

Problems in Transition to a Digital Library: An Indian Perspective

V.K.J. Jeevan & S.M. Dhawan

Abstract

This paper highlights the problems and issues in the transition to a digital library environment. The problems and issues relate to integration of information technologies in libraries and information centres, in particular to relevant aspects such as IT infrastructure planning, digital library tools and software, models for resource development, IT training needs, resource development, content development, and copyright management, etc. The paper also reflects on the changing roles of the libraries and librarians in the context of emerging digital library environment.

1. INTRODUCTION

The libraries and institutions in developed World are building digital libraries for providing digital information on the web. The trend is picking up fast. On the other hand libraries in India are slow in their march towards digital library environment. As per details available on the INFLIBNET site, at present there are only 26 libraries and institutions in the country which have put up their OPACs on the web. Given the fact that India is one of the IT leaders in the world, the pace of digital library development in the country is rather very slow. The paper attempts to undertake micro analysis of the current situation with a view to identify the problems and issues coming in the way of emerging into digital library environment.

2. DIGITAL LIBRARY:

NOMENCLATURE CRISIS

The concept underlying digital library is being expressed in the literature by several different terms such as electronic library, automated library, computerised library, cybrary, virtual library, hybrid library, and

internet library. These variants of digital library are not totally unrelated and isolated terms. They do bear close relationships since they, in a way, stand to characterise different stages in the evolution of digital library concept. However, divergent interpretation of these terms across different platforms is a source of constant confusion for the professionals especially when it comes to planning, implementation, and evaluation of digital library projects. To an extent, it also reflects our own lack of conceptual understanding of these terms. It is, therefore, important to evolve a common understanding on various digital library variants and the ideas they imply. Even otherwise it is important to address digital library nomenclature crisis since it is a hot topic for research in the IT intrinsic library and information environment. Several experts in the subject such as Lynch and Garcia-Molina¹, Seamans and McMillan², Fox³, and Cleveland⁴ have tried to address this issue in depth. Yet consensus on a single/standard definition of digital library and its variant forms is still elusive.

3. IT INFRASTRUCTURE PLANNING

Digital library projects are cost intensive and as such require extensive funds for the acquisition of necessary hardware, software, network, and humanware. At present, there is no national programme on digital libraries in the country which could offer support for IT infrastructure planning. Libraries interested in pursuing digital library projects have to struggle hard for financial support. They have to plan innovative strategies with a view to sensitize the parent management for budget planning and allocation. Not all libraries are able to succeed in such endeavours and as such planning funds for establishing digital library infrastructure is still a barrier. The country, therefore, needs a national programme on digital libraries for transforming traditional libraries into digital library environment.

4. DIGITAL LIBRARY SOLUTIONS

Typically, digital library resources include on-line catalogues, electronic journals, all types of materials in digital formats: from bit mapped page images to Standard Generalised Markup Language encoded texts and page images stored in proprietary formats, available on the intranet, internet and the web. A digital library has also to offer tools and interfaces for providing intranet access to CD-ROM databases, for archiving digital assets downloaded from internet, accessing and searching locally generated digital assets, creating web-based bibliographic and full-text databases, creating OPACs supporting the metadata standards like Dublin Core, RDF and Z39.50 web client protocol, etc. Besides, it must also offer tools for digitising print material for access and archival purposes.

Internet is offering several public domain databases for access to journal information. They allow access to table of contents, bibliographic information, abstracts, etc. These databases are from commercial as well as association publishers such as Academic Press, ACM, American Chemical Society, American Institute of Aeronautics and

Astronautics, American Institute of Physics, Cambridge University Press, Kluwer, Royal Society of Chemistry, Wiley Interscience, etc. The main difficulty in using such databases is that the information is scattered across various sites. One has to visit to all the sites one by one for retrieving information on a subject. Immediate desktop information from such databases is not feasible. Such solutions are, therefore, not user friendly.

A digital library has also to offer unified, open interfaces to make searches on all the library's digital collections, whether these material originate from a commercial publisher such as Elsevier, an association publisher such as American Physical Society, or internally generated online catalogues. Such a common, open interface certainly helps in freeing the user from working within several different proprietary softwares and ensures thereby seamless access.

Digital library solutions available in India are few and their scope of operations is limited to select few functions such as library management systems, OPACs, and documentation. They have yet to offer intranet-based solutions for accessing and searching locally generated digital assets, CD-ROM databases, archiving digital assets downloaded from internet, creating web-based bibliographic and full-text databases, and for digitising print material for access and archival purposes. Besides they have yet to offer tools for seamless access to information scattered across publishers' sites.

At present, digital library software solutions are marketed from abroad. But they are expensive, and many a times they even fail to provide effective customer support for software implementation, trouble shooting and customization. Free digital library software solutions have just started beginning to appear on the internet. They promise to provide tools for information organisation and publishing on the web but they have yet to establish their credibility.

If digital library movement is to pick up in the country, there is a need to pool efforts to produce indigenous, inexpensive digital library solutions. This is not a tall order, and

leading national organisations like the NISSAT, INFLIBNET, and INSDOC could very well come forward to take a lead in this regard. Digital library offers a lucrative market. Entrepreneurs in the private sector could take advantage of the vacuum and come up with attractive solutions. In India, Informatics (India) Ltd., a Bangalore based company, has offered a gateway on the Internet for providing seamless access to table of contents of 10000 journals. However this is not a public domain database. Its access is limited to the clients only.

5. MODEL FOR RESOURCE DEVELOPMENT

An increasing number of traditional publishing houses and professional associations have embarked on the electronic publishing of academic journals. They are now providing electronic versions of material previously issued only in print. They are also developing new products entirely in electronic form. Secondly, there is a growing trend towards uploading such electronic journals on web sites or making titles available from a variety of publishers. Initiatives in this regard are coming forth from commercial publishers, scholarly societies, and platform providers. They are collaborating with libraries for project implementation and evaluation. A sample of such initiatives includes Project TULIP, Project MUSE, OCLC's First Search Electronic Collections Online, the Association for Computing Machinery's Digital Library, and ISI's Electronic Library project. Currently available products include Springer Verlag's LINK, Elsevier's ScienceDirect service, IOP journals, ACS journals, SAE journals, Web of Science, American Physical Society journals, American Institute of Physics journals, and JSTOR. The ultimate mission of these projects has been to provide access to research material primarily via electronic form.

Commercial publishing houses are marketing their electronic journals under various pricing models. They are offering attractive offers to institutional subscribers. These are bundled offers comprising internet

access to all of their journals at a fractional cost of library print subscriptions. They are also offering attractive offers which are comparatively cheaper for consortia users. Libraries interested in broadening access to journals are resorting to such new offers and have started subscribing to such online solutions. For example, several libraries under the CSIR system have subscribed to Elsevier's ScienceDirect service. IIT libraries are able to avail web access options for over 60% of the subscribed journals.

Although accessing digital library resources via internet is an excellent solution, but not all libraries are still coming forward to take advantage of such new opportunities. They are still debating whether to go for access model or continue to maintain the holdings in print form. Digital asset management is another issue which must concern libraries. In the context of access model, libraries do not store digital resources on their servers. All the resources in electronic form reside on the servers of the publisher and are accessed online on the internet. In the event libraries decide to discontinue the subscription to electronic journals they are bound to suffer heavily since they do not own any digital asset in their collection. Access model amounts to compromising with archival role of the libraries. The choice of model for resource development in libraries continues to be an issue of major concern to librarians.

6. IT TRAINING AND EDUCATION

Developing digital library environment requires adequate knowledge and understanding about elements such as clients, servers, networks (LAN, WAN), internet technologies, and digital technologies. It also requires adequate knowledge and understanding of bibliographic and metadata standards and formats. Traditional library professionals lack understanding of these concepts and such limitations are coming in the way of planning and implementation of digital library projects.

There are several agencies in the country which are engaged in IT training and

education. These include the library associations and library networks, NISSAT, INSDOC, DESIDOC, INFLIBNET, DRTC, NASSDOC, NIC and NSIC. Their training programmes are focused largely on library management systems. There are less number programmes which focus on electronic library solutions or digital library solutions, covering topics such as XML, HTML, imaging, OCR, web designing, CD databases for access over web, LAN and intranet, web integrated databases, and selection and indexing of materials for digitising. The list of topics is only illustrative and not exhaustive. Such digital library topics are also needed to be covered in the university course curricula. But, it is too much to depend on university education system for training and education in the IT. Secondly, the training programmes are usually of short term duration, from one week to two weeks, and the duration is not sufficient to provide adequate hands on experience. Such training programmes do serve the purpose of giving exposure to IT concepts but fail to equip the working librarians with adequate training and knowledge that should enable them to lead digital library revolution. Thirdly, such training programmes are expensive and not all working librarians and information professionals can always afford them.

The impact of any IT training programme must be judged from its outcomes. It should enable librarians to publish their products and services on the web, share their resources on internet, and manage their digital assets for long term use. In reality, it is hard to visualise such outcomes. This is because we lack content and hands on practice in the course curricula of training programmes, lack trainers and the training agencies, lack infrastructure and funds for the purpose, and also lack motivation on the part of the working librarians to undergo training. We need to evolve a well structured programme for filling up the vacuum. The country needs to take initiative in case we want make visible impact in digital library revolution.

7. COPYRIGHT MANAGEMENT

Commercial publishers are offering online access to their subscribers through password validation or through IP address validation. For this the commercial publishers require subscribing institutions to sign license document. The license document is an instrument to protect copyright interests of the publishers. As a responsible social service institution, it is the prerogative of libraries to effectively guard access to its collection and to ensure that no copyright content is misused by anyone through its services. Therefore the librarians have to ensure that copyright document is honoured. However, of the commercial and association publishers insist that the signatory to license document has to be head of the subscribing institution. They do not regard the authority of the librarian as sufficient enough in this regard. Many a times, meeting such stipulated requirements lead to delays and unnecessary interruption in online service to e-resources.

Commercial publishers are using copyright as an instrument largely for commercial gains. For example they regard multiple access to e-products from a single site as copyright violation unless so permitted against additional payment. The charges for single site multiple access are higher than those for the single site single user access. Secondly, copyright conditions are coming in way of utilising digital resources for the purpose of resource sharing and information marketing. Libraries are being prevented from giving e-resources for a fee as otherwise they will invite punitive action under the copyright law. For subscribing libraries this is still a puzzle. How to get over such copyright restrictions is still a puzzle for subscribing libraries.

8. INFORMATION STORAGE & PRESERVATION

The storage and preservation of contents are the two key issues that have to be closely monitored in the digital library development. The contents once digitised are to be stored on the servers for a sufficient duration to stabilise the confidence of clients. The basic motivation behind preservation is not only to

ensure the permanency of contents, as is the case for printed contents, but also to make them amenable to the latest software to guarantee the persistent use. Preservation is the aspect of archival management that preserves the content as well as the look and feel of the digital object⁵. The data migration dictated by hardware/software upgrades or changes is another underlying concern in preservation. The intricacies of digital preservation are further complicated due to the presence of different formats and diverse systems to achieve similar tasks, and even the multiplicity of standards. After briefing about the virtues of formats like TIFF, PDF, and markup languages (HTML/SGML), Hodge observed that "the key preservation issue is the format in which the archival version should be stored".

9. RESOURCE DEVELOPMENT

The main concern in resource development is about the rising costs of e-journals and e-databases and on the other hand about the shrinking purchasing power of the libraries to afford such high rising costs. The other concern is about the approach libraries continue to follow with regard to resource development. There are hardly any efforts in evolving cooperation between libraries for developing comprehensive collections. Put together libraries in India are spending over Rs.150 crore per year on journal subscription in science and technology. They are still not able to offer comprehensive information in this area. At the best they are getting not more than 3000 to 4000 journals in science and technology whereas the country needs more than 15000 journals.

The libraries need to follow a multi-pronged strategy to ensure comprehensive access. This may involve adopting consortium approach to journal subscriptions or digitisation of print resources on cooperative basis or adopting new access model which should allow access to electronic journals and books for a fee. They can also establish subject information gateways such as in physical science, biological and medical science, and agricultural sciences. These

information gateways will have the mandate to build digital library resources and make them available country-wide via internet. The issue is who should give leadership in resource development.

10. CONTENT DEVELOPMENT

Our current efforts aimed at retro-conversion of cataloguing records at the library level have not succeeded in converging and pooling the outputs into a national union catalogue of books and journals available in the country. Identifying book availability at the individual network level might have become feasible but it is certainly not so feasible at the national level. The problem lies in our failure to eliminate duplication in our efforts in retro-conversion, Organise coordinated efforts in retro-conversion at national level, and the end less multiplicity in our opinions on standards and formats for retro-conversion. Content development activity at local library level has been the current approach to developing OPACs. This alone may not lead us to a digital library environment. There is an urgent need for a national bibliography utility in the country charged with the responsibility to offer bibliographic records of books and journals as per standard record formats. The utility may also offer facilities for data conversion into various record formats and meet thereby libraries demands without embroiling itself into any record format controversy.

The technical reports, conference proceedings, theses, etc., constitute an important institutional resource. They need to be preserved and archived for current and future use. The libraries should digitise them and make them accessible via internet without any fear of copyright infringement. In this regard the libraries will have to focus on standards for content creation, identification, organisation, storage, and hosting. For succeeding in such endeavours, the libraries will have to acquire high level skills and expertise in bibliographic and full text database development activity.

11. ROLES OF LIBRARIES AND LIBRARIANS

Publishing on the web is a highly visible activity with relatively little investment, as the software is available absolutely free or for a nominal charge, and the main overheads are hardware and personnel costs. The libraries may now be required to adopt this new role. This should be in addition to their role of computerising their routine tasks by using library management systems. They will also be required to use CD-ROM systems as an alternative medium for information access and delivery. These activities are going to make the role of libraries and librarians even more challenging.

In the evolution of digital libraries the professionals in the field have started apprehending a threat to their survival. They consider specialists from the computer field as their competitors who might eventually take their roles. They also fear threat from publishers and content aggregators who are offering online access via internet to electronic journals and journal archives. Some open access archives on the internet are making many primary materials available at no cost⁶. Admittedly, web has enabled us to accumulate a large quantum of digital contents. It is demonstrating the viability of paperless systems that could compete with the traditional setup. While such threat perceptions to the professionals are logical, what is not being understood is that digital library solutions are still very expensive and not affordable by all libraries outside the developed world. Even otherwise library skills in knowledge management are time tested ones. It is erroneous on our part to perceive that librarians could be replaced by computer specialists. Computer specialists are indeed capable of writing software for managing information overload. But it is only the library professionals who with their knowledge management skills are able to reduce noise in information retrieval and also improve precision in searching information from databases on the internet.

We have to understand that digital libraries are adding new dimensions to activities pertaining to information organisation and management. It is about managing digital contents including multimedia products and publishing them on the internet and intranet.

12. CONCLUSION

The digital revolution offers tremendous potential to add value to the process of information creation, dissemination, and access. It is soon going to reshape the roles of libraries and librarians. At present, the libraries and information centres are responding as intermediate users to the initiatives taken by commercial publishers, scholarly societies, and platform providers. This alone is not enough. They must also respond as platform providers. They should coordinate their activities in creating new information gateways and aim at providing access to their resources on the internet and other networks. They need to publish on the web, use IT as a tool for expanding their scope of the services from document delivery to information delivery and gradually graduate to address issues of knowledge management. In achieving this objective they must first address the issues and problems highlighted in this paper.

13. REFERENCES

1. Lynch, Clifford, & Garcia-Molina, Hector. Interoperability, scaling, and the digital libraries research agenda: A report on the 18-19 May 1995, IITA Digital Libraries Workshop, 1995.
<http://diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.html#2> (accessed on 17 June 2001)
2. Seamans, Nan, & McMilan, Gail. Digital library definition for DL12, June 16, 1998.
<http://scholar.lib.vt.edu/DL12/defineDL.html> (accessed on 15 June 2001)
3. Fox, Edward A., *et al.* Digital libraries. *Communications of the ACM*, 1995, **24**(4).
4. Cleveland, Gary. Digital libraries: Definitions, issues and challenges. (UDT Occasional Paper #8), March 1998.

- <http://www.ifla.org/VI/5/op/udtop8/udtop8.htm> (accessed on 25 June 2001)
5. Hodge, Gail M. An information life-cycle approach: Best practices for digital archiving. *Journal of Electronic Publishing*, 2000, 5(4).
- <http://www.press.umich.edu/jep/05-04/hodge.html> (accessed on 7 July 2001)
6. Arms, William Y. Key concepts in the architecture of the digital library. *D-Lib Magazine*, July 1995.
<http://www.dlib.org/dlib/July95/07arms.html> (accessed on 8 July 2001)

Contributors: **Shri V.K.J. Jeevan** is Assistant Librarian at Central Library, Indian Institute of Technology, Kharagpur-721302, West Bengal. E-mail: vkj@library
Shri S.M. Dhawan