

# **The Study on the Open Source Digital Library Software for Institutional Repository in India: A Survey**

**Goutam Biswas**

Librarian, Nagar College, Nagar, Murshidabad, West Bengal, India  
[goutambiswas8@gmail.com](mailto:goutambiswas8@gmail.com)

## **ABSTRACT**

*As Dspace and Greenstone embark on their second decade, this article offers a retrospective analysis of their development, highlighting key challenges encountered and lessons learned in building and implementing a comprehensive open-source system for constructing digital libraries on a global scale. The DSpace and Greenstone Digital Library Software have played crucial roles in promoting the practical application of digital library technology globally, with a particular focus on developing countries. Open-source digital library packages have gained significant recognition and usage in recent years. For cost-effective development of digital libraries, using open-source software is highly recommended. This paper presents the findings of an online survey examining the adoption and usability of open-source digital library software across various organizations. The survey aimed to assess the extent to which organizations have implemented these open-source solutions and to compare the features of the most popular OSS digital library software packages.*

**KEYWORDS:** Open Source Software, Digital Library, Greenstone Digital Library, and Dspace.

## **1. INTRODUCTION**

Open-source software (OSS) provides freely accessible source code that anyone can view and modify. Various licensing models fall under the OSS label, but the core principle is that the software's "license may not restrict any party from selling or giving away the software as a component of an aggregate software distribution." Additionally, the working software must either be distributed with its source code or have a well-publicized method for downloading the source code at no cost. This allows users to contribute to the software's ongoing development, fix bugs, and tinker with the code as they please. This contrasts with proprietary software, where the source code remains solely under the control of the software vendor and is not publicly accessible. In response to restrictive software copyright laws, the software developer community-created open-source software, which is available at no cost and allows users the freedom to run and distribute the programs without limitation. This has provided a cost-effective solution for smaller and medium-sized libraries, which often view the automation of administrative tasks as a financial burden due to the high prices of Integrated Library Systems (ILS). The development of open-source software offers an affordable way for these libraries to digitize their operations. [1]

## **2. OBJECTIVE OF THE STUDY**

The objective of the present study is to look into the technologies and tools available in the open-source world that can be used to improve the services within the libraries.

## **3. OPEN SOURCE SOFTWARE MOVEMENT**

The open-source software movement emerged alongside the early development of computers and software, as programmers and developers commonly shared their creations freely during this period. The rise of profit-driven software companies hindered the culture of openly sharing software source code. Key milestones in the history of open-source software include:

- 1983 - Richard Stallman formed the GNU project.
- 1985 – Creation of Free Software Foundation.
- 1991 – Development of Linux kernel by Linus Torvalds.
- 1998 – Open Source Initiative (OSI) formed by Eric Raymond.

The terms "free" and "open source" are often used interchangeably to refer to the free distribution of software. Some popular licenses used for this purpose include the GNU General Public License (GPL), BSD license, GNU Lesser General Public License, MIT License, Mozilla Public License, and Apache License. While these licenses have some differences in their specific terms and conditions, they all share the fundamental goal of ensuring users' freedom to use, copy, distribute, and improve the software. This philosophy aligns with the principles of the Free Software Foundation, which defines "free software" as software that gives users the freedom to run, copy, distribute, study, change, and improve it. [2]

## **4. REASONS FOR CHOOSING OPEN SOURCE SOFTWARE**

Price of the library software is very high; most of the libraries are not in a position to buy high-priced commercial software due to severe budget constraints. Biswas, Goutam and Paul, Dibyendu [3] have mentioned various advantages of open source like unrestricted use; free of cost; community involvement in the development and maintenance of software; competence compared to other commercial software; and the issues of copyright, etc. The primary reason organizations like libraries choose open-source software for automation is the lack of licensing costs. Open source software can be freely used, studied, modified, and distributed, regardless of the user's position, wealth, or social status. This social aspect is a significant benefit, as the development and maintenance of open-source software is often community-driven, allowing anyone to contribute. Open-source projects encourage innovation and collaboration among community members, who are actively involved in the process. Another key advantage is the interoperability and customizability of open-source software to meet specific needs and standards. Finally, the legal aspect is favorable, as open-source licenses are committed to preserving users' freedom to use, modify, and redistribute the software.

## **5. DEFINITION**

A digital library is a collection of information stored in digital formats, such as electronic documents or multimedia that can be accessed via computers. These digital materials may be stored locally or accessed remotely through

## ***The Study on the Open Source Digital Library Software for Institutional Repository in India: A Survey***

computer networks. Digital libraries are a type of information retrieval system. The term "digital library" was first used in print in a 1988 report to the Corporation for National Research Initiatives, and was later popularized by the NSF/DARPA/NASA Digital Libraries Initiative in 1994.

Though electronic library these days more often refers to portals, frequently offered by government entities, as in the case of the Florida Electronic Library, the older terms electronic library or virtual library are still infrequently used.

A digital library is described as follows by the DELOS Digital Library Reference Model:

An organization, which might be virtual, that comprehensively collects, manages and preserves for the long-term rich digital content, and offers to its user communities specialized functionality on that content, of measurable quality and according to codified policies

### **5.1 Advantage of Digital Library:**

Though electronic library these days more often refers to portals, frequently offered by government entities, as in the case of the Florida Electronic Library, the older terms electronic library or virtual library is still infrequently used. A digital library is described as follows by the DELOS Digital Library Reference Model:

- Preserve the valuable documents, rare and special collections of libraries, archives and museums.
- Provide faster access to the holding of libraries worldwide through automated catalogues.
- Help to locate both physical and digitized versions of scholarly articles and books through a single interface.
- Search optimization, simultaneous searches of the Internet makes possible, to prepare commercial databases and library collections.
- Offering an online learning environment.
- Making short the chain from author to user.
- Save preparation/ conservation costs, space and money.
- Digital technology affords multiple, simultaneous user from a single original which is not possible for materials stored in any other forms

## **6. DSPACE DIGITAL LIBRARY SYSTEM**

The Dspace digital asset management system is a joint project of the MIT Libraries and HP Labs. Dspace helps create, index, and retrieve various forms of digital content and is adaptable to different community needs. Interoperability between systems is built-in, adhering to international standards for metadata format. Dspace is an open-source technology platform that can be customized or extended to suit users' specific requirements.

✓ Dspace is a service model for open access and/or digital archiving for perpetual access.

✓ Dspace is a platform for building an Institutional The repository and the collections are searchable and retrievable by the Web.

✓ To make available institution-based scholarly material in digital formats. The collections will be open and interoperable.

An institutional repository is a suite of services offered by a research institution, organization, or university to its community members. These services facilitate the management and dissemination of digital materials created by the institution and its community.

### **Major Features of Dspace**

The following sections describe the three major features of DSpace.

- 1) Lucene Search Engine and query language
- 2) Handle System
- 3) OAI-PMH

#### **6.1.1 Lucene Search Engine**

Dspace uses Lucene Search Engine, which is a part of the Apache Jakarta Project (6). The syntax of the queries is given below.

#### **6.1.2 Exact Term**

The search term can be a word or a phrase. One can use a search word, e.g. "Information" or a phrase "information retrieval".

#### **6.1.3 Fielded Search:**

One can search for a term in a particular field. e.g.: author: jaba title: web keyword: ocr  
Abstract: Digital

#### **6.1.4 Wild cards:**

The symbol '?' is used for a single character, as in 'te?t' that matches words like 'test', 'text' etc. The symbol '\*' is used for multiple characters matching, as in "inf\*" matches with information, informatics, etc.

#### **6.1.5 Fuzzy Search**

One of the popular fuzzy search algorithms is the Levenshtein distance algorithm named after the Russian scientist Vladimir Levenshtein, who devised the algorithm in 1965. It is also called 'Edit Distance algorithm' Levenshtein Distance (LD) is a measure of the similarity between two strings, which we will refer to as the source string (s) and the target string (t). The distance is the number of deletions, insertions, or substitutions required to transform s into t. For example, • If s is "test" and t is "test", then  $LD(s,t) = 0$ , because no transformations are needed. The strings are already identical. • If s is "test" and t is "tent", then  $LD(s,t) = 1$ , because one substitution (change "s" to "n") is sufficient to transform s into t. The Levenshtein distance algorithm has been used in:

- Spell checking
- Speech recognition
- DNA analysis
- Plagiarism detection

In Dspace implementation, one can use in the following way: Example: author: sanker~

Can match Shankar. It can notice, the search word has 'sa' not 'sha' and also 'ker' not 'kar'.

#### **6.1.6 Proximity Search**

## ***The Study on the Open Source Digital Library Software for Institutional Repository in India: A Survey***

In a query, proximity search is used to find documents that contain two words or phrases that are close to one another. "Information system"~3 retrieves data in which the terms "information" and "system" are separated by no more than three words. So, the following titles are returned by the search above.

### **6.1.7 Range search**

When typing the search term, "author:[prasad to rao]," Documents authored by names that fall between "prasad" and "rao" are then retrieved by the system. In contrast, Prasad and Rao are not included in the query "author:{prasad to rao}."

### **6.1.8 Boosting a Term**

Based on the terms detected, Lucene gives the matched document's relevance level. Use the caret, "^", symbol at the end of the term someone is searching for, together with a boost factor (a number), to boost the term. The word will be more significant the higher the boost factor. By increasing a document's term, you can manage its relevance by boosting it. For instance, to make the term "internet" more relevant when someone searches for it online, use the ^ sign in conjunction with the boost factor adjacent to the term. You would write: By default, the Internet~5 web's boost factor is 1. The boost factor can be less than 1 (e.g. 0.2), even though it must be positive.

### **6.1.9 Boolean Search**

Boolean 'AND', 'OR', 'NOT' are used for Boolean combinations. Boolean operators should be caps.

- 'OR' is the default conjunction operator. One can use '|' instead of 'OR'.
- Either 'AND' or '&&' can be used for Boolean 'AND'.
- Either 'NOT' or '!' can be used for Boolean 'NOT'.

## **6.2 Handle System**

Dspace uses the global resolution function of the Handle system; to utilize it, you must first set up a Handle server, which may be found at the official CNRI Handle site. When a handle entered locally does not match any local material, the handle server runs as a separate process that accepts TCP requests from other handle servers and sends resolution requests to a global server or servers. Since the Handle protocol is based on TCP, a server capable of broadcasting and receiving TCP on port 2641 must host it. Keep in mind that the Dspace Handle server does not provide administrative actions like Handle creation and maintenance because the Dspace code maintains individual Handles.

## **6.3. OAI-PMH**

The de facto standard for metadata harvesting is now the Open Archives Initiative-Protocol for Metadata Harvesting. As a result, digital library service providers can gather metadata, index it, and offer improved search results. The purpose of this section is to demonstrate OAI-PMH verbs that service providers may typically utilize. The XML output is displayed by the OAI verb examples. The following verbs can be used to gain a better understanding of the harvesting process, but they are not intended for the end user.

To view the output of Dspace, try a few of the OAI-PMH verbs listed below. Determine: provides broad details about the archive and its rules (such as dates and timestamp granularity). As an illustration, see <http://drtc.isibang.ac.in/oai/?verb=Identify> List Sets: Provide a list of possible sets for organizing records (they can be flat, overlapping, or hierarchical). This is an example: [oai/drtc.isibang.ac.in/?verb=ListSets](http://drtc.isibang.ac.in/oai/?verb=ListSets).

## **7. GREENSTONE DIGITAL LIBRARY SOFTWARE**

The New Zealand Digital Library project's Greenstone Digital Library Software offers a novel approach to knowledge organization and online accessibility. Information collections consist of many documents (usually between a few thousand and a few million), and they are presented with a standard interface. Libraries contain a wide variety of collections, each one arranged differently but having a lot in common. A collection's structure is set by a configuration file. There are already collections of everything from technical papers to newspaper articles, oral histories to educational publications, movies to visual art, and ethnic folk tunes to MIDI mainstream music compilations. [4]

### **7.1 Special Feature**

- It suits both Windows and Unix (Linux SunOS) any of these systems can be used as a web server.
- Its built-in administration feature allows new users to create collections and secure documents so that only registered users with a password can access them.
- It creates collections with efficient metadata-based browsing and full-text searches. It is possible to build a collection of up to several terabytes that contains millions of documents. Because compression is utilized to minimize the size of the indexes, full-text search is quick, allowing text users to peruse the list of authors, titles, dates, class numbers, and other information.
- It is possible to write plug-ins to support new document types. The compilation may include images, sounds, music, film clips, and more. Moreover, multilingual papers are supported.
- Updates to collections and the addition of new ones can be made at any moment without crashing the system. [6]

There is also so many open source digital library software available like:

## **8. EPRINTS**

The University of Southampton in England is the developer of the free program EPrints. The IISc research community's digital research output is gathered, preserved, and shared by the ePrints@IISc repository. Through an online interface, the Institute community can organize and deposit preprints, post-prints, and other scholarly publications.

## **9. COMPARATIVE STUDY OF FEATURE OF GSDL, DSPACE**

## ***The Study on the Open Source Digital Library Software for Institutional Repository in India: A Survey***

The accompanying table compares the features of both digital library software programs. It is said that neither of the software costs anything. While GSDL supported Dublin Core, Dspace only supported qualified Dublin Core. MS Word, PDF, PPTs, JPEG, and GIF were all supported by Dspace and GSDL, whereas MS Word, PDF, HTML, PostScript, JPEG, and GIF were all supported by GSDL.

**Table no: 1:** Comparative study of Dspace and GSDL

	<b>Dspace</b>	<b>GSDL</b>
<b>Creator</b>	MIT libraries & Hewlett-Packard	University of Waikato
<b>Open Source and Free</b>	Yes	Yes
<b>Operating System</b>	Unices, Linux Windows	Unices, Linux Windows
<b>Language</b>	J2SDK v.1.4	Perl
<b>Database</b>	Postures 7.3	Its own
<b>Resource Identifier</b>	CNRI Handles	No
<b>Dublin Core</b>	Qualified Dublin Core	Dublin Core
<b>METS</b>	Implemented in Version 1.2	No
<b>OAI-PMH V 2.0</b>	Yes	No
<b>Subscription</b>	No	No
<b>Supported File formats</b>	MS-Word,PDF, PPTs, JPEG, GIF.	MS-Word,PDF, HTML, PostScript,JPEG, GIF.

## **10. ANALYSIS OF ONLINE SURVEY**

### **Distribution of Software Packages**

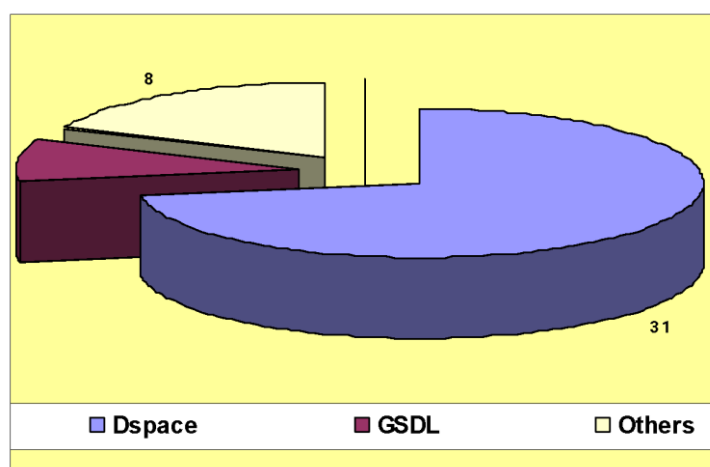
Many subject-oriented, institution-oriented, and mission-oriented digital library projects as well as a wide range of software packages for digital object management were found when searching the Internet for the availability of various digital libraries and software packages for digital libraries. Two of the most widely used digital library software packages have been chosen for the study because it was not possible to conduct a thorough analysis of them all owing to time constraints. Dspace Greenstone [GSDL] software packages have been chosen. Utilizing the Google search engine, an online poll has been carried out to learn more about the user or installation Institute of OSS digital library software.

Another useful tool is the [www.openoar.org](http://www.openoar.org) survey. It is an open-access repository's directory. Open-source digital library software packages have been discovered all over the world, particularly in India at institutions like DRTC, where Inflibnet Dspace is installed and is operational.

According to a web survey, 43 academic institutions across India have implemented the repository software. Additionally, it discovered 4 GSDL installations. Only the institutes that we discovered online have been selected. Out of all the digital library software packages, Dspace appears to be the most popular option with 31 installations, followed by Greenstone with seven installations. Programs such as E-Print are included in the other software package. The two most well-known open-source repository construction programs that we have mentioned here are GSDL and Dspace.

**Table 2:** Distribution of Software Packages under study

Dspace	GSDL	Others	Total
31	4	8	43



**Figure No: 1:** Distribution of Software Packages

## 11. INITIATIVES OF DIGITAL PRESERVATION SYSTEM AND DIGITAL REPOSITORIES IN INDIA

A digital archive of an organization's or institution's intellectual output is known as digital preservation. It allows people all over the world to access the organization's high caliber and extensive scholarship through the Internet. A university or organization provides a range of services to ensure that this treasure of information is preserved through digitization projects and digital preservation initiatives. With the adoption of Open Source Software Systems, diverse multicultural and multilingual contents are now being recorded, preserved, and made accessible through the widely recognized Digital Preservation and Repository initiatives, such as members of its community for the management and distribution of digital material created by the institution and its community members.

In essence, it is an organizational commitment to the long-term preservation and maintenance of digital materials. A collaborative effort between libraries, information technology, archives and record management, academics, university administrators, and policymakers is necessary to create an institutional repository of necessity and effectively preserve digital materials. [7]

With the use of Open Source Software Systems, diverse multilingual and multicultural materials are currently being recorded, maintained, and made accessible through widely recognized Digital Preservation and Repository projects including [10]:



## ***The Study on the Open Source Digital Library Software for Institutional Repository in India: A Survey***

### **National Level Digital Preservations/Repositories:**

- Catalysis Database [www.eprints.iitm.ac.in](http://www.eprints.iitm.ac.in)
- Software Used: EPrints
- Librarians' Digital Library (LDL) <https://drtc.isibang.ac.in/>
- Software Used: DSpace
- OpenMED@NIC <http://openmed.nic.in/>
- Software Used: EPrints

### **Institutional Repositories:**

- Digital Archive of National Institute of Technology Rourkela <http://dspace.nitrkl.ac.in/dspace/>
- Software Used: DSpace
- Electronic Theses and Dissertations of Indian Institute of Science (ETD@IISc) <http://etd.ncsi.iisc.ernet.in>
- Software Used: DSpace
- Open Access Repository of IISc Research Publications (ePrints@IISc) <http://eprints.iisc.ernet.in/>
- Software Used: EPrints
- IDRC Digital Library <http://idl-bnc.idrc.ca/>
- Software Used: DSpace
- Digital Repository of IIT Bombay <http://dspace.library.iitb.ac.in/dspace/>
- Software Used: DSpace
- DSpace at National Centre for Radio Astrophysics <http://ncralib.ncra.tifr.res.in:8080/dspace/>

### **Software Used: Dspace**

- DSpace@IIMK <http://dspace.iimk.ac.in/>
- Software Used: DSpace
- DSpace at National Chemical Laboratory <http://dspace.ncl.res.in/dspace/>
- Software Used: DSpace
- DSpace@INFLIBNET <http://dspace.inflibnet.ac.in/>
- Software Used: DSpace
- University of Delhi EPrint Archive <http://eprints.du.ac.in/>
- Software Used: EPrints
- Raman Research Institute Digital Repository <http://dspace.rii.res.in:8080/dspace/>
- Software Used: DSpace
- One World South Asia Open Archive Initiative <http://open.ekduniya.net/> Software Used: EPrints

### **Digital Library:**

- Archives of Indian Labour: Integrated Labour History Research Programme [www.indialabourarchives.org](http://www.indialabourarchives.org)
- Software Used: Greenstone Digital Library

### **Software**

- India Education Digital Library [www.edudl.gov.in](http://www.edudl.gov.in)
- Software Used: Greenstone Digital Library Software
- Vidyanidhi [www.vidyanidhi.org.in](http://www.vidyanidhi.org.in)
- Software Used: DSpace

## CONCLUSION

Though this does not mean that open source is exclusive to underprivileged areas, it does empower them in general. It is undeniable that Open Source Software (OSS) facilitates the closing of the digital divide in numerous ways. Because they can use OSS, libraries in underdeveloped nations can facilitate resource sharing, digital libraries, and electronic access. Even wealthy nations' libraries are moving closer to Open Source Software (OSS) to enhance their offerings. Dspace is a reasonably potent program. The software's primary benefit is that it enables members to submit digital documents. It will be far more powerful if it has METS (Metadata Encoding and Transmission Standard), which it does not currently have. Out of all the open-source digital library options, Dspace is the most widely used. Eprints and other digital library applications are also extensively utilized. Educational establishments are the primary users of these products. While many universities have adopted digital libraries, not all of them are accessible online. Knowledge cannot be freely accessed unless these repositories are made available online. India is gaining a lot from the open-source initiative.

## REFERENCE

- [1] Lee, Cal (2009) Open-Source Software: A Promising Piece of the Digital Preservation Puzzle. Available at [http://ils.unc.edu/Callee/oss\\_preservation.htm](http://ils.unc.edu/Callee/oss_preservation.htm) (Accessed on 09.06/2016).
- [2] Kuar, V Vimal. (2008) Selection and management of open source software in libraries. Asian School of Business, Padmanabha Building, Techno Park, Trivandrum. Accessed on 01.07.2016 from <<http://eprints.rclis.org/archive/00008739/01/OSS-selection-management.pdf>>
- [3] Biswas, Goutam and Paul, Dibyendu (2008) "NewGenLib, The First Indian Open Source Software: a Study of Its Features And Comparison With Other Software", 23<sup>rd</sup> National Seminar of IASLIC held at Bose Institute on Library Profession in Search of a New Paradigm , Kolkata, 10-13 December, 2008, Special Publication No.48, PP 333-340, ISSN- 09723668
- [4] Ian H. Witten, David Bainbridge, Stefan J. Boddie, (2001) "Greenstone Open-Source Digital Library Software", D-Lib Magazine, Volume 7 Number 10, October 2001, ISSN 1082-9873 [Available in <http://www.dlib.org/dlib/october01/witten/10witten.html>] [Access on 10.2.2016]
- [5] Devika P. Madalli, (2009) "A Digital Library of Library and Information Science using DSpace", Available in <http://drtc.isibang.ac.in> (Access on 5.02.2016)
- [6] Sharad Kumar Sonkar, Veena Makhija, Ashok Kumar, and Mohinder Singh, (2005) "Application of Greenstone Digital Library (GSDL) Software in Newspapers Clippings", DESIDOC Bulletin of Information Technology, 2005, 25(3), 9-17
- [7] Meitei, L Shanta and Devi, Purnima (2009), "Open Source Initiative in Digital Preservation: The Need for an Open Source Digital Repository and Preservation System", 7th International CALIBER-2009, Pondicherry University, Puducherry, February 25-27, 2009  
[Available in <http://www.inflibnet.ac.in/caliber2009/CaliberPDF/6.pdf>] (Access on 09.06.2016)

## ***The Study on the Open Source Digital Library Software for Institutional Repository in India: A Survey***

[8] Wikipedia: Available at [http://en.wikipedia.org/wiki/Digital\\_library](http://en.wikipedia.org/wiki/Digital_library) (Access on 12.12.2016)

[9] Madalli, Devika P (2008) "A Digital Library of Library and Information Science using DSpace" [Available in <http://drtc.isibang.ac.in>] (Access on 11.12.2008)

[10] Das, Anup Kumar (2008), "Open Access to Knowledge and Information. Published in Open Access to Knowledge and Information in South Asia: Scholarly Literature and Digital Library Initiatives-The South Asian Scenario". Edited by Bimal Kanti Sen and Jocelyne Josiah. New Delhi: UNESCO, 2008.

[11] Wikipedia, Available at <http://fedoraproject.org/wiki/Overview> (Accessed on 16/01/2017).

### **Annexure: Examples of Some Indian Repository**

<b>Software</b>	<b>Institution</b>
DSpace	Indian Institute of Science
DSpace	National Chemical Laboratory
DSpace	University of Hyderabad
Greenstone	Indian Institute of Technology - Mumbai
DSpace	Vidyanidhi - National e-theses repository
Greenstone	Indian Institute of Management, Kozhikode

<b>Software</b>	<b>Institution</b>
GNU EPrints	Indian Institute of Science
GNU EPrints + DSpace	Indian Institute of Management - Kozhikhode
GNU EPrints	Indian Medlars Centre, NIC, OpenMED@NIC
GNU EPrints	National Aerospace Laboratory
GNU EPrints	One World South Asia
DSpace	National Institute of Technology - Rourkela
DSpace	Indian Statistical Institute - Bangalore
DSpace	Documentation Research and Training Centre (DRTC)
DSpace	Indian Institute of Technology - Delhi
DSpace	INFLIBNET – Inter University Centre
DSpace	Indian Institute of Astrophysics, Bangalore

DSpace	Raman Research Institute, Bangalore
GNU EPrints	University of Delhi
GNU EPrints	Medknow Publications -MedknowEprints
DSpace	Indian National Science Academy (INSA)
GNU Eprints	Indian Institute of Information Technology
GNU Eprints	Rajiv Gandhi Centre for Biotechnology

---