importance has dawned upon us and we are now thinking of Digital Libraries, which will store information in the Information Infrastructures. Storage of Information or records or data for whatever requirements we can imagine has to use digital library technology. So librarians and their techniques have assumed more importance in the IT era in all advanced countries. The techniques they developed for knowledge and record management in library and information systems is now utilised in advanced countries for storing all types of records. Let the record be a government file, a book, a report, a video film, a sound recording, a map, a painting, or any thing that records information are stored in digital libraries. Access is controlled by the person who controls digital library and the Director of an institution only gets access to files that he only can see, finance department their files, and library users, their documents and information. Information infrastructures facilitate access, storage, further processing, transmission and any related process from any table and any place from in or outside the organization to the concerned person

Our information system developers are also currently thinking of the need for Digital Libraries. But like last decade in building up Information Infrastructures it is still behind another faulty complex that fail to recognise the role of library and information science in building up digital libraries. One thinks that it is the business of computer people. They will be able to provide the finished product without any involvement of library science or experts in subject content. And the finished product will perform magically whatever functions our traditional libraries used to perform and a thousand other things, which one could dream only in regard to that environment.

But will the creation of digitised records and using some retrieval systems available in our computing environments make a Digital Library? It is a question to be seriously enquired into at present. Failure will result in building digital Libraries, which will amount accumulation of digitised records that cannot be managed in future like the unorganised collections in traditional libraries. With the changes in technology that is very fast in IT Scenario the records of Digital Libraries that we create will become unscalable, unusable. The way in which we handle originals will send them to their final grave with the first digitisation of that record and we will lose the centuries old priceless wealth that withstood all the carelessness and disrespect that man has shown to them through centuries.

Several Question

In this context one has to think what is a digital library? Can it be developed without involvement of experienced librarians, including classifiers, archivists, conservation technologists, cataloguers and those who are familiar with the ways of providing information services on the subject areas covered by the content of the digital library to be developed? Is the protection and care of original document important while digitising it? Can the original or its conservation be disregarded once it is digitised? Can the optical character recognition and related editing and automatic indexing process be done in ancient documents to be kept in digital library? Is optical character recognition possible in ancient documents, if sophisticated equipments and programmes are used as computer science people who offer solutions claim to be using? In this context what is the importance of old local language scripts used in the ancient documents? Is the responsibility of librarians end with

selecting and providing documents to be included in the digital library and prescribing some norms for digitisation or use of original documents? Can the available general-purpose database management systems and image copying programmes manage the digital library development? Is Digital Library Technology more affiliated to Computer related subjects or Library and Information Science related areas.

An Example

These questions come to us when one of our professionals narrated an incident he witnessed in regard to the development of a digital library at a prestigious institution of the country. The institution has decided to digitise some very old documents available in it. As the institute has no facility or expertise for undertaking the work it has invited the organizations that can take up that work. From the organizations that have come forward it has selected a few professionals to be competent to do the work. After various discussions the organizations were requested to demonstrate the prototypes, programmes and methods they will use to digitise the collection before an expert body. The expert body consisted of computer and IT specialists as well as administrative officers and the librarian of the organizations. Of the programmes demonstrated there was a functional prototype. One or two groups brought actual digital library packages that will be used for the project. Of this one was developed with CDS/ISIS package at its front end and they are the group who demonstrated a functional prototype. Most of them proposed digitising the books and providing some popular DBMS programmes at the front end for enabling retrieval.

While demonstrating the equipments and methods for digitising one group dismantled the pages of a very ancient book for scanning and they told they will be neatly binding it after the digitisation. Another group used some automatic methods to turn pages that broke the pages of old brittle books while turning pages automatically. They also explained that after digitising the material they would do the conservation work for that document. Another solution provider told the pages of the book will be laminated after the digitisation to preserve it. The cost of digitising a page varied from Rs. 4 to Rs. 8. In this context we have to understand that restoring a page of an ancient document as per standards prescribed by conservation laboratory will come to about Rs. 50. Lamination, use

of general adhesives, disturbing the original binding in tact etc are not permissible and even if such work is done that may cost more than Rs.10 per page in addition to digitisation cost. So it is funny thing to make the real cost benefit analysis considering all these aspects. And a consideration of the fate of the original after digitisation work will point out a great crime one will be doing to our past and future generations. The majority of people who attended that programme never considered the importance of preserving the original after digitisation. Some even expressed the view that instead of wasting the space by keeping the original, once it is digitised the original can be destroyed.

LIBRARY SKILL V/S IT

The attitude of experts and all those concerned with the project were towards very popular DBMS programmes, and cost effectiveness and involvement of only computer and IT specialists in the work and they considered that librarian has nothing to do in the processes of digital library development. Interestingly the Librarian involved has no exposure to IT and computer. Even the incident narrated by the professional who happened to witness reveals that he is lacking the library science specialization and awareness of the value of ancient documents and their conservation needs to take a decision on such projects or support decision making. This is the case of most senior librarians in our country and they cannot be blamed for it. For these developments are recent ones and they never learned it in their formal courses or trainings. Government or organizations never did anything to give them the necessary minimum awareness and training in modern technologies while at service or took a positive attitude when someone showed interest to learn.

Non of the experts invited to assess the solutions of that project where librarians or those trained or experienced in Digital Library development. Non of the experts from computer side who were involved have earlier developed a digital library or associated in the development of any digital library that is functioning at that time. So our systems lack the human awareness capable to develop digital libraries or maintain them. But few attempts have succeeded in the country but could not be maintained due to the complexes and ignorance about the technology existed and that exists at present also.

So in the cited instance also the result was that no one could evaluate the solutions based on a perspective of technology as well as the future development and facilities for scaling further to be incorporated. Everything is mere formality for fulfilling the procedural requirements. Projects will waste funds, time and end without results. But the sad thing is not the loss of time or financial resources but the loss that may be caused to knowledge conserved priceless wealth which our forefathers have left for the generation to come.

We will develop digital libraries using computer infrastructures or computer solution providers. When the collection becomes large one will find that the problems of the Digital library are a thousand times more than such a traditional collection. So when we start our attempt to reorganize the digital Library or transform the content to new formats and systems we will understand that the digital libraries that we developed have not envisaged such changes. We will go on like this for another ten years or more and then we will come to an informed and technically feasible view of Digital Library. Then everything may need to be done from the start. But by the time we would have lost the originals that are to be again used for the content development.

OPTING ADVANCE TECHNOLOGY

Advanced countries opt right technology at the start itself and what they do stands. In our country it takes years to reach at right technology and by the time we could have wasted ten times resources than the advanced countries have used for a similar purpose. So it is not the lack of technology or lack of financial resources that prevent our slow development. It is the complexes and interests that prevent us from taking informed decisions on information technology application. Also it is lack of our competence, the lack of readiness of our professionals at higher levels to improve and accept the technology of the time and that too in time.

Anyhow the instance on digitisation that the professional narrated helped us to refresh our thoughts and knowledge on digital library initiatives. A look through Digital Library research and developments going on around us can very easily help us to be in the right track at least for our knowledge and to be capable to view things in their right perspective.

It is to understand that advances in digital libraries research and technologies have mainly occurred in huge traditional libraries and not in computer/software labs. Of the less than a dozen solutions that exist in the world about eight have originated in libraries with full or major involvement of librarians. Among them only one integrated solution has been known to have originated in India and was programmed by Mr. K.H. Hussain who is also a librarian by profession.

The funny thing is that we cannot respect it or accept it because a librarian developed it. More than that it is built around software developed by UNESCO for libraries, which is also accepted as a librarians' software. And the software is also a very cheap one costing less than five per cent of the common DBMS we get with the popular Office suits.

Who cares for the power of its search engine which is a hundred times powerful than most of the DBMS we propose for digital libraries. Who cares the resources an international organization like UNESCO continuously put in for research and development of this program. Who cares for the unlimited number of records that have to go into a database of the library, which has no parallel in most other projects or institutional contexts. We care only to ensure that the solution provider is an authentic computer scientist. Does a popular multinational software group develop the software? Is its price according to statuesque acceptable? And many more things, which we cannot say.

LIBRARY SOFTWARES

The research on digital library solutions all over the world, which were successful, reveal increased collaboration between librarians, computer technologists, and experts in the subject content of the documents which are to be digitised, in developing programmes. The activities of a few digital library solutions like Greenstone, TechlibPlus, Nitya Archive etc and materialised projects like that of UNESCO Memory of the World Programme, American Memory, and The NSF/DARPA/NASA Digital Libraries Initiative (DLI) reveal this fact.

There is no doubt about the fact that digital libraries are large-scale, organized collections of data recorded in various media with information management tools and methods capable of presenting data as useful information and knowledge to people in a variety of

social and organizational contexts. This implies that digital libraries both reveal and require the science and technology similar to those used in traditional libraries for collection, conservation, organization, and access/dissemination. Hence the concepts that dominated the instance quoted at the beginning of this article are faulty. But that will dominate our seen for years to come wasting huge quantum of resources without result and will also be destroying our priceless heritage preserved without applying any such technology.

Anyhow the aim of any digital libraries packages should be to serve diverse communities of users with a broad range of information needs and high levels of expectation. This necessitates digital library developers to look beyond technology into domain, social, legal, economic contexts and be informed by library user and usage studies at each stage of the technologies design and development cycle. Hence it is a misconception to think that if one provides the books to computer people they can develop digital libraries using some available packages or programmes to be made by them for it.

Conceptually, digital libraries mirror collections and library services in the physical world. A digital library is analogous to a 'traditional' library in terms of the diversity and complexity of its collection. Its contents, although residing on electronic media, are stored in familiar forms such as journals, reports, photos, maps, recordings, etc. and referred to using familiar terminologies. Digital libraries draw on many of the same methods of organizing, indexing and cataloguing information as traditional libraries, with software systems automating some of the services performed by librarians in indexing and interpreting users' requests for information.

Digital libraries and traditional libraries are not two competing forms. Digital library technologies will complement traditional library functions and services. It is the librarians who developed the major existing digital library technologies and they will continue to develop it further and will recreate the same feeling to the user as using a very efficient traditional library while using digital library. The goals of the librarians engaged in digital library solution development like Mr. K.H. Hussain familiar to us are to efficiently and effectively create, capture, store, search and retrieve information from electronic collections of text, images, maps, audio recordings, video and film clips and combinations of these. They are experimenting on how best to do this.

Paradigm used by Mr. K.H. Hussain to develop Nitya is adopted from the traditional library information storage and retrieval practices. When a user enters the library, he first goes to the catalogue and searches the catalogue. The user to get the record related to most relevant documents he wants he goes to the document then takes the content page, he locates the relevant chapters or pages or portions of the book. Like this in Nitya, from a dictionary of keywords similar to the library catalogue user goes to relevant documents and through the same process as in traditional library he reaches the peace of information or page of document that it contains in the digital library. So here Mr. Hussain has succeeded to make available to user the same process one uses in the library for his search for documents or information also in digital Library created in Nitya Archive.

DIGITAL LIBRARY DEVELOPMENT

Experience associated with digital library development which includes retrospective conversion of physical media, acquisition of native digital documents in forms that can be used and preservation and archiving of digital materials, indicates that this is possible in the presently available popular digital Library solutions including indigenous Nitya Archive. But the effort required exceeds prior expectations in terms of human labour and cost. And various aspects are to be looked into and thought over while developing digital libraries. The following are the aspects identified by those who have attempted digital library projects.

- Archiving digital material has proved to be as painstaking and arduous as printed material.
- Electronic media, like physical media, become obsolete and degrade.
- New software may not be able to use data prepared for different software or even earlier versions of the same.
- Storage devices and device types become obsolete and the physical media on which data is stored degrades.
- Selecting proper strategies for long-term preservation and use of complex digital objects is essential before undertaking the actual tasks.
- Scanning and indexing records is also human labour intensive.

- Creating new resources that meet quality and completeness requires meticulous handling and documentation of records.

The requirements vary for each type of media and digital object according to the intended use. For example, to scan a very old document as mentioned in the instance quoted at the beginning needs scanning using planetary type of scanners with base adjustments that can position the document in most appropriate way/angle without giving even causing mild pressure to original document, preparation of the digitised raw version of the document for initial processing, software for improvement of the digitised document to achieve maximum similarity with original by removal of distortions due to curves caused by binding, removal of marks and defects in original documents caused by earlier use through time, etc and their loading into the digital libraries. These processes as per experience in creating a digital library requires more than 10 different steps and takes approximately 30 minutes for ancient documents and five minutes for new or healthy documents. It requires manipulating the record, the scanning equipment and the software. Then indexes for the record if available must be digitised and if not available must be prepared in digitised form. The most difficult part of the digital library development is preparation of metadata, the decision of access points preparing abstracts, and the like which have to work in the front end of the digital library. These are technical works to be done by expert librarians. The incidents quoted here in which documents will be transferred to computer people have not envisaged such processes. Finally, the digital file is to be incorporated properly into the logical space of the digital library a parallel process to adding a volume to traditional collection.

One of the distinguishing features of a digital library is the attention placed on metadata collection and management. Metadata describes content, structure, context, and other aspects of a record in digital library and is critical to resource in digital libraries. Unesco and one or two other international organization have prescribed standards for this and following that is very important in digital library initiatives because future development as well as information exchange and cooperation for digital library at national and international levels will depend on this. A challenge for the digital libraries community is to define simple metadata elements that sufficiently describe a wide range of electronic information. All

research and development of this aspect occurred in library environments like that of OCLC and only packages that originated inside library profession efficiently incorporated the results of such research.

Let us conclude this discussion by quoting from an expert report which states: 'It should be noted that, at this relatively early stage in the evolution of digital library technology, it is of vital importance that projects strive for approaches that incorporate high functionality and extensibility. A high level of functionality in the standards and protocols used, even if not fully exploited initially, will postpone the time when the inertia of the installed base begins to confine research opportunities. Careful design of extensibility in digital library systems will facilitate continued research progress and understanding of the impact of new approaches on the user community without the need to attempt to displace an installed base.'

REFERENCES

1. <http://www.geocities.com/cirdinda>
2. RAVINDRAN ASARI, (K.); HUSSAIN, (K.H.) *Nitya archives: Innovative blending of Techniques for selective Access to Information from Digitally Organised Text (SAIDOT)* In: Parthan, S. (ed.). Proceedings of the National Conference on Information Management in e-Libraries, 26-27 February 2002, IIT, Kharagpur. New Delhi, Allied Publishers. 2002. pp. 275-285.