

Development of Some Selected Institutional Repository in India: A Study Based on Open Access Resources

Manas Kr Ganguli

Assistant Librarian cum Information Officer,
Documentation Centre, Philips India Ltd.

Mohit Banerjee

Manager & Chief Librarian,
Documentation Centre, Philips India Ltd.

Abstract:

Institutional Repositories (IR) are digital collections of the outputs created within a university or research institution. This paper describe about the IR(Institutional Repository) technology implementation in Indian institutes, and also discuss its objectives, uses of software, growth and development of IR in India. The paper seeks to provide an overview of Institutional Repositories its benefits to the institutions and also describes the role of library.

***Keyword:** Institutional Repository, Digital collection, Open access, Dspace, Eprint, Contentdm, Metadata*

1. Introduction

Institutional repositories are repositories designed to manage, host, preserve and distribution of the scholarly output of an institution. Institutional repository are a managed storage system with content deposited on a personal, departmental, institutional, national, regional or consortial basis providing services to designed communities with content drawn from the range of digital resources that support learning, teaching and research.

Digital repositories broadly to include:

- IR which aim to collect widely across a particular university or similar institution, possibly in a wide range of formats.
- Subject repositories based on collecting only within a certain discipline.probably across more than one institution.
- Format repositories whose scope is limited by collecting in a particular format, perhaps student dissertations and thesis or research data.

In addition researchers were also asked about their use of personal web pages to store and make openly available their own publications, a function similar in some respects to that of Institutional Repositories.

As per Wikipedia "A repository is a central place where data is stored and mined. A repository can be a where multiple databases or files are located for distribution over a network or a repository can be a location that is directly accessible to the user without having to travel across a network" en.wikipedia.org/wiki/repository

Essential Characteristics of an IR

An institutional repository in fact is a virtual collection consisting of single or multiple types of intellectual products created in digital form. The Scholarly publishing and Academic Resources Coalition (SPARC) position paper illustrates four essential characteristics that an IR should have:

- Institutionally defined
- Scholarly content
- Cumulative and perpetual
- Interoperability and open access

Key advantages v/s disadvantages

- Expansion of the range of knowledge that can be shared
- Opportunities to simplify and extend dissemination
- Enabling of IPR to be exploited more effectively at institutional level
- Leverage of existing investment in information and content management systems
- The highlighting of the quality of intellectual capital
- Opportunities for new forms of scholarly communication
- Flexible ways to develop existing scholarly communications

An IR can fail over time if the

- institution stops funding,
- management failure or incompetence,
- Or technical problems.

Any of these failures can result in the disruption of access, or worse, total and permanent loss of material stored in the institutional repository.

Content of IR

The content of an Institutional Repository could be:

- Pre-prints of articles or research reports submitted for publication
- The text of journal articles accepted for publication
- Revised texts of published work with comments from academic readers
- Conference papers
- Teaching materials
- Student projects
- Doctoral theses and dissertations
- Datasets resulting from research projects

- Committee papers
- Computer software
- Works of art
- Photographs and video recordings

2. Objective of the study

The general objectives for development of any IR are –

- To create global visibility for an institution's repository
- To collect content in a single location
- To provide open access institutional research output by self-archiving.
- To store and preserve other institutional digital assets, including unpublished or otherwise easily lost ("grey") literature (i.e. theses or technical reports).

The objectives of the present study are to understand the following based on 30 repositories which are ranked in website <http://repositories.webometrics.info/en/asia/india> (visited on 27/08/2014) –

- To understand the development of the inventory of repository of individual institutions
- To understand the availability of different types of resources in the individual repository
- To identify the specific users who are using information frequently
- To understand the application used in individual repository
- To understand the user-friendliness of the front end application and strength of back end database
- To understand the future of the selected repository

3. Scope in Indian aspect

India has adopted the open access much ahead of other developing countries. In India a number of scientific institutions, universities and corporate R & Ds produce high quality research accompanied by innumerable scholarly communications published by national and international journals and conference proceedings. More than 31 academic and research institutes have set up their IR as indicated by ROAR (Registry Open Access Repositories) viz., IIM Kozhikode, Indian Statistical Institute, National Chemical Laboratory, National Institute of Technology, National Informatics Centre, National Institute of Oceanography and so on.

Three kinds of repositories development in India:

- Designed for specific in house scholars

The institutions have established open access institutional repositories (IRs) that disseminate research outputs of respective institutions. Sometimes, these are self-archived. Otherwise, administrator of the repositories collects the research documents from different sources and submit the documents to the IR on behalf of the persons concerned.

- Designed for specific subjects

Few institutions repositories in India are designed to store and provide access to specific subject collections of documents. The reason behind organizing such a open repositories is that scholars with example of scholarly output but not affiliated to any specific institutes providing the IR facility can host their research articles which belongs to the respective subject field of interest. Exmple-DRTC is a subject specific repository. Openmed@NIC National Informatics Centre, New Delhi store and provide access to biomedical literature.

➤ Designed for specific documents

This IR is designed to store and provide access to documents pertained to specific type of collections. Vidyanidhi of University of Mysore is an example of document type specific collection that stores and provides access to theses and dissertations. Vidyanidhi accepts any thesis or dissertation from any researcher or student that is accepted in any of the Indian universities or institutions.

Availability of software for developing IR:

There are number of software's available for creating/developing institutional digital repositories; the brief of some IR are given below:

Open source digital repository softwares:

DSpace

EPrints

Fedora

Greenstone

Commercial Digital Repository Software's:

Apart from above Open Source Software, some commercially developed software's also available for digital repository. The name of few is mentioned below:

CONTENTdm@-

DigiTool-

EN Compass-

Hypersion-

Meta Source-

VITAL-

Steps involved in creation of IR:

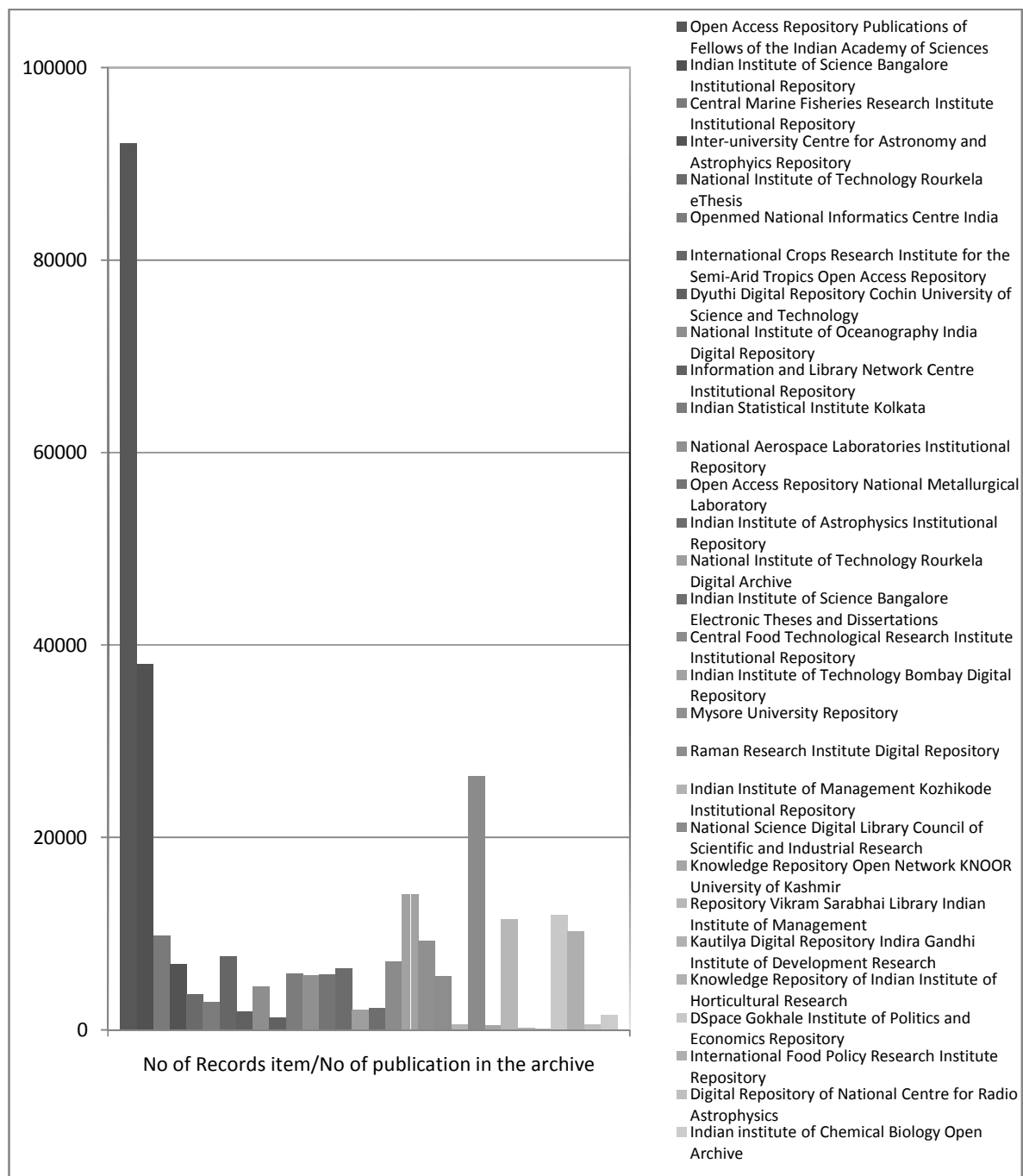
- Data identification and selection
- Data collection and Segregation
- Design and development
 - Software and hardware
 - Content creation and development
 - Content organization
- Web hosting
- Maintenance

4. Methodology

The data for the study was obtained using website of selected institutions. The details list of individual website is affixed in Annexure – I. A list of questionnaire is also been prepared and circulated to individual institutions among different stakeholders through e-mails to collect the data primarily. Then, the collected data were analyzed, tabulated and presented using MS- Excel 2007.

5. Development & analysis

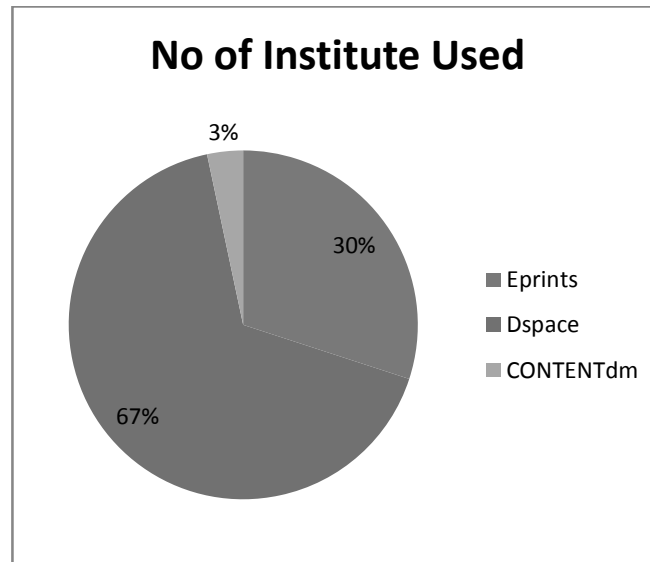
Institutional repositories enhance teaching, learning, research and are considered as a boon to the scholarly community. Institutional repositories have great potential for improving visibility and impact of institutional research.



Software Name	No of Institute Used	%
Eprints	9	30

Dspace	20	67
CONTENTdm	1	3

Table 1: Above table showing the Repository software uses



6. Problems & challenges

From the above study of 30 Indian repositories, it is found that growth rate of IR per year is quite low in compare to other developed countries. Only few have included learning object and multimedia documents. Theses and dissertation is the common object.

Data collected from 'Ranking Web of World Repositories', Retrieved on May 03 Sept 2014, from <http://repositories.webometrics.info/toprep.asp>

The problems and hurdles which implementation teams face in building a repository include the following-

- Lack of awareness
- Ignorance of users in the absence of appropriate promotion program
- Poor bandwidth
- Inadequacy of generation of digital resources
- Non availability of telecommunication infrastructure
- Lack of Institutional repositories expertise
- Insufficient fund of IT infrastructure and manpower

- Difficulties in managing Intellectual Property Rights
- Problems related to customization of open source software is bottle neck

Library staff needs training about software installation and customization, understanding the service goals, user interface search method. In order to promote IR one should-

- Increase awareness and training through followup workshop.
- By advertising its services on its website and in the print
- Develop FAQ and online power point presentations.

Finally be proactive in publicizing repository developments via institutional newsletter and bulletins, seminars and workshops and e mail alerts etc.

Functional Area	Testing Criteria	Summary of Findings
Data Model	<ul style="list-style-type: none"> • object structures • collections structures • external aggregations 	<p>DSpace and EPrints have restrictive data models (although DSpace is currently more flexible than EPrints) which reduces their ability to hold complex, structured objects. Both would require modification or significant customisation to support the range of digital objects in scope. Optimized for customization, CONTENTdm also accommodates institutions that want to make more advanced customizations.</p>
Ingest, Data Management and Administration	<p>Ingest</p> <ul style="list-style-type: none"> • command-line interface • Web interface • machine interface • batch import • custom workflows • metadata capture 	<p>All repositories provide sufficient functionality to support flexible human- and machine-ingest for a range of materials with different workflows, although configuration would be required for all.</p> <p>All repositories provide sufficient functionality to support digital object management and administration operations on content stored within the repository.</p> <p>Additionally, records in existing</p>

		<p>databases can be easily imported into CONTENTdm, while exporting metadata is simple using the XML export function. Import and export capabilities are compatible with legacy, local, regional and national systems. CONTENTdm also fully supports the OAI-PMH protocol for harvesting metadata from your digital collections into other systems such as your integrated library system.</p>
Descriptive Information (Metadata)	<ul style="list-style-type: none"> • persistent identifiers • human-readable, hierarchical identifiers • support for standards • OAI compatibility 	<p>DSpace and EPrints suffer from a lack of support for identifier schemas (although EPrints is significantly worse) which tends to reduce options for making collection and object identifiers independent from the repository software and human-readable interpretations of collection hierarchies. CONTENTdm also fully supports the OAI-PMH protocol for harvesting metadata from your digital collections into other systems such as integrated library system.</p>
Storage	<ul style="list-style-type: none"> • file system structures • integrity checking 	<p>DSpace and EPrints suffer from a lack of attention to preservation storage, both requiring the maintenance of the repository software and environment to retrieve digital objects in the event of a hardware or software failure.</p>
Access	<ul style="list-style-type: none"> • batch export • Web front-end 	<p>DSpace and EPrints have limitations in their access</p>

	<ul style="list-style-type: none"> • API/URI schemas • indexing engine • access control approach • authorisation • authentication 	<p>functionality. Both provide out-of-the-box front-ends which are designed for open access publications, and which would require modification to support digitised collections and born-digital archives. Their machine interfaces to support the development of independent Web applications are also limited.</p> <p>The Gateway's Web-based interface facilitates synchronization of metadata from all OAI-compliant repositories, including CONTENTdm, with WorldCat. End users are able to search, discover and retrieve your digital items through WorldCat.org™, WorldCat® Local and search engines—like Google and Yahoo! Search—and then view them in theCONTENTdm collection. With WorldCat, your digital collections will have unparalleled visibility on the Web.</p>
--	--	--

Table 2: Repository software comparison by functional area, testing criteria and summary of findings

Feature	Contentdm	Dspace	Eprints
Cost:	Commercial	Free	Free
Database:	XML/Text based	Oracle, PostgreSQL	MySQL, Oracle
News Feeds:	-	RSS	RSS,Atom
Metadata Formats:	Dublin core, Qualified DC,VRA Core	Dublin core, Qualified DC	Dublin core, Qualified DC,METS,MARC

Table 3: Repository software comparison by cost database News feeds Metadata <http://www.rsp.ac.uk/documents/Repository-Software-Survey-2009-03.pdf>

Repository softwares are not equal. Competing solutions have relative strengths and weaknesses in different functional areas, which is most likely because they were designed to solve different problems. Specific aspects of our context further supported our selection of repository software:

- The heterogeneous nature of our collections and the pattern of emerging requirements naturally disposes us to a modular architecture, where functionality in different collection and technical areas is independent;
- Although digital preservation is of increasingly urgent concern, we have systems already in place handling aspects of collection management, so we are not starting from scratch and attempting to develop a solution in all functional areas in one go;
- Despite competing demands for resources, increased in intensity by the financial situation in the sector, we do have access to members of staff to carry out technical development and maintenance.

CONTENTdm provides a comprehensive solution for local history archives, slide libraries, “born-digital” items, newspapers, books, letters, maps, electronic theses and dissertations, and audio/video files.

Benefits

Flexibility. CONTENTdm is easy to learn, easy to use, fast to start and scalable for growth as your collections evolve. **Discoverability.** By publishing your collections on the Web and adding metadata to WorldCat, your collections are more visible and discoverable by local and global users. **Interoperability.** CONTENTdm’s import/export

capabilities are compatible with legacy, local, regional and national systems.

Standards-based. CONTENTdm supports numerous industry standards including Unicode, Z39.50, Qualified Dublin Core®, EAD, VRA, XML, JPEG2000 and OAI-PMH.

Security options. Restrict access to collections or items by user name or IP address, or limit access to images while allowing access to metadata.

7. Conclusion

The Indian researchers, library and Information professionals, administrators of academic and R & D institutions must be aware of the new opportunities provided by Information and Communication technologies and the advantages of IRs. The awareness programs, training and workshops funded by the leading organizations like UGC, CSIR etc., to educate and instruct the faculty members, researchers and scientists and all other working bodies to adopt the “open access” approach must be more widespread. Most the universities in India are lacking in infrastructure for establishing institutional repositories but today with the availability of several free, open source repository software packages it has become relatively easy to establish

IRs. So, all the universities, R & D institutions in India which are being run by public or private fund should also establish their own repositories to make their research widely accessible by scholarly community throughout the world and to ensure its long term preservation for future use.

8. References

1. Directory of Open Access Repositories(DOAR) <http://www.doar.org> accessed on 25.10.2014
2. E prints@ NML Digital Repository <http://eprints.nmlindia.org/> accessed on 03.09.2014
3. Hirwade, M. & Hirwade, A. (2006). Institutional repositories: challenge and opportunity for LIS Professionals in digital age. *Library Herald*, 44(2), 146-151. Retrieved May 3, 2014 from <http://eprints.rclis.org/bitstream/10760/7987/1/7D311B68.pdf>
4. <http://repositories.webometrics.info/toprep.asp> accessed on 25.10.2014
5. <http://www.rsp.ac.uk/documents/Repository-Software-Survey-2009-03.pdf> accessed on 20.09.2014
6. Nazim, M. & Maya Devi (2008). Open access journals and institutional repositories: Practical need and present trends in India. *Annals of Library and Information Studies*, 55,27-34. Retrieved May 8, 2014 from <http://nopr.niscair.res.in/bitstream/123456789/825/1/ALIS%2055%281%29%20%282008%29%2027-34.PDF>
7. Poornima, N., Biradar, B.S. & Goudar, I.R.N. (2008). Enhancing the impact of Indian
8. scholarly communication through institutional repositories. ILA Platinum Jubilee
9. International conference on quote; Knowledge for all: Role of libraries and information
10. centers quote, 12-15 November, Mumbai, India. Retrieved May 10, 2014 from <http://nal-ir.nal.res.in/4924/1/poornima.docx>
11. Registry of Open Access Repositories (ROAR) <http://archives.eprints.org>
12. Robinson, M. (2009). Promoting the visibility of educational research through an institutional repository. *Serials Review*, 35(3), 133-137. Retrieved May 3, 2014 from <http://libir1.ied.edu.hk/pubdata/ir/link/pub/Robinson IR Paper Final.pdf>
13. Sahu, A K, Goswami, N G, Chowdhury, B K (2013). Institutional Repository of CSIR-NML and global information seeker. *Annals of Library and Information Studies*. V.60,36-40. ([http://nopr.niscair.res.in/bitstream/123456789/17689/1/ALIS%2060\(1\)%2036-40.pdf](http://nopr.niscair.res.in/bitstream/123456789/17689/1/ALIS%2060(1)%2036-40.pdf))(Accessed on 03.09.2014)
14. Thomas, G. (2007). Evaluating the impact of the institutional repository, or positioning

15. innovation between a rock and a hard place. New Review of Information Networking,
 16. 13(2), 133-146. doi:10.1080/13614570802105992
 17. www.oclc.org/contentdm/ accessed on 30.10.2014

Annexure-1:

Data collected from 'Ranking Web of World Repositories', Retrieved on Sept 03, 2014

from <http://repositories.webometrics.info/toprep.asp>

Indian ranking	World Rank	Institution	No of Records item/No of publication in the archive	Software used	User	Collection
1	236	Open Access Repository Publications of Fellows of the Indian Academy of Sciences	92160	Eprints	Fellows of the Indian Academy of Sciences	meetings, discussions, seminars, symposia and publications.
2	276	Indian Institute of Science Bangalore Institutional Repository	38014	Eprints	limited to the IISc research community only	preprints, postprints and other scholarly publications using a web interface
3	283	Central Marine Fisheries Research Institute Institutional Repository	9809	Eprints	CMFRI scientists who do research on fisheries and related areas	journal papers, conference papers, reports, theses, patents etc. - are uploaded/self-archived by CMFRI scientists who do research on fisheries and related areas.
4	284	Inter-university Centre for Astronomy and Astrophysics Repository	6829	Dspace	Inter university students & scholar	Conference paper, Presentation, research publication
5	345	National Institute of Technology Rourkela eThesis	3765	Dspace	Students	theses
6	365	Openmed National Informatics Centre India	2904	Eprints	NIC community	including biomedical, medical informatics, dentistry, nursing and pharmaceutical sciences.
7	432	International Crops Research Institute for the Semi-Arid Tropics Open Access Repository	7636	Dspace	Open to User	Research areas
8	473	Dyuthi Digital Repository Cochin University of Science and Technology	1944	Dspace	Open to User	PhD theses such as scientific publications, conference E-proceedings, multi media contents, Books etc.
9	553	National Institute of Oceanography India Digital Repository	4548	Dspace	Research community	journal articles, conference proceeding articles, Technical reports, thesis, dissertations, etc
10	582	Information and Library Network Centre Institutional Repository	1326	Dspace	Open to User	articles published in all conventional proceedings of INFLIBNET Centre
11	682	Indian Statistical Institute Kolkata	5854	Dspace	Open to User	Conference paper, Dissertation, theses, Institutional docs and records, lectures, Scientific contribution

12	699	National Aerospace Laboratories Institutional Repository	5728	Eprints	Open to User	Theses Dissertation
13	735	Open Access Repository National Metallurgical Laboratory	5825	Eprints	Open to User	Locally produced research outputs - Journal papers, conference papers, reports, theses, patents etc. - are uploaded here by NML authors.
14	858	Indian Institute of Astrophysics Institutional Repository	6396	Dspace	Open to User	Theses Dissertation
15	862	National Institute of Technology Rourkela Digital Archive	2048	Dspace	Open to User	Theses Dissertation
16	903	Indian Institute of Science Bangalore Electronic Theses and Dissertations	2304	Dspace	Scholar or Faculty & others	Theses Dissertation
17	932	Central Food Technological Research Institute Institutional Repository	7157	Eprints	Scholar or Faculty & others	research output created by the CFTRI scientists, researchers, staff and students community.
18	982	Indian Institute of Technology Bombay Digital Repository	14100	Dspace	Scholar or Faculty & others	best research including full-text of book chapters, conference/proceeding papers, technical reports, journal pre-prints & post-prints, working papers, Patents and others like annual reports etc.
19	1035	Mysore University Repository	9291	Eprints	library staff, the faculty and research scholars	cover scholarly publications covering journal articles, conference papers, books, book reviews, presentations, reports and patents
20	1074	Raman Research Institute Digital Repository	5601	Dspace	Scholar or Faculty & others	research publications of the faculty and students of the Raman Research Institute. The collected papers of C.V. Raman and the historical records of the institute (Annual Reports and Newspaper Clippings) housed here.
21	1123	Indian Institute of Management Kozhikode Institutional Repository	532	Dspace	Scholar or Faculty & others	archive their preprints, post prints and other scholarly publications.
22	1135	National Science Digital Library Council of Scientific and Industrial Research	26413	Dspace	Open to User	S&T information to students of science, engineering and technology in the country.
23	1161	Knowledge Repository Open Network KNOOR University of Kashmir	450	Dspace	Scholar or Faculty & others	Theses, dissertation, research output etc.
24	1427	Repository Vikram Sarabhai Library Indian Institute of Management	11496	Dspace	Scholar or Faculty & others	Theses, Working Papers, Annual Reports and bibliographic information on book chapters, monographs, research project etc
25	1449	Kautilya Digital Repository Indira Gandhi Institute of Development Research	249	Dspace	Scholar or Faculty & others	conference proceedings, thesis and dissertations, working papers and research articles, etc
26	1518	Knowledge Repository of Indian Institute of Horticultural Research	160	Dspace	Scholar or Faculty & others	research output of the institution.
27	1625	Dspace Gokhale Institute of Politics and Economics Repository	11989	Dspace	Scholar or Faculty & others	Research outputs theses

28	1661	International Food Policy Research Institute Repository	10247	CONTE NTdm	Scholar or Faculty & others	collection of literature, including research articles, working papers, policy documents, journal articles, donor reports, presentations and other publications, related to agriculture, food security and nutrition in Asia
29	1680	Digital Repository of National Centre for Radio Astrophysics	529	Dspace	Scholar or Faculty & others	Lectures notes Theses workshop etc.
30	1717	Indian institute of Chemical Biology Open Archive	1536	Eprints	IICB researchers	Research documents, which include peer-reviewed journal articles, conference papers, theses and reports, produced by are archived here.