Chapter 1

INSTITUTIONAL REPOSITORY
INSTITUTIONAL REPOSITORIES

INTRODUCTION

In India, there are a number of internationally reputed institutions, which are producing a good number of research documents that are expanding the frontier of knowledge and scope of technological innovation. The recent phenomenon of outsourcing of scientific research, applications development as well as business-intelligence-related research to India has generated a good deal of enthusiasm in advanced studies. Recently some internationally famous journals published cover stories on Indian scientific and technological research. The success and achievements of these institutions can give rise to promising research environment in India that may attract sponsored or collaborative research in all disciplines of study. These institutions essentially use modern information and communication technologies for information management and dissemination. Other than the research activities, these institutions also produce intellectually mature graduates and scholars in some disciplines. Other areas of studies in India are also getting international attention proportionately. Some of the institutions provide access to their research documents and learning materials initially to the Indian scholars in other institutions as well as external scholars in institutions across the globe. The sharing of knowledge may lead to further development in the same discipline or related disciplines. Institutional repository, which may be called an extension of digital library, is now becoming a platform for the sharing of knowledge.

In a general sense, an “institutional repository” can mean many things a library, an archive, a museum, or even a warehouse that stores for use and safekeeping an organization’s records or artifacts falls under the broad definition of institutional repository. In recent years, however an IR has taken on a more specific, but still evolving meaning that refers to the
storage and preservation of an organization’s digital information or knowledge assets.

It is a set of services offered by a university to its community members for the stewardship of scholarly publication generated by the faculty, staff, and research scholars by preserving it for long term. Since the resources are generated digitally and electronically it is very easy to build a collection of any specific subject discipline or any targeted user group. It has dual role so far its functions concerned. It may be self generated knowledge-base by the university itself on one hand, on other hand it may be the substitute model for the publication channel. With the coalition of these efforts at different demographic strata, it impacts very powerfully to the scholarly world.

An Institutional Repository consists of formally organized and managed collections of digital content generated by faculty, staff and students at an institution. This is the collective intellectual output of an institution, recorded in a form that can be preserved and exploited. There are the result of the vision to collect, secure and provide access to scholarly publication in a novel, digital way, mostly initiated by the institutional library. Institutional repositories are spreading, as they have become a dispensable component for information and knowledge sharing in the scholarly world.

Institutional Repository may be defined as information systems capable of capturing, preserving and providing access to the intellectual output produced by the members of an institution. Within the specific context of academia, an Institutional Repository may contribute to the increase of the institution prestige and nourish the idea of an “Institutional Repository” because it manages and preserves relevant informational items that otherwise would remain scattered, unattended or unaccessible.
Institutional Repository provides tools that help faculty, students, and researchers to disseminate the Institutional Repository work to audiences outside the institution. Institutional Repository may serve as a complement to traditional forms of publication or as an alternate. Institutional Repository enable information seekers to find faculty and student work more easily by organizing and indexing it, making it more visible to colleagues, fund providers, and employers. The main purpose of Institutional Repository is to bring together and preserve the intellectual output of a laboratory, department, university or any other entity, the incentives and commitments to change the process of scholarly communication have also begun serving as strong motivators.

DEFINITIONS OF INSTITUTIONAL REPOSITORIES

According to Lynch (2003) “a university based Institutional Repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members.”

Foster and Gibbons (2004) defined Institutional Repository as, “an electronic system that captures, preserves, and provides access to the digital work products of a community.

Raym Crow (2004) defined as Institutional Repository as a “Digital Archive of intellectual product created by the faculty, research staff, and students of an institution and accessible to end users both within and outside the institution, with few, if any, barrier to access. The content is institutionally defined, scholarly, cumulative, and perpetual, open, and interoperable.

Chang (2003) defines an Institutional Repository as a new method for capturing, collecting, managing, disseminating, and preserving scholarly works created in digital form by the constituent members of an
institution. For the present study, the term ‘digital libraries and repositories’ include digital collection, digital archives developed using digital library and Institutional Repository software packages.

GENESIS OF INSTITUTIONAL REPOSITORIES
Repositories have existed ever since human begin collecting and storing important information and artifacts for safekeeping and long term use. The long and rich history of libraries, museum and archives provides the foundation for any type of repository program, but two contemporary developments in particular have helped shape the nature of today’s institutional repositories.

(a) The emerging knowledge management movement.

(b) The maturing, but still rapidly advancing technology of content or asset management in the digital information management.

(a) Knowledge Management Movement
In 1988, Peter Drucker published an influential paper entitled “The coming of the new organization” in the Harvard business review. In this paper, Drucker argued that for the modern Organization, knowledge had become its most important asset, and that these organizations best able to manage and exploit their corporate knowledge assets would be the most successful in the marketplace. Drucker did not explain in any detail what he meant by “knowledge” but he was clear that knowledge manifest itself in many forms in the organization, ranging from patents and trade secrets, to operational routines, to expertise inside the heads of employees. Subsequent books, articles and conferences by researcher in the multidisciplinary fields of information science and business management have elaborated on Drucker’s ideas about the importance of knowledge management in the organization. These researchers have pursued the meaning and implication of knowledge as it is differentiated from data
and information, and they have developed models, conceptual structures, and best practices for managing knowledge in the modern organization. Repositories, and that is the commonly used in the knowledge management field, play an important and supporting role in knowledge management system. Davenport and Prusak in their book “Working knowledge: How organization manage what they know, list “knowledge repositories” first in their review of knowledge management projects in practice. They claim to have come across three basic types of knowledge repositories.

1. External knowledge repositories (example: competitive intelligence.);
2. Structured internal knowledge repositories (example: research reports, production oriented marketing materials and methods.)
3. Informal internal knowledge repositories (example: discussion database.)

The knowledge management movement of the 1990s influenced the development of IR in a number of significant ways in addition to establishing the nomenclature management consultants and senior administrators endorsed the movement's emphasis on the competitive value to an organization of paying attention to its knowledge assets. And the movement's broad view of knowledge as diverse and dynamic made the identification, capture and management of knowledge assets much more complex and challenging. Books, articles and any other type of published documents were viewed as only one obvious manifestation in a wide range of explicit and tacit knowledge assists that need to be managed in an organization.

(b) Maturing digital assets management (DAM) technology:-
By the year 2000, several broad technological developments in the digital information system were also pushing individuals and their organization towards the creation of IR. First it was becoming easier for individuals or
small groups to create and to disseminate digital assets through the use of microcomputer, desktop tools and computer networking. At a university, for instance, it would not be common by 2000 to find faculty members or small disciplinary centres around campus creating digital text documents, digital multimedia, websites, or online courses.

While the highly decentralized and distributed nature of microcomputer desktop publishing was empowering to the individual, it presented some managerial or stewardship challenges for organizations interested in coordinating, sharing and preserving its units or employees digital assets. Some of the most important new developments in the technology of digital content management include:

1. The open archival information system (OAIS) model developed by an international group of information technology organizations spearheaded by NASA’s consultative committee for space data systems, which offers a comprehensive logical model describing all the functions requirements in a digital repository.

2. The open archival initiative from the library and scientific community which has developed an open archive metadata harvesting protocol (OAI-PMH) that defines a mechanism for harvesting XML-formatted metadata from repositories.

3. A metadata encoding and transmission standard (METS) developed under the sponsorship of the digital library federation, which provides a schema for encoding descriptive, administrative and structured metadata in a digital repository or library.

4. Shareable courseware object reference model (SCORM) developed by the federal government agency advanced distributed learning to provide guidance for the preparation and storage of digital educational material so that such material is “reusable, educational, interoperable and durable.”
5. Publishing requirements for industry standard metadata (PRISM), a schema under development by the publishing industry to create a common language for the metadata that describes published digital assets.

6. Open source and proprietary software such as Dspace, ePrints, Fedora, Documentum, contentdm, IBM’s content management and Arstesia Teams that offer technical infrastructure options for implementing all or part of an IR.

GENESIS OF ACADEMIC INSTITUTIONAL REPOSITORIES
Disappointment and inability created by the closed access to scholarly literature and the anomalies in journal publishing paradigm in the areas of pricing, access, copyright and host of other monopolies paved the way for open access initiatives. The core essence of open access initiatives is to make research articles in all academic fields freely available on the internet among these initiatives are the working models of open access journals and institutional repositories. These initiatives were directed to rescue the obliterated research world with the sole aim to provide a complimentary role by reforming the current scholarly communication impasse and to re-assert control over scholarship by the academia and to provide increased and uniform visibility of its researchers. While the open access journal models were experimented in the publisher and organizational level, institutional repositories were experimented by the educational institutions and the universities. Universities with its numerous intellectual activities in its scientific and scholarly arena and the most affected party in the closed and profit oriented publishing scenario grabbed this opportunity to build the repositories and eventually succeeded in its endeavor. With the mandate to bring together and preserve the intellectual output of individual institutions many
universities came forward to experiment this new flame. In these experiments, unlike the traditional expectations of resistance from the publishing monopolies, the strong resistance came from within the institution’s faculties, mainly because of its inexperience with the open archiving and open access concepts and the further copyright commitments with the publishers. While the academia feared over the uncertainty of intellectual theft, the administrators feared the exposure of their institution’s intellectual scarcity. With its initial resistance from the conservatives, slowly the academia learnt the benefits of Open Access and many came forward to publish their works in their institution’s servers. This marked the significant success of the Institutional Repository model of scholarly publishing towards creating a world of free and open access scholarly society.

**CHANGING SCENARIO OF SCHOLARLY COMMUNICATION**

There are three models in scholarly communication which the world has generally accepted, viz., i) the traditional paper based journal publishing process which has a track record of over 300 years ii) E-publishing on commercial basis and iii) the Open Access mode of publishing. The developments in IT and internet have contributed considerably to shrink the supply chain of information and it is indeed a welcome change. For the information industry, the supply chain extends from the source of information to the point of usage. The traditional journals in paper format took 36-52 weeks for journals to publish. Surprisingly enough, in this total cycle time of 1 year the value addition (generation, review, correction and printing) takes place in not more than 2-3 weeks, indicating a huge wastage of time and money contributed mostly by non-value adding links. In this context, the advent of the E-publishing has really been a boost to the scholarly publishing domain, bringing down the
publishing time frame to a remarkable 3-4 weeks. A noticeable departure from the traditional systems here is that the printing is delinked from the publishing process. It is important to note here that the much appreciated and respected scholarly value systems are not compromised even by a single degree in relation to quality checks in the new process. In fact it improves and strengthens them with its inherent advantages of being online with respect to processes and procedures. The third category is the growing sets of open access publishing and scholarly archive initiatives, which are the offshoots of the novel open access movement, catching up globally. Authors are now able to publish their findings at an astoundingly fast pace such as 10-15 minutes or even at a lesser time. The relative features and merits of the three systems are illustrated in the following figure below:-

Fig 1- Traditional and Upcoming Scholarly Communication Model
IMPORTANCE OF INSTITUTIONAL REPOSITORIES

Institutional repository is the marquee of an institution in the world, where institution displays its worthwhile research programmes, projects, and initiatives to the broad spectrum of audience in the world. An institution outreaches its findings that in turn encourage other institutions and organizations to collaborate and to share their knowledge, expertise and skills. Institutional repositories offer seamless access to documents and reflect past and present research interests of the institution as well as its future research goals. It makes the publications more usable by contemporary and future scholars as well as other professionals like policy makers and social workers. The pace of scholarly communication would be highly accelerated if the IR holds research papers, research reports, etc as soon they are made public. This also have publications in receiving more citations, since the research findings are quickly available to the fellow scholars. The IR can be used throughout the institution and collaborative institutions. Some institutional repositories in India are only providing access to metadata to the external communities who are accessing their repository through internet, whereas internal members who are accessing their repositories through intranet are getting full text information besides metadata. These restrictions exist due to various reasons involving copyright issues, bandwidth issues, permissions from the grant providing agencies (GPA’s), and so on. In India there are number of research grant providing agencies. Sometimes some areas may overlap with each other. Using institutional repositories, the GPA’s can evaluate the novelty of a research proposal and come to know whether any study has been already undertaken in a particular area or discipline.
OBJECTIVES

Following are the objectives of an Institutional Repository:-

1. To create global visibility for an institution's scholarly research.
2. To collect content in a single location.
3. To provide access to institutional research output by self-archiving it.
4. To store and preserve others institution’s digital assets.

CHARACTERISTICS

Gibbons identifies the five core features of an Institutional Repository:-

1. It is institutionally defined – unlike a subject repository; the Institutional Repository captures only the intellectual output of the host institution.
2. Content may be purely scholarly, or may comprise administrative, teaching, and research materials, both published and unpublished.
3. It is cumulative and perpetual. Once items are submitted, they should not be withdrawn. This carries with it a long-term obligation on the host institution to preserve Institutional Repository content.
4. It is open and interoperable – a primary goal of an Institutional Repository is to disseminate the institution’s intellectual output.
5. In collecting, storing, and disseminating information, it contributes to the process of scholarly communication.

FUNCTIONS

Gibbons lists the six core functions of Institutional Repository:-

1. Material submission
2. Metadata application
3. Access control
4. Discovery support
5. Distribution
6. Preservation
CONTENTS
An Institutional Repository may contain a variety of material produced by the researchers of the institution like-
1. Pre print of articles or research reports submitted for publishing the text of journals articles accepted for publication.
2. Revised text of published work with comments from academic readers.
3. Conference papers.
4. Teaching material.
5. Student’s projects.
6. Doctoral thesis and dissertations
7. Database resulting from research projects.
8. Committee papers, administrative papers.
9. Computer software work of art.
10. Photographs and video recordings.

BENEFITS
The main primary advantages appear to include:-

(a) For users
1. Expansion of the range of knowledge that can be shared.
2. Opportunities to simplify and extend dissemination.

(b) For institution
1. Enabling of intellectual property rights to be exploited more effectively at institution level.
2. Leverage of existing investment in information and content management systems.
3. The highlighting of the quality of intellectual capital.

(c) For all
1. Opportunities of new forms of scholarly communication.
2. Flexible way to develop existing scholarly communication.
DRAWBACKS

Following are the drawbacks of IR:-

1. They affect the balance of institutional power as some departments proceed faster than the others.
2. They rely on unproven methods for long term digital preservation.
3. They may need quick wins to sustain institutional support.
4. Initial costs may be high as contributors perceive high risks and duplicate effort to reduce them.

TECHNICAL ASPECTS OF AN INSTITUTIONAL REPOSITORY

With regards to technical aspects of institutional repository development, specifically the software and hardware used to run repositories, there is a surfeit of information (McKay 2007). There is a tradition in the literature of repository managers publishing case studies of their institutional repository deployment (Barwick 2007, Bevan 2007) including technical requirements of software, hardware and support. The software user groups peer support through email lists and wikis (DSpace 2008, Eprints 2008).

The general consensus gleaned from the literature is that there are two key software platforms, DSpace, and Eprints, and several other lesser-used platforms. Discrimination between the two key platforms tends to be on grounds of preference or existing technical abilities (University of Bath 2008). In addition there are organisations that provide a managed solution based on the open source software for a fee, for example Open Repository (Open Repository 2008), and Digital Commons (Berkley Electronic Press 2008) which can also offer additional related services, for example faculty liaison materials to encourage deposit. These managed services remove or minimise the need for in-house technical expertise when deploying a new institutional repository. This in effect
means that the technical aspects of institutional repository development and deployment that dominated the experiences presented in the early literature have now been essentially sidelined, and creating a new repository is, from a technical point of view, reasonably straightforward. This increased ease of implementation is reflected particularly in newer studies looking at institutional repository 'ecology' or "the interactions between repositories and other systems, processes, and people" (Robertson 2008), which is reflected in the focus of this investigation. However, it is worth noting that institutional repositories are regarded as still in their infancy, both as a technology and a resource (Aschenbrenner 2008), and so further changes of both the software and role of the repository is likely in the future, perhaps leading to further specialisation of repository functions and fracturing of the definitions of institutional repositories. Information regarding the evaluation of the performance of institutional repositories will therefore be necessary to ensure they are able to achieve the roles and targets set for them.

SOFTWARE FOR ACADEMIC INSTITUTIONAL REPOSITORY
There are many world renowned open source software used to create repositories are EPrints, DSpace, FEDORA, CDSware etc. They are issued either under GNU public license or the BSD license and can be downloaded from their own sites or open source software directories such as SourceForge. Each of the software has a host of features, unique facilities and excellent capabilities, which the users could explore and experiment.

(a) Greenstone digital library software
The New Zealand Digital library Project at the University of Waikato produces greenstone digital library software. This project is a research program aiming to develop the underlying technology for digital libraries
and make it available publicly so that other can use it to create their own collections. The main architects of the software are Roger McNab and Stefan Boddie. Greenstone is a suite of software for building and distributing the digital library collection. It organizes the collection digitally for publishing it on the internet or on CD-ROM. This software is developed and distributed in cooperation with UNESCO and the human info NGO. It is a open source software available from http://www.Greenstone.org under the term of the GNU General public license.

(b) GNU E-Print Archiving software (Version 2.2.1)
E-Print software is developed as a part of digital library project at University of Southampton, UK. It is available free under the term of GNU General public license. It runs under the Linux and creates online archive libraries of electronic prints. The default configuration creates a research paper archive, but could not be modified and use for other purpose. The document can be stored in any format and each individual research paper (e-print) can be stored in more than one format. The eprint can be submitted through powerful www-based interface. The data integrity checks are performed.

(c) Dspace
DSpace is a platform that allows you to capture items in any format – in text, video, audio, and data. It distributes it over the web. It indexes your work, so users can search and retrieve your items. It preserves your digital work over the long term. DSpace provides a way to manage your research materials and publications in a professionally maintained repository to give them greater visibility and accessibility over time. DSpace is typically used in an institutional repository. It has two main roles:

- Facilitate the capture and ingest of materials, including metadata about the materials
What are the benefits of using DSpace?

- Facilitate easy access to the materials, both by listing and searching
- Getting your research results out quickly, to a worldwide audience
- Reaching a worldwide audience through exposure to search engines such as Google
- Storing reusable teaching materials that you can use with course management systems
- Archiving and distributing material you would currently put on your personal website
- Storing examples of students' projects (with the students' permission)
- Showcasing students' theses (again with permission)
- Keeping track of your own publications/bibliography
- Having a persistent network identifier for your works, that never changes or breaks.
- No more page charges for images. You can point to your images persistent identifiers in your published articles.

EVALUATING INSTITUTIONAL REPOSITORIES

Performance indicators are used to evaluate how well an organisation or project is meeting its expected targets. Ideally, standard performance indicators will be used across related organisations. However, in new areas of practice there may not be enough evidence to validate the use of a particular set of indicators or tools. This is certainly the case with institutional repositories (Kim and Kim 2006). This lack of common methodology for evaluation is reflected in institutional repository literature aimed at suggesting and evaluating methods of evaluation for Institutional repositories. In particular, Westell (2006) proposes a series
of qualitative measures designed to evaluate different areas of institutional repository implementation that have been based on Canadian institutional repositories. Fuhr (2007) note three kinds of evaluation; formative, carried out in parallel with development, summative, carried out after an initial release, and comparative, whereby systems and components are evaluated against each other. Although not implicit in many investigations, the literature relating to established institutional repositories is largely comparative.

Thomas and MacDonald (2007) summarise a number of both qualitative and quantitative measures proposed in the literature, before outlining a framework of performance indicators for different institutional repository functions (i.e., inputs, outputs and impact). In particular, criticism is levelled at a 'bean-counter' outlook on evaluation, where quantitative evaluation methods are used without critique. Particularly, attention is drawn to Carr and Brody's (2007) investigation of a 'sustainable deposit' profile, confirming that in assessing performance indicators more attention is paid to authors/depositors than information seekers. In Thomas and MacDonald (2008), they go on to discuss the possible future evaluative measurements of institutional repositories, suggesting that usage and impact will be important evaluative factors. However, no suitable tools to achieve such an evaluation are presented.

In particular, as part of the web they are tightly linked with search engine technologies. Markland (2006), which looks at how available institutional repository articles are via Google, and case studies such as Organ (2006), that states Google as being identified as the primary access and referral point for an institutional repository, re-enforce the importance of search engines to repositories, and emphasize that institutional repositories are a web-embedded technology. So we can see that search engines are important points of discovery for institutional repositories, and an
understanding of how search engines direct users to repositories is useful. In contrast, the email survey component of Zuccala (2006) reports the majority of respondents claimed to discover the institutional repository via colleagues, and a negligible number via search engines. This may point to a lack of institutional repository impact when users are being referred from a search engine.

In summary, evaluation of institutional repositories is currently not standardised, and is generally comparative. It is recognised that repositories are exclusively accessed online, and so an examination of evaluation using internet relevant methodology is appropriate.

CONSTRAINTS OF ACADEMIC INSTITUTIONAL REPOSITORIES

The Open Access Movement has a big challenge especially the publishing industry of journals under commercial sector challenging their sovereignty. The management support, availability of IR expertise, willingness of authors to participate are very important factors for the success and sustainability of IR.

- Absence of a well defined institutional policy is a serious constraint for IR development in academic libraries. Uncertainty will exist about the norms to be adopted for inclusion of documents regarding the person depositing the document, the need for review and technical evaluation of the document, types of documents to be included and the level access control.

- IR being a new development in academic libraries, there is serious lack of IR expertise especially in a developing country like India. Many institutions although serious to set up IR failed due to non-availability of IR expertise from both library and IT staff.
• The management and the authors concerned about forms a serious bottleneck in building the content of an IR. Many institutions fail to allocate sufficient funds for IR. The basic necessities like IR infrastructure availability of expertise can not be fulfilled without adequate funds.

• Another important constraint is apathy of authors towards time consuming and lengthy deposition procedure.

• Ignorance of users in the absence of appropriate literacy program is another constraint with viz. one cannot expect any developments in IR.

• In case of journals and conference proceedings usually copyright of a research publication lies with the publishers. The publisher’s rigid attitude for allowing the published item in IR and the authors concerned in this matter is another constraint to be sorted out appropriately.

• A good number of institutions in India although have set up the IRs, but made them available only on the LAN of their institute or on a single system due to various reasons like copyright problem from publishers or reservation of their management to throw open their publications. Apathy of Creators/authors for depositing content.

• They affect the balance of institutional power as some departments proceed faster than others.

• Diversity of content and the language used in the full texts.
References


http://iam.inflibnet.ac.in:8080/dxml/bitstream/handle/1944/43.pdf?sequence

- Hirwade, Mangla and Hirwade, Anil (2006), op.cit, p.147


Yeates, Robin, op.cit. p. 97-98.


