

# Pragmatic and Usable Approach for Digital Library Initiatives in India

Dinesh S. Katre

National Multimedia Resource Centre

**Centre for Development of Advanced Computing (C-DAC)**

Ministry of Communications and Information Technology, Government of India,

Agriculture College Campus, Near District Industries Centre,

Shivaji Nagar, Pune 411005, India. Tel/Fax: 91+020+5533250

[dinesh@cdacindia.com](mailto:dinesh@cdacindia.com)

## **Abstract:**

Cultural and heritage organizations in India have to undergo a major phase of transformation in terms of their mindset, style of working and objectives, procedures, tools and technologies while proposing to build Digital Libraries (DL). Sponsors of DL initiatives in India must notice the rate of technology obsolescence and the pace of our cultural organizations in adoption of new methods. DL technology can never be absorbed and sustained unless all transformational aspects are addressed or it will end up as yet another artifact in a museum. In short, it is extremely important to study the socioeconomic conditions while choosing the correct approach or roadmap to digital library. Success of digital libraries in the developed countries is due to continued efforts of modernization in the past several decades. Effective partnership between the prime stakeholders such as cultural, technological and educational organizations has helped in achieving greater level of acceptance to digital libraries. DL technologies without such interdisciplinary and collaborative efforts will not be able to capture the requirements of curators, historians, archeologists, critics, research scholars, educationists, students and readers.

## **Introduction:**

The myths described in this paper highlight the concerns of Human Computer Interface (HCI) issues relating with Digital Library (Dillon, 2002). Insufficient awareness of HCI in India (Nielson, 2002) is leading to non-usable technological artifacts. Lessons have to be learnt from the failure of some of the e-governance applications in India considering its similarity with the Digital Library domain. Technology developers, content providers and government bodies in India have their own viewpoints of Digital Library. A platform for all them to influence each other's viewpoints is not available and that has resulted in mythical perceptions. Some of these Myths are enlisted in this paper. The paper also presents some of the experiments carried out by C-DAC's National Multimedia Resource Centre, Pune that attempt to counter the HCI issues relating with large information systems and also demystify the myths.

Following myths are prominently discovered among the DL technology developers in India.

**Myth 1.** Always, it is best to go for highest possible resolution while scanning.

**Myth 2.** Bigger the storage needs better the digital library.

## **Reality:**

Most DL technology developers have propagated the need of high resolution scanning of all images in a rather unrealistic manner. Increased storage needs gives them a business advantage. The large collections of manuscripts amounting high resolution scanning of several millions of folios inflate the Digital Library infrastructure requirements that include High Speed Scanners, Storage Area Network and Large Databases. Such infrastructure is very demanding in terms of its effective usage, maintenance and sustenance. The organizations dealing with Indian heritage do not have the much-needed IT culture, technical manpower and business model for sustaining such a facility. Many private museums cannot even afford such unwieldy infrastructure. Considering these prohibitive factors, experiments are carried out to find out the rational for justifying the allotment of DPI resolution for objects while scanning.

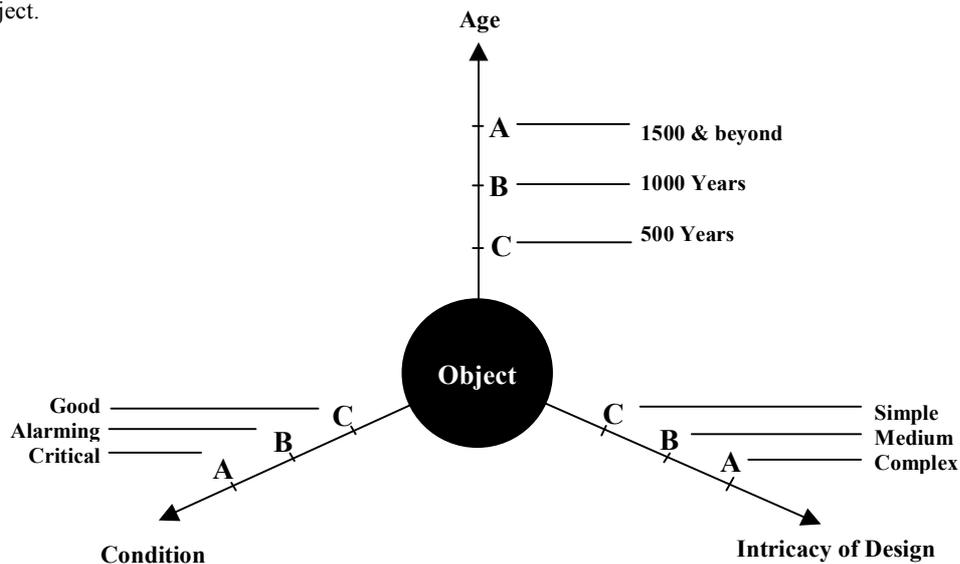
**Variable Allotment of DPI Resolution:**

Most digital library projects tend to choose uniform DPI resolution for scanning or they change it randomly. Contrary to this, the idea is floated to allocate appropriate DPI resolution on the basis of importance and need of the object. This helps in regulating the storage needs of a Digital Library.



**Figure 1. A pot and a vase with varied amount of intricacy of design**

Museum collections of antiquities were observed very closely to discover that curators attach different degree of importance to different objects (Chapman and Kenny, 1996). Objects, which are very ancient, precious, fragile and in critical condition are given special attention and care compared with other objects. Similar logic could be applied while scanning as well. As shown in Figure 1. Pot A need not be scanned at very high resolution, as its surface does not have any designs. Whereas, Vase B has a lot of intricate designs on its surface which may be captured well if the image is scanned at higher resolution. Following instrument is devised for logically deciding the amount of DPI resolution to be allotted for scanning of an object.



**Figure 2. Instrument for allotment of DPI resolution for scanning**

The three major aspects of an object namely its Age, Condition and Intricacy of Design are chosen as parameters for calculation. Every parameter is measured on the scale of A (1), B (2), C (3). The values mentioned in brackets are used for calculation. Hypothetically, if one is going to scan the images max. at 300 DPI resolution then the above parameters help in gradient allotment of DPI for every object. For example if an object is rated as under-

$$\text{Condition} = C (1) + \text{Intricacy} = C (1) + \text{Age} = C (1) = 3$$

The rating is 3 out of 9. Proportionately, the object is allotted 100 DPI resolution for Scanning. The highest ratings (9 out of 9) will permit you to scan an image at 300 DPI. Other ratings such as [Condition = A (3) + Intricacy = C (1) + Age = B (2) = 6] indicate intermediate DPI sizes commensurate with the needs of a particular object (i.e. 200DPI). The denominations of Age parameter can be fragmented further and they could be beyond 1500 years. For the simplicity of presentation, Age is denominated up to 1500 years only.

This approach helps in logical allotment of DPI resolution and effective usage of storage space available. It also disproves the myth that all images have to be scanned at highest possible DPI resolution. C-DAC's National Multimedia Resource Centre has applied this technique while designing a low cost Digital Library solution for a museum. It also saves on time required for scanning of high res. images. A manageable Digital Library with modest storage needs can keep the collections in circulation and thus preserve better.

**Myth 3:** Be it antiquities, manuscripts or whatever, after scanning they all end up as JPG files.

**Myth 4:** Support of various file formats is enough to satisfy the needs of museums, libraries or whatever type of collections (There can be one digital library solution to all type of contents.)

**Myth 5:** We don't need to indulge in content creation for DL technology development.

**Reality:**

Most museum artifacts end up as images during digitization. Though they are all images, every artifact has a different story to tell. It is necessary to go beyond typical library attributes such as Accession No. Classification, Author, etc. Scanned image of a manuscript, a sculpture, a miniature painting, an architecture, a book, a weapon, etc. serve different purposes and experts can extract different information from each of them. Therefore, it would be wrong to treat them just as JPG files.



Figure 3 Measuring a folio

**Need to see beyond file formats**

C-DAC's National Multimedia Resource Centre is developing a Digital Library solution to handle large collection of manuscripts. It has a measurements module that helps in capturing true properties of a manuscript folio. User interface for measuring vertical and horizontal margins, letter height and width, leading space, line width, no. of letters in a line and a folio, stroke width and angle, border thickness and

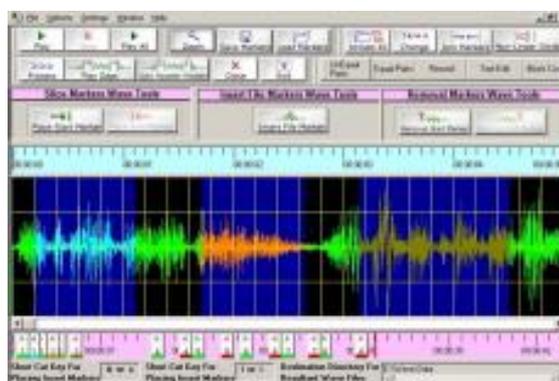
color of a folio. These are the true attributes of a folio unlike the associated library attributes. The search parameters are based on the folio properties in addition to library parameters. This example clearly brings out the fact that every image has different properties depending on its content although the file format may be JPG. These requirements were observed from the viewpoint of a manuscript expert. Usually, it is perceived that every museum artifact is a single entity but a manuscript contains several folios though it may be treated as a single entity from the library perspective.

This also brings forth the fact that specialized tools and templates would be required for extracting / describing the true attributes of contents and managing different types of collections. Common DL software supporting only library attributes will handle the contents at a superficial level.

Myth no. 4 & 5 are typical examples of the attitude of most DL technologists towards content creation. DL technology is also partly responsible for content creation processes. Following example enumerates how content creation tools can be discovered for meeting the requirements of Digital Library.

#### **Productivity tools for content creation**

We carried out a full-scale case study of content creation on *Dnyaneshwari* and *Bhagavadgita* in order to design DL software for Saint Literature. During this effort, we had to digitize and edit the recording of vocal rendition of verses. It was a mammoth task as the total duration of recitation was more than 240 hrs. We had to edit each verse and create a separate WAV file to be integrated along with the text (All together 13,000 audio files were to be edited). This task became very tedious, as the existing audio editing tools are not designed for such voluminous nature of editing. It was eye straining to mark start and end points of verses in a WAV form. Why should one strain the eyes during audio editing? It's an HCI problem. Therefore, we designed an audio editing tool that allows defining multiple segments of editing by key pressing (In and Out markers) while listening to audio. One does not require to even looking at the WAV form. The tool also supports segment wise annotations, incremental alphanumeric naming standards and auto addition of fade in and out patches while saving the file. The tool can slice hundreds of segments in no time.



**Figure 4. Editing tool for rendition of verses**

Such productivity tools may prove useful only in certain contexts and situations faced while creating large amount of content for Digital Library and they can never be conceived without indulging into content creation.

**Myth 6.** Presentation and comprehensibility of information are not our problems.

#### **Reality:**

##### **Library Atmosphere**

Comparison between conventional and digital libraries opens up a variety of HCI issues. If we recall our experience of conventional libraries, we predominantly remember the environment of library, which includes the architecture, the garden, the quiet corridors, and various sections partitioned by cupboards and book shelves, cool breeze coming through open windows, readers sitting in comfortable chairs, some readers discussing in low voice and a librarian with his assistant in one place, etc. This atmosphere attracts readers to library. Similar environments are created for museums as well. Present digital libraries lack such environmental qualities that create right kind of mood. We remember the books or artifacts through spatial

clues, colors, shapes and titles. We recall a book with red cover, kept in the topmost floor of the shelf in 'History' section of library. 'History' is the fourth section towards the right side of second floor. Such spatial clues are essential for our cognitive convenience. Most contents in a digital library are presented in the form of a hyper linked structure. Contributing spatial feel to digital library interface is a challenge.

### 3D conceptual model of *Bhagavad-Gita* Temple to Serve as Index

The metaphoric temple is constructed as 3D model in VRML format to serve as 3D index (Figure 5.). Links provided in three-dimensional space offer spatial clues while accessing the contents of Digital Library. We are experimenting this idea in our lab. In the final, eighteenth chapter of *Dnyaneshwari*, Saint *Dnyaneshwar* says that this chapter is like the pinnacle of the *Bhagavad-Gita* Temple. From this highest point you get the Aerial view of previous seventeen chapters. Then he describes earlier seventeen chapters by linking them with the remaining architectural parts of temple. Other philosophical literatures such as *Upnishadas*, *Vedas*, *Mahabharata*, *Navrasas*, etc., are visualized while describing the premise and surroundings of *Bhagavad-Gita* Temple (Katre, 2002).

#### Following advantages are observed in this 3D metaphoric index:

- 3D model in VRML format allows creation of hyper links over objects.
- The user is able to recall the links with visual association e.g. *Vedas* on the steps of the temple at entrance, *Nyayshastra* on the pillars inside the temple, *Navrasas* as colorful lamps, etc.
- The effort reduces cognitive load on users.

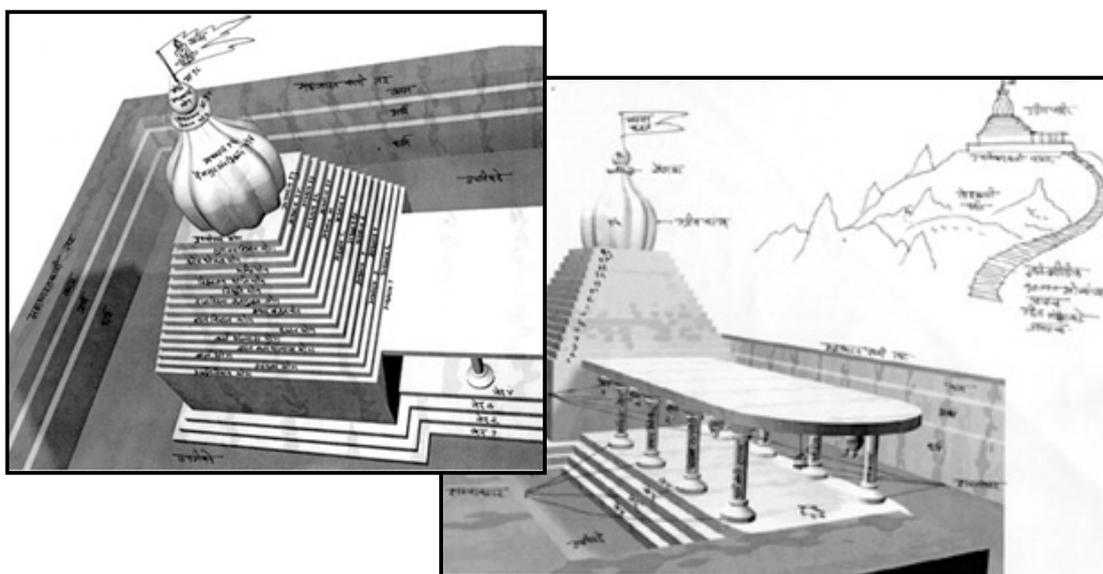


Figure 5. 3D Index for spatial navigation through digital library

Most contents after integrating as part of DL lose their usual look & feel. During our experiment of presenting *Dnyaneshwari* as a sample of Saint Literature we have attempted very hard to retain its cultural face in terms of layouts and other interface components (Figure 6.). In summary, the usability concerns of DL technology should be extended further to meet the comprehension needs of its end users.

Following myths are about content providers.

- Myth 8.** Content providers will-
- have complete information about every item that is digitized.
  - identify the information in single effort.
  - always give final and correct information.
  - give just the required amount of information.

**Myth 9.** A training program for content providers will help them internalize the content integration processes.

**Reality:**

It is observed that DL technology developers succumb to Myth 8, and provide a plain form for data entry. They assume that every record is bound to be complete and correct at the time of integration. One tends to visualize such idealistic scenarios using UML (Unified Modeling Language) but real life scenarios are influenced by diverse human and non-human factors.



**Figure 6. Presentation of saint literature**

Our close interactions with domain experts and observation of procedures practiced by museums have revealed that information records are prepared through iterative processes. A record involves contributions from many individuals who specialize in different subjects. Content is checked and approved at various levels before it gets integrated in Digital Library. Also, it takes a number of days to get a set of records completed. The Digital Library software designed by NMRC allows inputs from multiple experts and iterative revisions before a record gets integrated in the database. In addition to this, progress reporting for Data Entry Operators and Curators is also provided.

The chances of DL technology not jelling with museums or libraries are high due to foreign body syndrome. For its seamless integration, linkages with concerned manual processes become important. One should not rule out a particular transaction of a museum because it has no direct connection with the content prepared for Digital Library. If the DL processes are logically interlinked with manual processes the chances of success are higher. Training of content providers alone is not enough for them to internalize the DL processes.

Following myths are about the subscribers of Digital Library.

**Myth 10.** All subscribers / visitors-

- are computer literate.
- precisely know what they are looking for in a Digital Library.
- have adequate vocabulary to input correct keywords and phrases for searching.
- will search and get what was wanted.
- read books or visit museums from digital library.
- will enjoy reading from digital library.

**Reality:**

**Navigation Model 1. 'From Known to Known'**

On visiting a Digital Library one comes across a small input box wherein he/she is expected type their query for searching. This is like when I started looking at books in a shop at Mumbai Airport, the shopkeeper asked me the name of book that I was looking for. When I told him that I did not have anything specific in my mind, he requested me to leave the shop, thinking that I will be wasting his time. Most Digital Libraries treat you in a similar manner. This is a 'from known to known' model of navigation. Of course you also get some 'unknown' information as a search result.

**Navigation Model 2. ‘From Unknown to Unknown’**

One can also visit the library without any specific book or title in mind. We often visit the library without any objective, spend a lot time looking at variety of books at random and then select something interesting. Digital Libraries should permit such free wondering through its collection. This would be ‘from unknown to unknown’ model of navigation. In this model, it is not possible to start by typing a keyword for searching. One should be able to access the contents directly like we pull out books from the shelf. The book attracts our attention by its cover design, title, author, etc. Digital Libraries should offer both models of navigation.

Following myths are found among Indian museums, libraries and organizations that manage large collections of manuscripts.

**Myth 11.** Subcontract entire digitization / integration work and get a readymade digital library.

**Myth 12.** Our collections cannot generate revenue.

**Myth 13.** The business model for digital library will succeed immediately.

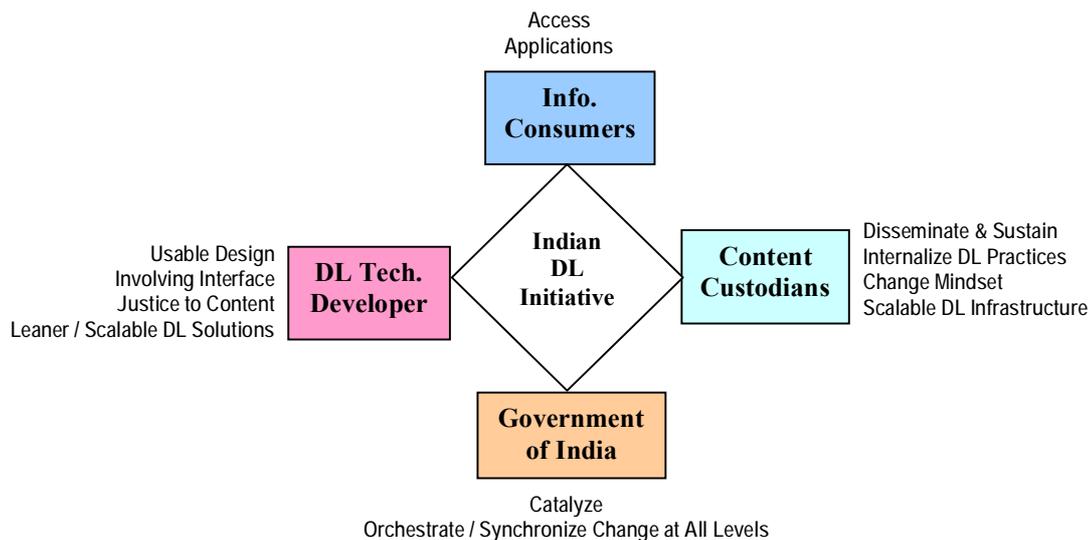
**Reality:**

Digital libraries will sustain only if they become part of our regular work culture. It is important for the library staff to internalize these work processes through participation, as digitizing does not replace the need to perform all the traditional tasks like acquiring, organizing, cataloging, and preserving materials (Erway, 1996). The staff must get familiarized with content integration process and the idiosyncrasies of computers. Outsourcing entire digitization and integration work will not give this experience to library staff. Managing and sustaining such digital library will be difficult for inexperienced library staff. Regular maintenance and upgradation of hardware and software can happen only if the DL setup generates some revenue. The collections are valuable and they must be made accessible to outside world for some cost. Persistent efforts should be launched for building a strong base of subscribers. Indian collectors must become aware of the value of IPR and modes/models of sharing the IPR for a price. The bottom line is ‘cultural heritage is best preserved by keeping it in use / circulation’. Digital mummification alone is not enough.

Following myths are sensed from government policies and actions.

**Myth 14.** Giving funds to organizations for building library information systems is enough.

**Myth 15.** Cultural, technical and educational organizations will build synergy on their own.



**Figure 7. Pragmatic approach for DL initiatives**

**Description of participants shown in Figure 7 -**

**Custodians:** Museums, Libraries, Archives, Manuscript Collections, Private Collections of Heritage Artifacts and other valuable items, etc.

**Info. Consumers:** Research Scholars from India and Abroad, Research Organizations, Educational Institutes, Schools and Colleges, Teachers, Students, General Visitors

**Government of India:** Funding Organizations, Policy Makers, Ministries, Controlling Bodies dealing with culture, education, technology, planning

**DL Tech. Developers:** Computer Professionals

All four participants of the activity have to be brought into play in a synchronized manner in case of building information systems for preservation and dissemination of our national heritage. Scottish Cultural Resources Access Network (SCRAN) founded by national museums in Scotland mentions about National Grid for Learning in its report. The report compares educational networking to an electricity supply system (Royan, 1998). The educational syllabus for schools and colleges should be appropriately redesigned and linked with proposed digital library initiatives. Educational CD ROMs, e-learning programs for transfer of Indian History, Culture, Sciences should be developed as extensions of these information systems. Present Digital Library initiatives in India are implemented in isolation and without the required partnerships and orchestration.

**Conclusions:**

Leaner and scalable approach to Digital Library will be most suitable for Indian conditions. Technology developers have to be sensitized about Human Computer Interface (HCI) design approach. Government of India should initiate, catalyze and synchronize the collaborative efforts between technology developers, content providers and educational institutes.

**Acknowledgements:**

I am very grateful to Dr. Mukul K. Sinha, my Ph.D. Supervisor and Managing Director of Expert Software Consultants, Pvt. Ltd. for his valuable guidance. Also, I acknowledge the support of my group members at C-DAC.

**References:**

Chapman, S. and Kenny, A.R. (1996). Digital conversion of research library materials: a case for full informational capture. *D-lib Magazine*, October, 1996.

URL: <http://www.dlib.org/dlib/october96/cornell/10chapman.html> (Accessed on 12 Dec. 2003)

Dillon, A. (2002), Technologies of Information: HCI and the Digital Library, Human Computer Interaction in the New Millennium edited by John M. Carroll, Published by Pearson Education Asia, pp 457-474.

Erway, R.L. (1996). Digital initiatives of the Research Libraries Group. *D-Lib Magazine*, December, 1996. URL: <http://www.dlib.org/dlib/december96/rlg/12erway.html> (Accessed on 12 Dec. 2003)

Katre, Dinesh, S. (2002). "Unconventional Inspirations for Creating Software Interface Metaphors" Proceedings of International Conference on Media and Design (ICMD 2002), Volume I, Mumbai, India, pp. 1-15.

Nielson, J. (2002). Offshore Usability

URL: <http://www.useit.com> (Accessed on 12 Dec. 2003)

Royan B. (1998), SCRAN: a case study of Networked Cultural Multimedia for Education, e-Proceedings of INTERNATIONAL CONFERENCE ON MULTIMEDIA FOR HUMANITIES, IGNCA, New Delhi.