

Institutional Digital Repositories/e-Archives: INFLIBNET initiatives in India

Yatrik Patel

Scientist-B (Software R&D Group)

J K Vijayakumar

Scientific Technical Officer-I (Database Management R&D Group)

Dr T A V Murthy

Director

Information and Library Network (**INFLIBNET**), An IUC of University Grants Commission,
PB 4116, Ahmedabad-38 009, INDIA

E-Mail: yatrik, vijay, tav@inflibnet.ac.in URL: <http://dspace.inflibnet.ac.in>

Abstract

The technological advances today make it possible to think in terms of storing all the knowledge of the human race in digital form and several organizations worldwide are experimenting with less-expensive ways to create Institutional Repositories. For long-term preservation of our knowledge base and cultures, we have to find out an economical way to save digital content for future generations. **INFLIBNET** (*Information and Library Network Centre*) decided to opt DSpace for its Institutional Repository and archive its publications, conference proceedings, lecture notes etc. DSpace, jointly designed by MIT and HP, is a groundbreaking digital library system that captures, stores, indexes, preserves and redistributes the intellectual output of a university's research faculty in digital formats. This paper narrates the practical experiences and provide an overview of **INFLIBNET's** institutional repository and **dArchive-India** developed for Indian academic and research community to archive their intellectual work.

Keywords: Institutional Repositories, Digital Repositories, E-Archives, DSpace, **INFLIBNET**, dArchive-India

0. Introduction

Digital technology can make all the significant literary, artistic, and scientific works of mankind permanently digitally preserved and made accessible to the billions of people all over the world. The technological advances today make it possible to think in terms of storing all the knowledge of the human race in digital form and several organizations worldwide are experimenting with less-expensive ways to archive and disseminate scholarly information and in-house knowledge as Institutional Repositories. For example, the California Digital Library's e-Scholarship program at the University of California, ArXiv.org at Cornell University etc support open-access distribution of scholarship in cross-disciplinary through subject-based approach. U.S. Library of Congress' US \$100 million National Digital Information Infrastructure and Preservation Program (NDIIPP), which is developing a standard way for institutions to preserve their digital archives. Librarians and computer scientists are working together at MIT for Dspace digital asset management system, which aims to create an institutional repository that will include digitized versions of lecture notes, videos, papers, and data sets—in short, everything produced by faculty and staff.

1. Institutional Repository:

An Institutional repository is a digital archive of the intellectual output of an Institution. It makes the quality and breath of scholarship produced at that Institution accessible to others world wide over the Internet. It is a set of services that a University/Organization offers to the members of its community for the management and dissemination of digital material created by the institution and its community members. It is most essentially an organizational commitment to the

stewardship of the digital materials including long term preservation. An effective Institutional repository of necessity represents collaboration among librarians, Information technologies, archives and record managers, faculty and University administrators and policy makers. It is a new channel for structuring the organisation's contribution to the border world and as such invites policy and cultural reassessment of this relationship.

It also form part of a larger global system of repositories which indexed in standardizes way and searchable using one interface, providing the foundation for a new model of scholarly publishing. Institutional repositories often employ a method of article submission known as author self-archiving and rely on user communities to control the input of content. These can include academic departments; research centers administrative groups and other sub-group who will determine what content merits inclusion and act as arbiters for their own research communities. Institutional repositories also have roles beyond disseminating and managing the works of individuals scholars that are part of the dialog of scholarly communication and it is also a place where we can put much of the material that research libraries identify as worth collecting and they offer a framework for organized stewardship and accessibility of these materials. It can encourage the exploration and adoption of new forms of scholarly communication that exploit the digital medium in fundamental way. Benefits are:

- Improved visibility and impact-Institutional, National;
- Contribute to institutional knowledge management;
- Improve Research collaboration-interdependent, inter-Institutional, International;
- Enhanced status and reputation;
- Enhanced research capacity.

2. Initiatives by INFLIBNET

INFLIBNET Information and Library Network is an Autonomous IUC of the University Grants Commission (UGC) started in 1991 with its Headquarters located at Gujarat University campus, Ahmedabad. The Programme is directed towards modernization of libraries and Information centres, and establishment of a mechanism for information transfer and access, to support scholarship, learning and academic pursuits. It also aimed at establishing a National Network of Libraries and Information Centres in Universities, institutions of higher learning and R & D Institution in India. It is basically a co-operative endeavor in resource development, sharing and its utilization at national level. The main objective being to evolve a National Network, interconnecting various Libraries and Information Centers in Universities, deemed universities, colleges, UGC information centers, Institutions of National importance and R & D institutions.

In India also leading institutions started creating of Digital Libraries in different disciplines by utilizing open sources and provide open access to their scholarly literature. For long-term preservation of our knowledge base and cultures, we have to find out an economical way to save digital content for future generations. Institutions in India started using variety of open source archiving solutions like Green Stone, E-Print, DSpace etc since 2000. After experimenting all popular solutions, **INFLIBNET** decided to opt DSpace for its Institutional Repository and archive its publications, conference proceedings, lecture notes etc. The capability of Dspace to handle multilingual content, even at Metadata level using globally accepted UNICODE standard was the important issue for selecting this solution, especially a country like India with its mullti-lingual dilemma in digital content creation and storage.

1. About D-Space

D-Space is a groundbreaking digital library system that captures, stores, indexes, preserves and redistributes the intellectual output of a university's research faculty in digital formats.[<http://dspace.org/>]. In March 2000, Hewlett-Packard Company (HP) awarded \$1.8 million to the MIT Libraries for an 18-month collaboration to build DSpace™, a dynamic repository for the intellectual output in digital formats of multi-disciplinary research organizations. HP Labs and MIT Libraries released the system worldwide on November 4, 2002, under the terms of the BSD open source license, one month after its introduction as a new service of the MIT Libraries. As an open source system, DSpace is now freely available to other institutions to run as-is, or to modify and extend as they require to meet local needs. From the outset, HP and MIT designed the system to be run by institutions other than MIT, and to support federation among its adopters, in both the technical and the social sense. The DSpace Federation will be explored in a later section [<http://www.dlib.org/dlib/january03/smith/01smith.html>].

2. DSpace@INFLIBNET

At **INFLIBNET** we have downloaded Dspace from <http://dspace.org/>. It has been installed in test-bed and experimented its capabilities and performance. After all sorts of testing required, Dspace was customized according to our requirements and it was installed on one of the WWW Server on Linux (RedHat 9) platform. Then we have requested Corporation for National Research Initiatives (CNRI) site for providing *Persistent Identifiers (CNRI Handles)* which promotes interoperability among open archives through Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). The Handle System® covers assignment, management, and resolution of these persistent identifiers and are compliant with the IETF's Uniform Resource Name (URN) specification. The customized Dspace Archive is accessible through Web at <http://dspace.inflibnet.ac.in> (Fig. 1)

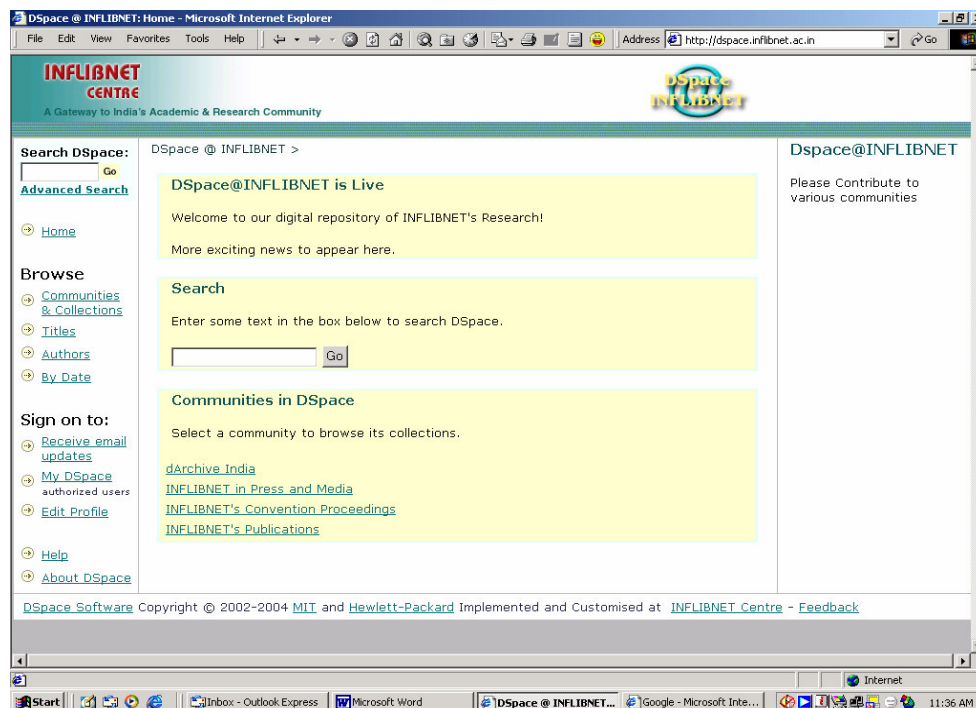
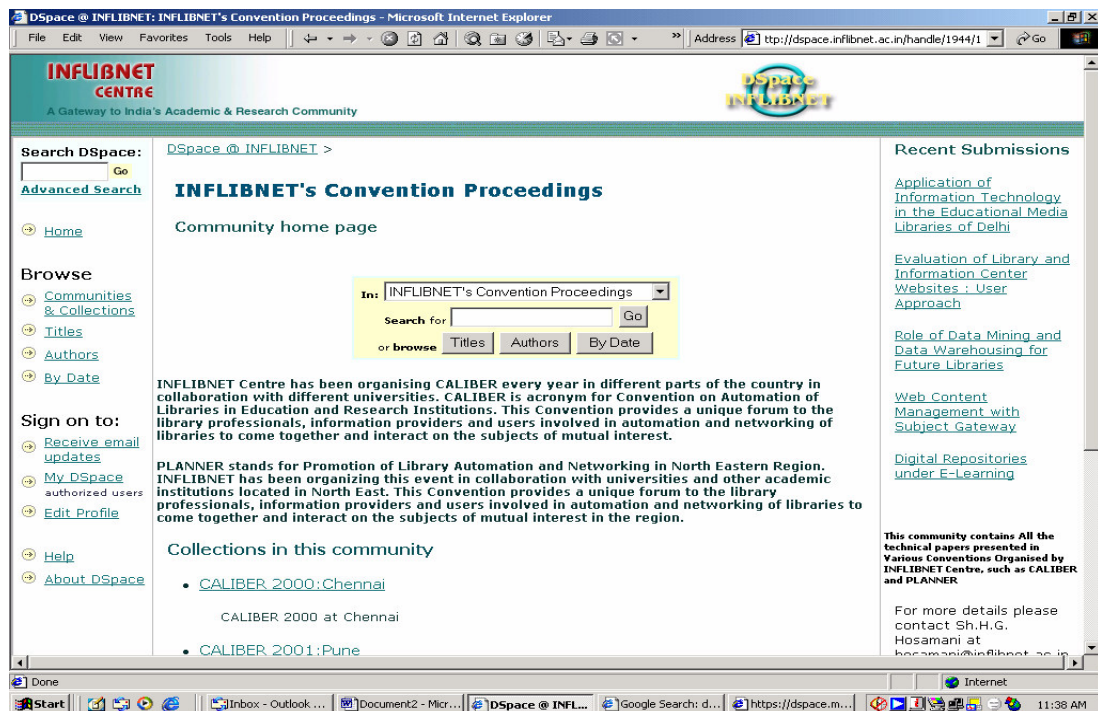


Fig. 1: Snapshot of **DSpace@INFLIBNET**

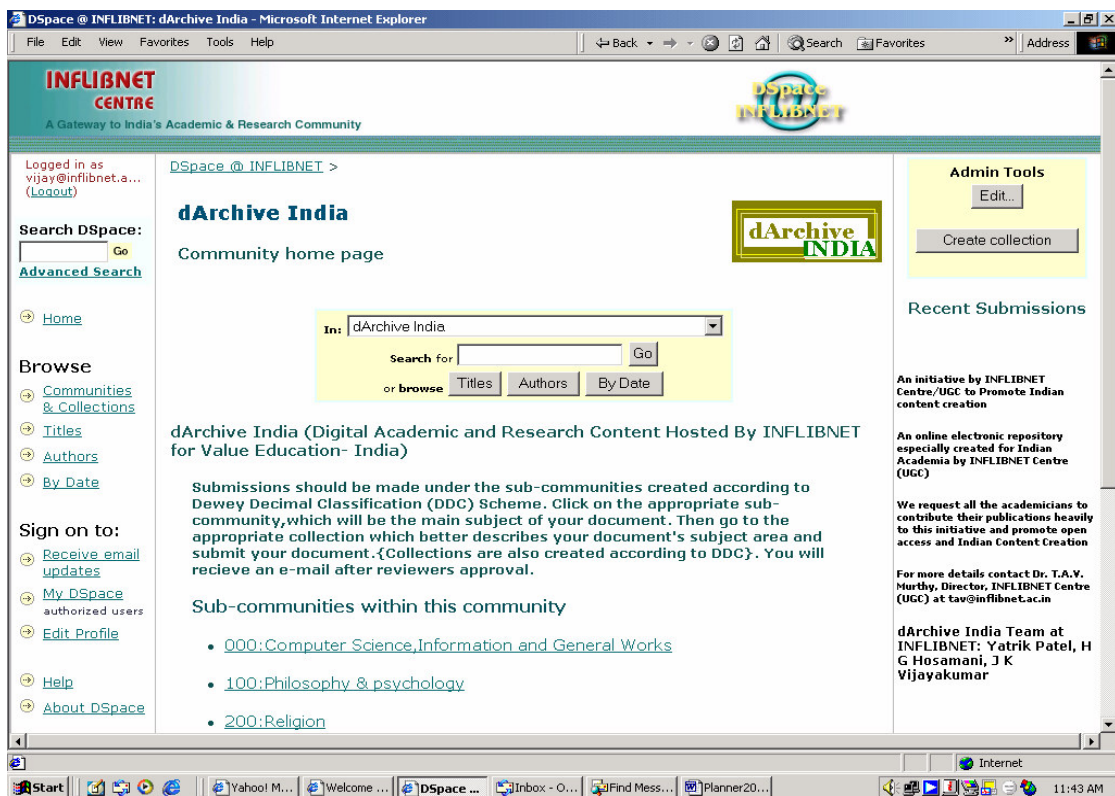
2.1 Communities in Dspace@INFLIBNET

Each DSpace site is divided into *communities*; these typically correspond to a department or laboratory or research center of an organisation, which can be organized into a hierarchy. “Communities” of DSpace contain *collections*, which are groupings of related content. At **INFLIBNET** we have extended this definition and categorized communities according to types of collections or area of interest. These are further divided into Sub- communities/Collections collections. The following table will provide you the details.

Comunities	Description
INFLIBNET's Convention Proceedings	INFLIBNET Centre has been organizing two conventions called CALIBER and PLANNER every year in different parts of the country in collaboration with different universities. These Conventions provide a unique forum to the library professionals, information providers and users involved in automation and networking of libraries to come together and interact on the subjects of mutual interest. The convention proceedings published in print form is digitized and archived under this community, under each collection separately named as different conventions.
INFLIBNET in Press and Media	Coverage of INFLIBNET activities appeared in different media is being collected and archived here
INFLIBNET 's Publications	INFLIBNET has been bringing out the various publications on regular basis, like Newsletter, various course materials etc. Lecture notes, presentations etc made by INFLIBNET staff will also be available here.
dArchive-India (Digital Academic and Research Content Hosted By INFLIBNET . for Value Education- India) Promotion of Indian content creation and archiving of Indian scholarly out put are the basic objectives of this initiative. Any faculty member or researcher in India can made submissions under the ten sub-communities created according to Main Classes at Dewey Decimal Classification (DDC) Scheme.	Sub-communities are created according to Dewey Decimal Classification (DDC) Scheme as follows; 000:Computer Science,Information and General Works 100:Philosophy & psychology 200:Religion 300:Social Sciences 400:Language 500:Science 600:Technology 700:Arts & Recreation 800:Literature 900:History & Geography



Screenshot of **INFLIBNET** Convention Proceedings Community



Screenshot of **dArchive-India** Community

3. Creation of Metadata

DSpace holds three types of metadata about archived content:

3.1 Descriptive Metadata

Each *Item* has one qualified Dublin Core metadata record. The set of elements and qualifiers used by MIT Libraries is the default configuration included in the DSpace source code. These are loosely based on the Library Application Profile set of elements and qualifiers, though there are some differences. Other descriptive metadata about items may be held in serialized bitstreams. *Communities* and *collections* have some simple descriptive metadata (a name, and some descriptive prose), held in the DBMS.

3.2 Administrative Metadata

This includes preservation metadata, provenance and authorization policy data. Most of this is held within DSpace's relation DBMS schema. Provenance metadata (prose) is stored in Dublin Core records. Additionally, some other administrative metadata (for example, bitstream byte sizes and MIME types) is replicated in Dublin Core records so that it is easily accessible outside of DSpace.

3.3 Structural Metadata

This includes information about how to present an item, or bitstreams within an item, to an end-user, and the relationships between constituent parts of the item. In addition to some basic technical metadata, bitstreams also have a 'sequence ID' that uniquely identifies it within an item. Additional structural metadata can be stored in serialized bitstreams, but DSpace does not currently understand this natively.

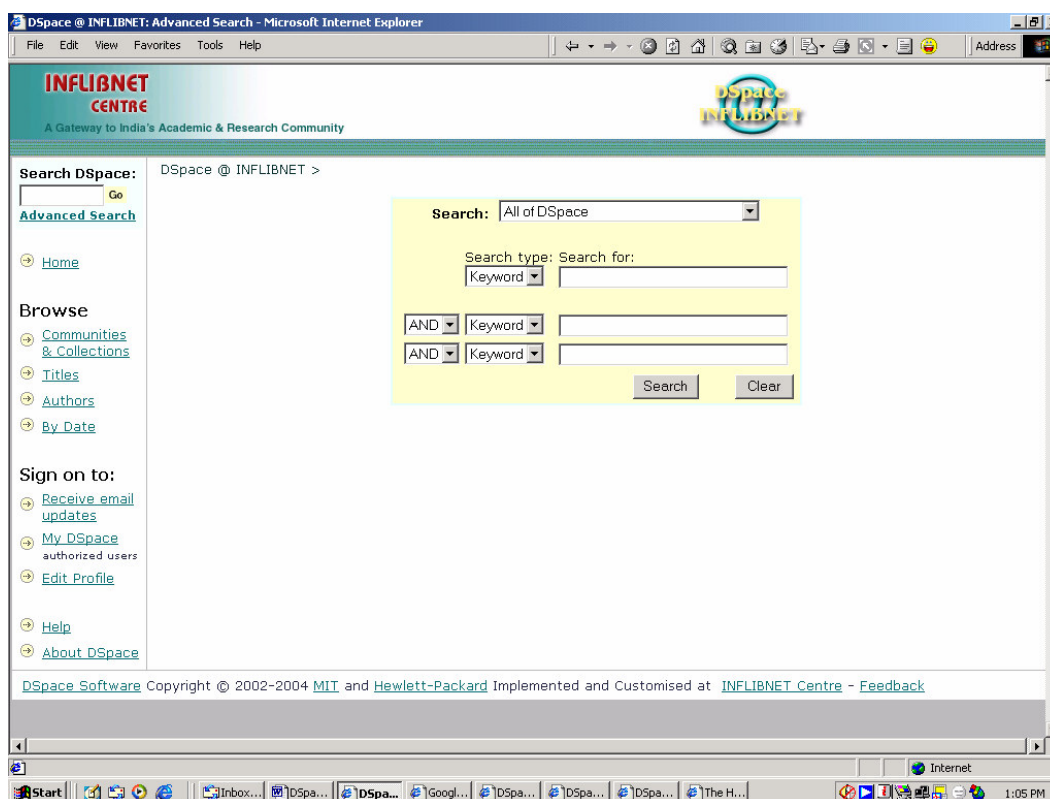
4. Search and Browse in DSpace

DSpace allows end-users to discover content in a number of ways, including:

- Via external reference, such as a Handle
- Searching for one or more keywords
- Browsing through title, date and author indices

Search is an essential component of discovery in DSpace. Users' expectations from a search engine are quite high, so a goal for DSpace is to supply as many search features as possible. DSpace's indexing and search module has a very simple API which allows for indexing new content, regenerating the index, and performing searches on the entire corpus, a community, or collection. Behind the API is the Java freeware search engine [Lucene](#). Lucene gives us fielded searching, stop word removal, stemming, and the ability to incrementally add new indexed content without regenerating the entire index.

Another important mechanism for discovery in DSpace is the browse. This is the process whereby the user views a particular index, such as the title index, and navigates around it in search of interesting items. The browse subsystem provides a simple API for achieving this by allowing a caller to specify an index, and a subsection of that index. The browse subsystem then discloses the portion of the index of interest. Indices that may be browsed are item title, item issue date and authors. Additionally, the browse can be limited to items within a particular collection or community.



Snapshot of Advance Search Option

5. Conclusion

Higher-education leaders in India invariably have long lists of difficult issues to confront and they should consider several strategies to comply with the latest technological developments and availability of free solutions for a better and effective implementation of ICT. They can develop and support new models of scholarly publishing that cut the costs of distributing and retrieving information. If universities continue to operate as we do now, our library collections will grow -- but their scope and depth will diminish precipitously. **INFLIBNET** believes that it is our responsibility to lead the charge for a realistic assessment of how we can head off an otherwise inevitable loss of academic resources. Awareness building in Open Archives and Institutional Repositories in Indian academic Institutions will be the main focus of **INFLIBNET** in coming future.

6. References

1. Yatrik Patel, J K Vijayakumar and T A V Murthy (2004). DSpace@INFLIBNET: an Indian initiative to develop Digital Repositories/e-Archives (in press)
2. MacKenzie Smith and others (2003). DSpace: An Open Source Dynamic Digital Repository, D-Lib Magazine, 9(1), Available at <http://www.dlib.org/dlib/january03/smith/01smith.html>
3. <http://dspace.inflibnet.ac.in>
4. <http://www.dlib.org>
5. <http://dspace.library.cornell.edu>
6. <http://dspace.mit.edu>

7. <http://www.dspace.org>

Published in In M G Sreekumar [et al.]; ed. **Digital Libraries in Knowledge Management: Proceedings of the 7th MANLIBNET Annual National Convention**, Calicut, India, May 5-7, 2005: New Delhi, Ess Ess, 2006. Pp. 312-318. ISBN: 81-7000-442-X
© ESS ESS Publications, New Delhi