

---

## BUILDING AN IMAGE REPOSITORY FOR INDIAN CULTURAL HERITAGE MATERIALS – A DIGITAL LIBRARY APPROACH

P Lalitha      T A V Murthy      M T M Khan      Manoj Kumar K

### Abstract

*Digitization has immense potential to store, preserve, disseminate and redistribute cultural materials, if the materials are captured in a bit stream format so as to enable to access through network for learning and understanding the history and culture of a civic society. Generally these materials are stored in museums or in libraries, but they can be spread over the globe and with the spread of the Internet, such contents can be made available to any corner of the world. Universal access implicitly assumes a largely unidirectional flow of information from the developed world to the developing world but the latter is often richer in traditional wisdom, crafts, and cultural artifacts: digital technologies have the potential to preserve that heritage and make it accessible.*

*For developing an Image Repository for photographs of Ancient cultural material by using the advanced digital technology is a challenge, which include description of the photographs of Indian cultural Heritage Materials, Creating Metadata from the sample, Mapping the created metadata with Dublin Core, Developing an Image Repository System using Digital Library open source software etc. This paper is the result of the case study of such an attempt by the authors on exploring Open Source Digital Library Software, DSpace, for preserving the cultural heritage materials.*

**Keywords :** Cultural Heritage Material, Digital Library, DSpace, Dublin Core, Photographs, National Repository, Images

### 1. INTRODUCTION

India is the home of world's 'great civilizations' with a social structure that can be traced back for thousands of years. The Indus valley civilization was the India's first major civilization that flourished a thousand years around 2500 B.C. This civilization is known to be the birthplace of Indian art and sculpture which depicts enormous skills and talents of those people. Indian art portrays virility and vivaciousness, which is refined and sophisticated, bold and robust. Cultural heritage materials have a lot of importance, as culture is something we do, the performance which fades into memory then disappears, but the record of culture consists of artifacts which we make, which may persist but inevitably decay. [3] The artifacts may be different materials like manuscripts, photographs, drawings, Paintings and three-dimensional sculpted materials etc. Out of these photographs play a vital role in disclosing the facts of history. A picture is worth of a thousand words, and object for object, pictures are several orders of magnitude larger and more subtle information carriers than written language. [1] Due to their importance in different fields like education, research, history, archeology and anthropology, most of the cultural and heritage materials are being converted into the digitized forms knowing that permanent access to this heritage will offer broadened opportunities for creation, communication and sharing of knowledge among different communities, as well as protection of rights and entitlements and support of accountability. [4] 2005)

### 2. IMAGE DATABASE RETRIEVAL SYSTEM

There has been an increase use of images as there is decrease in the cost of acquiring and storing them. Image Retrieval problem is concerned with retrieving images that are relevant to user's requests from an Image Database.[5]

---

An Image Database System is set of images, which are collected, analyzed and stored in multimedia information system. They may comprise different kinds of systems like office systems, Information Retrieval Systems, earth resources systems, medical databases, virtual reality systems, robotics systems, art galleries museum catalogues, animal and plant atlases, sky star maps and meteorological maps etc. [8]

### **3. IMPORTANCE OF METADATA IN IMAGE DATABASE RETRIEVAL SYSTEM**

The full and seamless accessibility, according to the end user's perspective, would require that these materials should be made accessible through automated searches from a number of entry points. For the effective management of image information; image database retrieval systems have been constructed. During past years a large number of Content-Based Image Retrieval (CBIR) systems have been developed. It is a challenging task to select most appropriate system, as it is almost impossible to do an exact comparison of database image with each query image. It is also difficult to evaluate how successful the CBIR systems are in terms of effectiveness, efficiency and flexibility. It has been noticed by the researchers that there are significant *gaps* between the features we can extract from the images and the meanings of the images. According to Smeulders et al. these gaps are the *sensory gaps* between the object in the object in the world and the information in a (computational) description derived from a recording of that scene. The *Semantic gap* is the lack of coincidence between the information that one can extract from the visual data and interpretation that the same data have for a user in a given situation.[7] It is important to bridge the gap by referring to descriptive metadata. The descriptive markup provided by the subject specialists remain the most precise and reliable recourse and will continue to be an invaluable guide to any development of automated search strategies.[2]

For the fast and meaningful retrieval, it is essential to develop Image Repository System that supports good metadata standards. Metadata is also useful in identification, representation, interoperability, technical management, performance, and use of data contained in an information system.

### **4. BUILDING OF IMAGE REPOSITORY SYSTEM (IRS)**

An attempt has been made to collect photographs of different kinds of materials, which depict the rich culture heritage of India. Amongst the collected samples there are photographs of manuscripts, drawings, paintings, sculptures and heritage sites. Digitization of the photographs is an integral part of building of IRS for photographs. The process comprises following steps:

#### **4.1 Image Capture**

Before going for imaging it is very important to note the characteristics of scanning devices such as optical resolution, dynamic range, registration of colour channels, bit depth, noise characteristics, and quantization controls are to be evaluated carefully considering the final use of the digital image.

#### **4.2 Image Analysis**

After scanning the image, it is to be evaluated by viewing it on a monitor. While doing so the different defects like dirt, half images, skew and so on can be detected.

#### **4.3 Image Enhancements**

In case of any discrepancies have been noted while analyzing the images, then they have to be enhanced at the cost of image authenticity and fidelity. The typical image enhancement features are tonal reproduction, curves and colour management, crop, image sharpening, contrast, transparent background etc.,

#### **4.4 Image Storage**

The digital archives should adopt a different storage strategy. Digital data has to be safeguarded and regularly maintained. Storing of images on CD-ROM is recommended.

---

As on date, there is a great need to provide information on web interface, as it helps in finding the information accurately, can be located easily and fastly. The increasing use of the Internet and World Wide Web has developed awareness about the access and retrieval of information across networks. The major boost to digital libraries comes from the web technologies that enable instantaneous online access to repositories. Digital libraries encapsulate a whole range of information services, such as organization of digital information, information retrieval, user interfaces, archiving and preservation, services and social issues, evaluation and application to particular areas and a set of standards for interoperability and value-added services.

#### **4.5 DSpace Digital Library Open Source Software**

In the present study Dspace Open Source Digital Library Software has been selected as it has the following features:

- Dspace is open source software that supports trouble free Digital Library operations. It can capture, store, index, preserve and redistribute content in digital formats.
- Dspace is a service model for open access and /or digital archiving for perpetual access.
- Apart from the above features, it also has:
  - Architectural design – Modular and Open
  - Backend Database- Scalable, robust, data formats
  - Network capabilities- web-based and seamless operations, persistent Ids, security and authentication
  - Metadata and Interoperability – compatible with world standards such as Dublin Core and OAI-PMH
  - Search, Retrieval and Display system- Type of Indexing, Ontologies, etc
  - User interfaces and customization- Annotations, Subscription
  - Ingest – XML based import/export [6]

The digital preservation in DSpace is in the form of bitstreams. For images DSpace accepts bitstreams in many file formats like jpeg, tiff, gif etc., which is most convenient for incorporation of the images into the software.

#### **4.6 Features and functions of the DSpace**

- Metadata
- User Interface
- Work flow

##### **4.6.1 Metadata**

DSpace basically uses a qualified Dublin Core Metadata for describing items intellectually. This metadata is displayed in the item record in DSpace, and is indexed for browsing and searching the system. Dublin Core is a set of 15 attributes divided into three groups, i.e. content, intellectual property and instantiation. Associate to Dublin core are Dublin Core qualifiers that enhance the identification of items. Most of the institutional repositories use Unqualified Dublin Core metadata to ensure interoperability.[9] Since OAI is based on the exchange of metadata, getting the metadata right is fundamentally important for a repository. The OAI compliant DSpace software automatically produces the necessary Dublin Core metadata for harvesting by service providers.

Description of the metadata according to Dublin core Registry of DSpace :

In the present study out of 65 *Dublin Core qualifiers*, of Dublin Core Registry of DSpace, 16 *qualifiers have been selected*. The metadata created from the original photograph with its target subject had been mapped with these qualifiers.

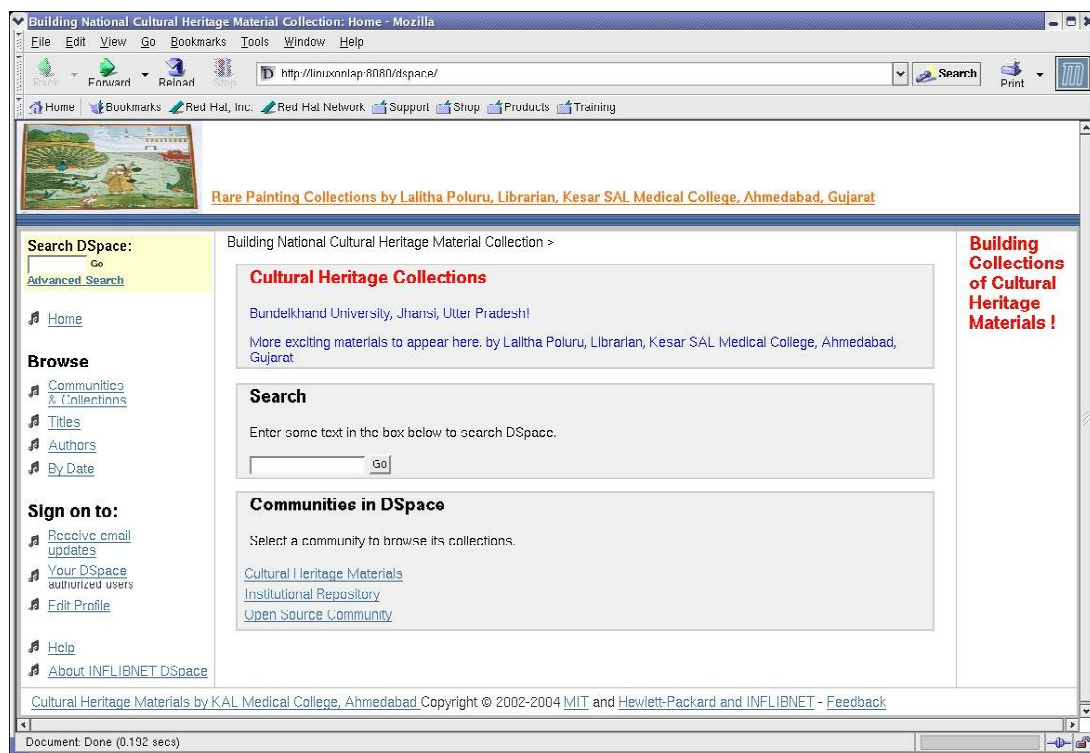
S.NO	ID	ELEMENT	QUALIFIER	SCOPE NOTES
1	5	Contributor	Illustrator	
2	1	Contributor		
3	7	Coverage	Spatial	Deccan
4	12	Date	Available	1700 Century
5	27	Description	Abstract	ATortured by separation wearing a yellow garment, gone to the forest carrying flowers and plays the reciting his beauty, she returns not; her fair body excites everyone's mind". In an open wood land she sits on a mound and plays the vina. The peacocks are attracted enamored by the music. Kakubha t be sung in the month of March.
6	28	Description	Provenance	Deccan Provenance
7	34	Format	Extent	6 X 9 ½ in.
8	35	Format	Medium	Painting
9	36	Format	mime type	JPEG
10	37	Language		Not applicable
11	47	Relation	Is based on <b>(relation of the material with particular subject like painting, drawing, heritage site etc.,)</b>	The present image is based on the Ragamala Painting.
12	42	Relation	Is part of <b>(the material is a part of so and so book)</b>	The present image has been extracted from the book: "Ragamala Painting by Klaus Ebeling, New Delhi, 1973.
13	53	Rights		Kankroli Art Collections
14	63	Subject	Other	Indian Miniature Paintings, Ragamala Paintings, Music, Ragas and Raginis
15	64	Title		Kakubha Ragini
16	65	Type		Photograph

#### 4.6.2 User Interface

DSpace works on Web User Interface, which consists of several Interfaces:

- User Interface for submission
- User Interface for submission Process
- End-users browsing Information

It supports search and retrieval of item by browsing or searching the metadata. After locating the item in the system, the retrieval is accomplished by clicking a link that causes the archived material to be downloaded to the user's web browser. It is highly scalable to customize the user interface according to the requirement and the home page can be tailor made to have a look and feel of the project it adheres to[6]. Following is an example of the screenshot of the case study done



- User Interface for System Administration is used for creating communities, additional non-qualified Dublin core metadata elements, users' creation, groups and community creation etc.

Collection can be created after creating a community. Collection can appear under a community or sub community. They hold digital items, which are digital objects.

For creating a collection click on 'create collection' and by default the first two options i.e., 'New items should be publicly readable' and 'some users will be able to submit to this collection' are to be ticked.

The fields for creation of collection are similar to the fields for creation of community. In the present study, the following collections have been created on experimental basis. All Collection holding are currently under the main community Cultural Heritage materials.

- Drawings
- Manuscripts
- Paintings
- Heritage sites
- Sculptures

An example of a Collection about 'Manuscripts' which is created in DSpace with description is given below :

Name	: Manuscripts
Short Description	: Rare and Invaluable collection of Manuscripts.
Introductory Text	: Indian manuscripts are the basic materials for Research for different fields like Mathematics, Medicine, Astronomy, Jyothisya, Religion and Art.
Copyright Text	: Any part of this collection cannot be used for any kind of purposes other than Education and Research purpose.
Side bar text	: Paintings of 17 <sup>th</sup> , 18 <sup>th</sup> & 19 <sup>th</sup> centuries
License	: Museologists, Archivists, Members of Archeological Society of India and Librarians (as defined in e-groups or e-people).
Provenance	: From Any Region of India
Collection's Authorization	: National Museum of India, National Archives of India and Archeological Survey of India.



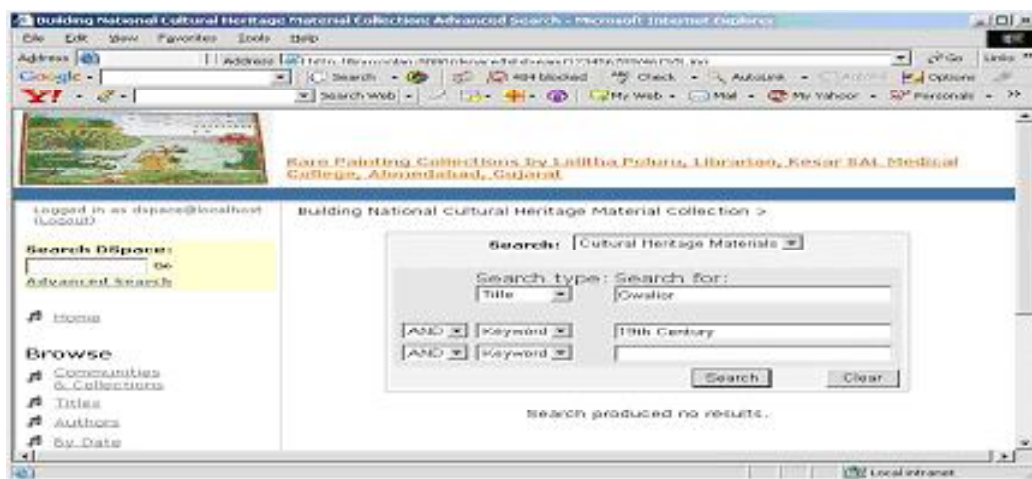
## Search Process

Any Image Repository System should have a powerful search process through which the images can be searched and displayed instantaneously. DSpace uses Jakarta Search engine Lucene, which is simple and performs high. It also gives the capabilities of fielded searching, stop word removal, stemming and adding of newly indexed content without regenerating the entire index.

Browsing in DSpace allows going through a list of items in some specified order. It can be by:

- Community/Collection - Title - Author -Date

Search result with Boolean Operators in DSpace is given below :





DSpace Information Model for Image Repository of Indian Cultural and Heritage Materials

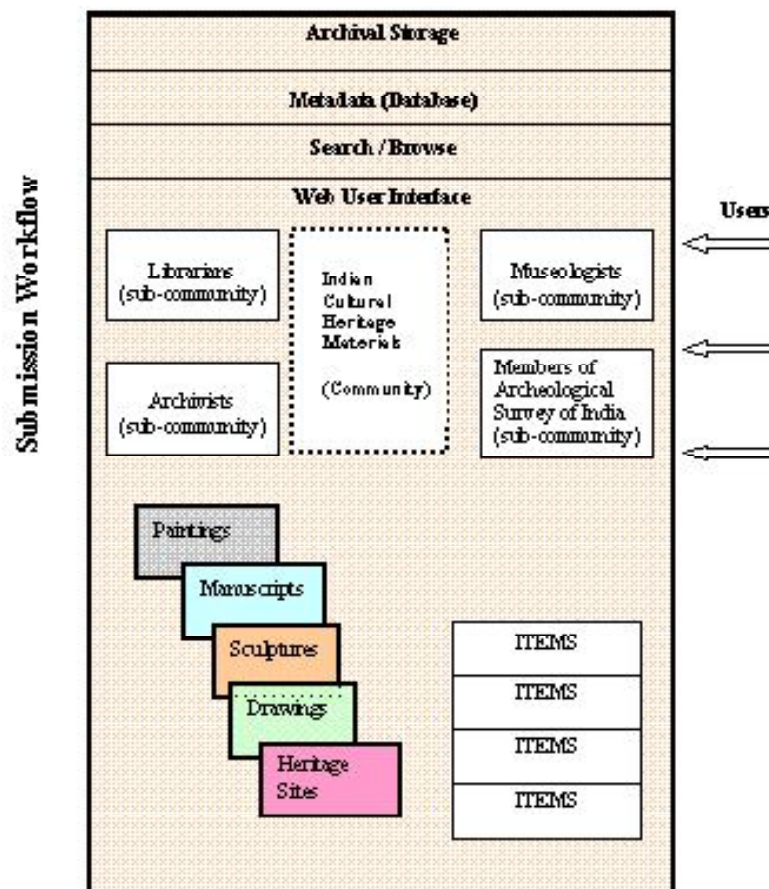


Fig. No. 1



### 4.6.3 Workflow

DSpace is the first open source digital repository which handles multidisciplinary system. Basically it follows a hierarchical structure. In other words it consists of Communities representing different groups. In the present study it can be referred as Indian Cultural Heritage Materials as a *Top Community* under which The Librarians, museologists, Archivists and Members of Archeological Society of India are *Sub-Communities*. Under Community or Sub-community there may be different collections of photographs of paintings, sculptures, manuscripts; Heritage sites etc., In each collection again there would be so many items. So who should submit these items? What type of items can be deposited? Who needs to edit or approve the submission? To what collection a particular item can be deposited? Who can see these items after deposition? The community representatives who work with the repository address all of these issues. The 'e-people' in the system have 'roles' in the workflow of particular community with particular collection. They are also members of 'groups' who can make administrator's functioning easier when manipulating authorization policies.

## 5. CONCLUSION

The digital technology holds the promise of a drastic change in different fields like education, culture and research. Our cultural, scientific and information heritage is increasingly converting to the digital forms. Modern digital technologies have made it a reality to exhibit large collections of work from multiple cultures. Most of the cultural and heritage materials are being converted into the digitized forms knowing that permanent access to these heritages will offer broadened opportunities for creation, communication and sharing of knowledge among all people, as well as protection of rights and entitlements and support of accountability for the better understanding of the tradition, culture and history of the nation. It is concluded that there is an urgent need to create a National Repository for Indian Cultural Heritage materials that would provide authenticity and ownership and will enable preservation of intellectual property rights.

## 6. REFERENCES

1. Boeri Robert J; Hensel Martin. "Here's Waldo: Content-Based Image Retrieval". At <http://www.std.com/~bboeri/pdfs/emedial/i9802.pdf> Accessed on 06.08.03.
2. Chen Ching-chih, Wactlar Howard D, Wang James Z, Kiernan Kevin. "Digital Imagery for Significant Cultural and Historical Materials- An Emerging Research Field Bridging People, and Technologies". At <http://www.db.stanford.edu/wangz/project/insearch/ART/JODL04/imaging.pdf> Accessed on 21.01.05.13.06.03.
3. Lyman Peter; Kahle Brewster. (1998). "Archiving Digital Cultural Artifacts: Organizing an Agenda for Action", D-Lib Magazine. At <http://www.dlib.org/dlibjuly98/07lyman.html> Accessed on 28.03.03.
4. Lalitha P, Murthy TAV. (2005.) "Preservation of Digital Cultural Heritage Materials". In *Proceedings of 3rd International CALIBER- 2005*, 2-4 February 2005. Ahmedabad:INFLIBNET, p.420-427.
5. Lalitha P; Muthy TAV (2004). "Search Query Process of Image Database Retrieval Systems: A Review" In, *University and College Librarianship in India in the 21<sup>st</sup> Century*. Edited by Neela J Deshpande and S.K.Patil, P-208-214.
6. Prasad ARD. (2005). International Workshop on Building Digital Libraries Using DSpace, 7-11<sup>th</sup> March, 2005, Organized by DRTC, Bangalore and HP Labs, India.
7. Smeulders AWM, Worring M, Santini S Gupta A, Jain R. (2000). "Content-Based Image Retrieval: The end of the Early Years". *IEEE Trans. On Pattern Analysis and Machine Intelligence* 22(12), p-p.1349-1380.

8. Stanchev Peter L; Green Jr. David. (2002). Current State and Research Trend in the Image Data based Systems", Mathematics and education in Mathematics, Borovez,2002, p.67-76. At <http://www.gmi.edu/~pstnache/borovez.pdf> Accessed on 21.02.03.
9. <http://www.dspace.org>.

### About Authors



**Smt. P Lalitha** is working as Librarian at Kesar SAL Medical College, Bhadaj, Ahmedabad. She holds B.SC., MLISC., and pursuing Ph.D. in Library and Information Science from University of Bundelkhand, Jhansi. She is also involved in teaching LIS Students of IGNOU at Regional Center of Ahmedabad. She has published 4 papers at National and International level. Her areas of interest are Digitization of Indian Heritage and Cultural Materials and Repositories.



**Dr T AV Murthy** is Director of INFLIBNET Centre Ahmedabad. He has over 30 years of professional experience both in the field of Library & Information Science including Computer Applications in India & USA. He has held high prestigious positions in reputed Government and Academic institutions across the Country. Published over 110 technical papers and produced over 8 doctoral students. His areas of specialization include: Multi-media, Networking, E-learning and Information Systems. Visited over 22 countries for professional pursuits and was one of the architects of the group of INFLIBNET right from its conceptualization stage. Member on several Government of India bodies/committees. He has proven record of administering large Libraries and Information Systems and organized several national and international conferences.

**E-mail** : [tav@inflibnet.ac.in](mailto:tav@inflibnet.ac.in)

**Prof. MTM Khan** is Professor and Head of the Department of Library and Information Science, University of Bundelkhand, Jhansi, Uttar Pradesh. He has over 31 years of Administrative and Professional experience. He has published 40 technical papers, 3 books and produced over 25 doctoral students.



**Manoj Kumar K.** is currently working as Scientist-D in INFLIBNET Centre at Ahmedabad since 2004. He has more than 10 Years of experience in Information and Technology. It includes 5 Years of service with Indian Institute of Management, Kozikode (IIMK). Prior to joining IIM, he was working as Asst Engineer in CEDTI, Calicut and also as Systems Officer in Coal Indian Ltd, Ranchi, Bihar. Presently he is entrusted with the task of modernization of college libraries and networking in Jammu & Kashmir and North Eastern Region under the aegis of Prime Ministers' Office. He has contributed number of papers in seminars and conferences and also written articles in popular Malayalam Dailies.

**E-mail** : [manoj@inflibnet.ac.in](mailto:manoj@inflibnet.ac.in)