

## Building an Institutional Repository: A TIFR Initiative

Rekha Dabholkar

R Prabakaran

B T Kurahatti

### Abstract

*Building an Institutional Repository has twin objectives. It serves as a digital archive of the total research output created by the academic community of an institute. It makes freely accessible over the internet this intellectual output to all those who are interested in it, thereby promoting the knowledge development. TIFR too is therefore planning to build its Institutional Repository and as a prelude to this has undertaken a pilot project. This paper discusses some of the details of this initiative.*

**Keywords:** Institutional Repository, DSpace, Digitization

### 1. Introduction

With the process of globalization in knowledge activities, the demand for information has been growing steadily in all spheres of work. The concept of access to information free of charge in electronic format is gaining momentum. Creating Institutional Repository is a step in this direction. Institutional Repository is a very powerful idea that can serve as an engine of change for any institution. Institutional Repository by capturing, preserving and disseminating collective intellectual capital, serves as meaningful indicators of an institution's academic quality. An Institutional Repository concentrates the institutional product credited by an academic or other institutions researchers, making it easier to demonstrate its scientific, social and financial values. The benefit of institutional repositories is that they enable free sharing of information, encouraging collaboration and the widespread communication of institutional education and research activity. Digital institutional repositories have become a hot topic over the last few years, and many institutions around the world are now actively considering or working towards implementing them. TIFR too is planning to create an Institutional Repository and has presently undertaken a pilot project. This paper discusses some of the details of this initiative.

### 2. About TIFR

The Tata Institute of Fundamental Research (TIFR) was established in 1945 and has the mandate of providing leadership in fundamental research. There are about 400 scientists working in the core areas of Physics, Astronomy, Biology, Chemical Sciences, Computer Science and mathematics. The various research departments function under three major schools:- the School of Mathematics, the School of Natural Sciences and the School of Technology and Computer Science.

#### 2.1 Scientific Information Resource Centre

The Scientific Information Resource Centre (SIRC) (earlier library) was established in 1945, almost simultaneously with the Institute with the objective to cater to the scientific literature requirements

---

of the academicians. Though, SIRC is primarily responsible for catering to the information needs of scientific community in TIFR, it also serves scientists of other Institutes in the country, as it is a National Centre of the Government of India for Nuclear Science and Mathematics.

## **2.2 Resources**

SIRC has a collection of 1,52,000 volumes including books and journals. It has subscription to 559 current journals. Besides, it has access to another 4000 electronic journals made available through full-text online databases such as Science Direct, Springer LINK, ACM Digital Library, IEL Online, etc. Of late, it has also made available to users the e-books published by some leading publishers. It also has access to bibliographic databases such as Web of Science, MathSciNet, CiteSeer, etc. Some of the reference sources are also made available via CD Mirror Server. In addition to these, it has theses, maps, a special and rare collection of "Manuscripts of Prof. Harish Chandra and Prof. C.P. Ramanujan" and original drawings of Dr. Homi Bhabha etc.

## **3 Resources Planned for Institutional Repository**

TIFR, where the frontline research is carried out, produces a good number of research papers. Any faculty, Ph.D. student working at TIFR can submit their research papers to the Institutional Repository. The students can submit Theses and Dissertations. Many academic lectures are held in TIFR and these are published in the form of Lecture Notes. TIFR also brings out regularly technical reports which are being planned to be included in the repository. Besides, as already mentioned above, SIRC has rare collection of manuscripts of some well known Mathematicians. These would also be part of the repository. Dr Homi Bhabha has published several research papers. There is also invaluable correspondence made by Dr Bhabha with well-known scientists and other personalities. All these are planned to be made accessible to the academic world.

### **Objective**

To build an innovative institutional digital repository to collect, preserve, and enable distribution of research, teaching and learning material generated by TIFR scholars, faculty and researchers and materials of historical importance.

### **Requirements**

For the full-fledged implementation of Institutional Repository, there is a need for proper infrastructure and other requirements, some of which are already in place. The details of these requirements are as follows.

#### **3.1. Hardware requirements**

Two Personal Computers with the P-IV configuration are enough for undertaking digitization processes. A separate high-end server which is functional 24/7 is required for an Institutional

Repository. More disk space is required for back-up of the data. RAM capacity should be adequate to store the graphic data. Scanners of various capacities from simple flat-bed scanners to heavy duty scanners are available. However, it should be ensured that the software attached to the scanner has the facility of Optical Character Recognition (OCR) conversion. For the present pilot study, a flat-bed HP Scanjet scanner is available.

### 3.2. Software Requirements

There are several open source softwares available which can be used for building an Institutional Repository. For the present purpose, a comparative study of three open source softwares viz., Greenstone Digital Library Software (GSDL), EPrint and DSpace was carried out. However, for building an Institutional Repository in TIFR, it was decided to opt for DSpace for the following reasons.

- a) DSpace is completely customizable to fit the needs. It has flexibility, functionality and can be maintained with minimum staff time. It has Qualified Dublin Core.
- b) Its structure allows to organize the repository collections according to various departments in the Institute.
- c) DSpace is being used by many libraries. Since this software is updated regularly, it helps in enhancing the performance of the retrieval system.

The software in-built in scanners is suitable to carry out digitalization work. However, Abbey Fine reader will be a better option among other OCR softwares available. The scanned images can be edited for clarity in the Photoshop software. Besides, softwares such as SQL server software, database management software, web designing softwares like java, front-page, XML etc., and Adobe Acrobat Distiller Professional 8.0 are needed.

Hardware Requirement	Software Requirement
2 P-IV computers, High-end server, Scanner e.g. Minolta, Kinoca, BOOK scanner, CD/DVD writer, web servers and FTP servers, UPS (10-20 KV), high speed LAN and internet connectivity.	OCR Software (ABBYY Fine Reader OCR 6.0 Professional), Adobe Acrobat Distiller Professional 8.0, CD/DVD read / writer software, windows NT networking software, SQL server software, database management software and web designing software like java, front-page, XML etc.

### Hardware and Software Requirement

### 3.3. Human Resources

Human resources are one of the most important requirement for building an Institutional Repository. The staff has to be dedicated. They need not be IT professionals. However, they should possess some skill sets needed for different activities such as installation of various software, scanning and digitization, uploading of files and other activities involved with DSpace and etc.

### 3.4. Financial Requirements

One needs PIV computers, a separate server for the installation of DSpace and Scanners. Besides, tape drives for backups, CD/DVD writers and laser printers are also needed. The open source software, obviously, comes free. There is also a cost involved in training of the staff on digital library software (DSpace). A rough estimate of the present project is given below:

Category	Item	Cost
Hardware	PIV Computers, High end server, Scanner and CD/DVD Writer	Rs. 6,25,000.00
Software	ABBYY Fine Reader OCR 6.0 Professional and Adobe Acrobat Distiller Professional 8.0	Rs. 50,000.00
	Total	Rs. 6,75,000.00

The pilot project is being carried out by the existing staff. However, for full-fledged Institutional Repository, the staff requirement and the associated cost are as follows:-

Human Resource	5 staff to be hired for 12 months(5X12X7000)	Rs 4,20,000.00
----------------	--	----------------

## 4. Implementation of the Project

### 4.1. Installation of DSpace

DSpace has been installed in one of the computers for the pilot project. This essentially involved installing of four prerequisite softwares i) Java SDK, ii) Apache Ant, iii) PostgreSQL, iv) Tomcat. These software are installed in sequence. DSpace has an in-built feature for customization and administration.

### 4.2. Digitization

Digitization includes acquiring, converting, storing and retaining information in standardized and organized manner with technological support. The documents are being scanned in the resolution of 300 to 600 dpi (dot per inch) depending on the physical condition of the documents. The next

process, which is an important task involves rectifying the errors and make text files error free with the help of OCR process. Scanned images are then being stored in the computer in PDF (for text) and JPEG (images) formats. These files need processing in order to make them suitable for access through a network. This sometimes involves compression of PDF and JPEG file formats.

### **4.3 Workflow in DSpace**

Submission of any document has to pass through a seven-step workflow process. The first three steps include basic descriptions about the materials and Dublin Core metadata elements like author, title, publisher etc. The fourth step involves uploading of the digital document i.e. PDF and image files which are earlier stored in the computers. In fifth step, includes verification and making of corrections, if any, to the data submitted in the earlier stages. The sixth step presents the license agreement and the last step confirms the submission process. After submission is complete, the submitted item goes through other formalities like review, edit, or approval in accordance with the established policies.

DSpace also has a remote publishing facility, where authorized users can submit their items directly from their desktop (This is yet to be implemented).

## **5. Browsing and Searching through DSpace**

The built-in Lucene Search Engine in DSpace facilitates to browse and search the collections in Institutional Repository. It also, has the facility of full-text search.

### **5.1. Browsing**

Browse allows to go through a list of items in some specified order. DSpace allows to browse through Community / Collection by Title, Author, Subject and Date.

### **5.2. Searching**

It allows to make a search by the title, author, subject, series and date. Search can be limited to a specific community or collection.

### **5.3. Advanced Search**

The advanced search allows to specify the fields one wish to search, and to combine these searches with the Boolean "and", "or" or "not".

## **6. Copyright Issues**

This is an important issue that one has to look into while creating Institutional Repository. Author's rights and interpreting publishers' copyright policies vis a vis the institutional repositories are currently being hotly debated. Internet and the digital revolution pose even more complex problems. As it is still an evolving subject, SIRC, initially has decided to make available the digitized materials on

---

intranet to internal users only. Once, the concrete laws emerge, it can think of making the content available to the outside academic world.

## 7. Conclusion

Building an Institutional Repository is a natural task for librarians as technology continues to evolve exponentially and as the amount of electronic information mushrooms. An institutional repository can boost the prominence of the institution and optimize dissemination of the institution's publications to a worldwide audience. Repositories can provide significant benefits after an initial investment of time. TIFR with availability of a sound infrastructure is gearing up to take up this task on a full-fledged way and its qualified staff will rise to the challenge of building a repository.

## References

1. Barwick, Joanna (2007). Building an institutional repository at Loughborough University: some experiences. Program: electronic library and information systems, 41(2), 113-123
2. Dspace Website (<http://www.dspace.org/> )
3. IASLIC (2006). Open Source Movement- Asian perspective: papers XXII National Seminar of IASLIC and Special Interest group (SIG) Meetings. Kolkata: IASLIC.
4. Jain, Sanjeev Kumar; Shrivastava, Anurag (2008). Academic Institutional Repositories in India: Global Visibility for an Institution's Scholarly Communication. (<https://drtc.isibang.ac.in/handle/1849/413>)
5. Kumar, K. Manoj (2006). Digital preservation, management and access to information in the twenty-first century. Ahmedabad: INFLIBNET Centre.
6. Sreekumar, M.G (et. al.) (2007). Institutional repositories for knowledge management in academic and research institutions. ( <https://drtc.isibang.ac.in/handle/1849/404> )
7. Witten, I.H. and Bainbridge, D. (2003). How to build digital library. Amsterdam: Morgan Kaufmann Publishers.

## About Authors

**Ms. Rekha Dabholkar**, Tata Institute of Fundamental Research, Colaba, Mumbai 400005

**Ms. R Prabakaran**, Tata Institute of Fundamental Research, Colaba, Mumbai 400005

**Ms. B.T. Kurahatti**, Tata Institute of Fundamental Research, Colaba, Mumbai 400005