CHAPTER 3

HISTORY AND GROWTH OF INSTITUTIONAL REPOSITORIES

3.0 INTRODUCTION

The concept of Open access can be found in the distant past, from the very beginnings of publishing, re-emerging with every innovation in publishing technology. The printing press allowed the written word to be printed and distributed, thereby extending literacy to the population at large. Moving from vellum to paper made it possible to print more cheaply. The invention of the postal system provided a means of widespread distribution. The modern Open access movement traces its history in 1960s, but became much more prominent in the 1990s with the advent of the Digital Age. With the spread of the Internet and the ability to copy and distribute electronic data at no cost, the arguments for Open access gained new importance.

The last three decades have been boom years for academic scholars, who have been producing an increasing amount of published academic research output year-on-year (Branin, 1998). It estimates that the number of scientific papers published annually, has been doubling every 10 to 15 years for the last two centuries (Odlyzko 1995). With the growing internationalisation of research, scholars needs to be able to access research wherever they are in the world and whenever they need it . Rapid advances in technology have opened the door for new methods of accessing and publishing research. Academic output is also linked to the Research Assessment Exercise that occurs in Indian Universities, which allocates government funding to Universities according to the quality of their research. With the dramatic rise in scholarly output, academic libraries have been finding it harder and harder to fund all the purchase of publications that would satisfy their researchers' needs. Even some of the world class academic and R&D libraries only have access to 10-20% of the relevant published papers (Derek 2007). This problem of rising journal costs is often termed the "serials crisis".

3.1 OPEN ACCESS

Researchers send their articles to academic publishers who edit and package the articles and publish them as journals, selling them at a high charge to the academic libraries. Publishers ask authors to assign their copyright over to them and then the publishers have overall change of the research output and its dissemination and accessibility. Researchers often lose out in this situation, having to pay for access to their own research output. There is also the fact that much of the research at Universities is funded by the public, and the public currently do not have access to the research findings when they are published (Allen 2005). Scholarly communication is changing; there has been a move over recent years towards increased diversity in the location of research activities and greater emphasis on multidisciplinary and collaborative research and communication (Houghton, Steele and Henty, 2004). Academic institutions and their libraries need to develop to support their researchers and their changing scholarly practices. The Open access movement started in response to these problems; its aims are to provide free, unrestricted access to research output for all (Swan 2006).

To further the development of knowledge, scholars require access to relevant scholarly literature. Increasingly, this literature is interdisciplinary, global, expensive, digital, and hidden behind technical walls to comply with license restrictions. It is also burgeoning. Little wonder that even scholars at the richest universities in the world have difficulty accessing the specialized literature that they need, while those at the poorest barely have any access at all. What can be done? The Open access movement believes it has an answer to this critical question. Many of its prominent figures have little or no interest in reforming the existing scholarly communication system. Rather, they are interested in transforming it so that it can function effectively in the rapidly changing technological environment.

Open access is well supported by many academic institutions who believe that the increased access to the research will promote the visibility and impact of the institution. Open access takes advantage of the opportunities provided by the Internet and the digital world, provides a solution to the “serials crisis” by releasing

the publisher's strong-hold on academic output, and supports efficient and effective research by increasing accessibility to scholarly information.

3.1.2 Open access: History and Concepts

An explosion of interest and activity in Open access journals has occurred since the 1990s, largely due to the widespread availability of Internet access. It is now possible to publish a scholarly article and also make it instantly accessible anywhere in the world where there are computers and Internet connections. The fixed cost of producing the article is separable from the minimal marginal cost of the online distribution. These new possibilities emerged at a time when the traditional, print-based scholarly journals system was in a crisis. The number of journals and articles produced has been increasing at a steady rate; however the average cost per journal has been rising at a rate far above inflation for decades, and budgets at academic libraries have remained fairly static. The result was decreased access - ironically, just when technology has made almost unlimited access a very real possibility, for the first time.

3.1.3 Historical Perspective

OA was physically and economically impossible in the age of print, even if the copyright holder wanted it. Prices were not only unavoidable for print journals; they were even affordable until the 1970's, when they began to rise faster than inflation. Journal subscription prices have risen nearly four times faster than inflation since 1986. Fortuitously, just as journal prices were becoming unbearable, the internet emerged to offer an alternative.

1997 The Association of Research Libraries developed the Scholarly Publishing and Academic Resources Coalition (SPARC), an alliance of academic and research libraries and other organizations, to address the crisis and develop and promote alternatives, such as Open access.

1980 The first online-only, free-access journals (eventually to be called "Open access journals") began appearing.

1991 The first free scientific online archive was arXiv.org, started.

1997 U.S. National Library of Medicine (NLM) made Medline, the most comprehensive index to medical literature on the planet, freely available in the form of PubMed.

1999 The Journal of Medical Internet Research (JMIR), one of the first Open access journals in medicine was created. The American Scientist Open access Forum was. OAI-PMH protocol for metadata harvesting was launched in order to make online archives interoperable.

2000 BioMed Central, a for-profit Open access publisher, was launched by the then Current Science Group (the founder of the Current Opinion series, and now known as the Science Navigation Group.

2001 34,000 scholars around the world signed "An Open Letter to Scientific Publishers", calling for "the establishment of an online public library that would provide the full contents of the published record of research and scholarly discourse in medicine and the life sciences in a freely accessible, fully searchable, interlinked form".

2002 The first major international statement on Open access was the Budapest Open access Initiative in February 2002, launched by the Open Society Institute . This provided a definition of Open access, and has a growing list of signatories. Two further statements followed: the Bethesda Statement on Open access Publishing in June 2003 and the

2003 Berlin Declaration on Open access to Knowledge in the Sciences and Humanities and the World Summit on the Information Society included Open access in its Declaration of Principles and Plan of Action.

2004 IFLA Statement on Open access to Scholarly Literature and Research Documentation. This statement was adopted by the IFLA Governing on December 5, 2003 but not published until February 2004.

2005 Several U.S and International Library associations released their Library related Principles for the International Development Agenda of the World Intellectual Property Organization.

2006 Federal Research Public access Act was introduced in US Congress by senators John Cornyn and Joe Lieberman. The act continues to be brought up every year since then, but has never made it past committee. The idea of mandating self-archiving was mooted at least as early as 1998.

2007 India's National Centre for science Information launched CASSIR (Cross Archive Serach Services of Indian Repositories). At the time of launch, CASSIR indexed 15 of India's OA, OAI-compimant Repositories and was working to index the rest.

2008 ENCES (European Network for Copyright in support of Education and Science) an organization dedicated to the adoption of OA-Friendly copyright laws in Europe, officially launched.

2009 UNESCO, the Library of Congress and dozens of other partners launched the OA World Digital Library.

3.1.4 Definitions

There are a variety of definitions of "Open access," and the concept is still evolving; however, several key documents, which build upon each other, collectively comprise the best current definition of this term. In December 2001, the Open Society Institute convened a meeting of prominent scholarly communication change agents in Budapest that strongly influenced the nascent Open access movement. The result of this meeting was the "Budapest Open access Initiative" (BOAI). Its definition of Open access (OA), while refined by subsequent documents, remains the most influential one to this day: The literature that should be freely accessible online is that which scholars give to the world without expectation of payment. Primarily, this category encompasses their peer-reviewed journal articles, but it also

includes any unreviewed preprints that they might wish to put online for comment or to alert colleagues to important research findings.

There are many degrees and kinds of wider and easier access to this literature. By "Open access" to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited. To achieve Open access to scholarly journal literature, we recommend two complementary strategies.

OA serves the interests of many groups.

Authors: OA gives them a worldwide audience larger than that of any subscription-based journal, no matter how prestigious or popular, and demonstrably increases the visibility and impact of their work.

Readers: OA gives them barrier-free access to the literature they need for their research, unconstrained by the budgets of the libraries where they may have access privileges. OA increases reader reach and retrieval power. OA also gives barrier-free access to the software they use in their research. Free online literature is free online data for software that facilitates full-text searching, indexing, mining, summarizing, translating, querying, linking, recommending, alerting, "mash-ups" and other forms of processing and analysis.

Teachers and students: OA puts rich and poor on an equal footing for these key resources and eliminates the need for payments or permissions to reproduce and distribute content.

Libraries: OA solves the pricing crisis for scholarly journals. It also solves permission crisis. OA also serves library interests in other, indirect ways. Librarians

want to help users find the information they need, regardless of the budget-enforced limits on the library's own collection. Academic librarians want to help faculty increase their audience and impact, and help the university raise its research profile.

Universities: OA increases the visibility of their faculty and research, reduces their expenses for journals, and advances their mission to share knowledge.

Journals and publishers: OA makes their articles more visible, discoverable, retrievable, and useful. If a journal is OA, then it can use this superior visibility to attract submissions and advertising, not to mention readers and citations. If a subscription-based journal provides OA to some of its content (e.g. selected articles in each issue, all back issues after a certain period, etc.), then it can use its increased visibility to attract all the same benefits plus subscriptions. If a journal permits OA through post print archiving, then it has an edge in attracting authors over journals that do not permit post print archiving. Of course subscription-based journals and their publishers have countervailing interests as well and often resist or oppose OA. But it oversimplifies the situation to think that all their interests pull against OA.

Funding agencies: OA increases the return on their investment in research, making the results of the funded research more widely available, more discoverable, more retrievable, and more useful. When funding agencies disburse public funds, OA helps in a second way as well, by providing fundamental fairness to taxpayers or public access to the results of publicly-funded research.

Governments: As funders of research, governments benefit from OA in all the ways that funding agencies do (see previous entry). OA also promotes democracy by sharing non-classified government information as widely as possible.

People: OA gives them access to peer-reviewed research, most of which is unavailable in public libraries, and gives them access to the research for which they have already paid through their taxes. But even those with no interest in reading this literature for them will benefit indirectly because researchers will benefit directly. OA accelerates not only research but the translation of research into new medicines,

3.1.5 Salient features of Open access

The salient features of Open access are

- Open access literature is digital, free of charge and free of copyright
- OA is compatible with copyright, peer review, revenue, print, preservation, prestige, career advancement, indexing and supportive services associated with conventional scholarly literature
- OA campaign focuses on the literature that authors give to the world without expectation of payment
- OA literature is not free to produce or publish
- OA is compatible with peer review and all the major OA initiative for scientific and scholarly literature insist on its importance

3.1.6 Vehicles of Open access

There are two primary vehicles for delivering Open access to research literature viz.

- Open access Journals
- Open access Archives

3.1.6.1 Open access Journals

OA Publishing is just like any other journal publishing. Like traditional publishing, it involves peer reviewing of submitted articles from authors and publishing. Published content is freely accessible over Internet and the users have right to download, use and further distribute it with proper attribution. In traditional publishing model, it is the “end-user” that pays to access the paper. Open access journals cost money to produce and distribute, especially since they are peer reviewed and edited like conventional journals. Various funding strategies are in use like, direct author fees, Institutional memberships to sponsor all or part of author

fees, funding agency payment of author fees, grants to Open access publishers and Institutional subsidies.

3.1.6.2 Open access Archives

OA Self-Archiving model is liberal on peer review. It simply provides persistent digital Repository where authors / owners of the content may archive their documents (Pre-referred or post-referred). Self archiving can be achieved in three ways viz.

- Articles on author web sites
- Depositing articles in disciplinary archives
- Depositing articles in Institutional archives and Repositories

3.2 OAI-PMH (Open Archives Imitative Protocol for Metadata Harvesting)

Open access works are scattered across many disciplinary archives, Institutional e-print archives, and Institutional Repositories and Open access journals. Therefore, it is difficult for scholars to locate all needed works on a particular subject. One important international movement to solve this problem is the Open Archives Initiative (OAI), which aims to develop and promote the use of a standard protocol, known as the Open Archives Metadata Harvesting Protocol (OAMHP), designed for better sharing and retrieval of e-prints residing in distributed archives.

3.2.1 Brief History of OAI-PMH

The OAI-PMH has its technical roots in the Universal Preprint Service (UPS) and Dienst protocol. In turn, Dienst is based on the Kahn-Wilensky Framework (KWF). Thus, KWF led to Dienst, Dienst to UPS and UPS to OAI-PMH. In the late 1999, a meeting was convened in Santa Fe, New Mexico to identify the key issues preventing the implementation of services such as linking and searching across large, diverse, distributed E-print archives. Attendees of Santa Fe Convention developed a consensus to adopt a UPS Prototype-based metadata harvesting model as a workable technical and organizational framework for delivering digital archive content and services to end users. The harvesting model allowed “E-print

(content) providers to expose their metadata via an Open interface, with the intent that this metadata be used as the basis for value-added service development".

Meeting participants also agreed upon the basic definitions, concepts, technical components and organizational aspects of interoperable E-print archives. Theses agreements became known as the "Santa Fe Convention". Shortly after the meeting in Santa Fe, members of UPS changed the name to the Open Archives Initiative (OAI) to refer to the overall group of people and its philosophy, and named the protocol itself, the "OAI-PMH". The authors did not plan to make changes to the protocol version 1.0 for a period of 12-18 months after the initial release, but they adopted the newly released World Wide Web Consortium (W3C) XML Standards, and upgraded the OAI-PMH in July 2001.

Nelson and Warner (2002) defined an OAI-PMH record and the process for a Service Provider to obtain it from Data Provider as "metadata expressed in a single format. A record is returned in an XML encoded byte stream in response to an OAI-PMH request for metadata from an item".

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) provides an application independent interoperability framework based on metadata harvesting. Metadata harvesting is a formalized framework for the co-ordinate exchange of metadata in distributed and decentralized electronic information environments. It enables automated distribution of any kind of metadata, which may be aggregated into searchable databases by "harvesting" systems. It is independent of both the types of content offered and the economic mechanisms surrounding that content, and promises to have much broader relevance in Opening up access to a range of digital materials. The OAI technical framework is not intended to replace other interoperability standards (for example Z39.50) but to provide an easy-to-implement and easy-to-deploy alternative for different constituencies or different purposes than those addressed by existing interoperability solutions.

3.2.2 OAI-PMH Concepts

1. Harvester: a client application that make a OAI-PMH request.
2. Repository: network accessible server able to process a OAI-PMH request.
3. Resource: the stuff the metadata is about.
4. Item: a constituent of a Repository, conceptually, it is the container of the metadata.
5. Identifier: Unique identifier that unambiguously identifies an item within a Repository.
6. Record: an XML-encoded set of metadata expressed in a specific format.
7. Date stamp: date of creation / modification / deletion of a record.
8. Set: optional construction for grouping items in the purpose of selective harvesting.

3.2.3 OAI-PMH Requests

The OAI-PMH is based on HTTP (Hypertext Transfer Protocol). Request arguments are issued as GET Or POST Parameters. OAI-PMH supports six request types.

- Identify: retrieve Repository information.
- List Metadata Formats: what metadata formats in Repository.
- List Sets: retrieve Repository set structure.
- Get Record: retrieves a single metadata record.
- List Records: harvest records from a Repository.
- List Identifiers: harvest record headers only.

3.2.3 OAI-PMH Responses

Responses are encoded in XML syntax. OAI-PMH supports any metadata format encoded in XML. Dublin

Core in the minimal format specified for basic interoperability.

- General Information
- Metadata formats
- Set structure
- Record identifier
- Metadata

3.3 OPEN ACCESS JOURNALS

Open access Journals have the following characteristics:

(1) Scholarly materials,
(2) Utilize quality control mechanisms like those of conventional journals
(3) Digital in nature
(4) Freely available,
(5) Allows authors to retain their copyrights, and
(6) Use Creative Commons or similar licenses

There is some dispute as to whether Open access journals must utilize peer review as a quality control mechanism. Most do, but there are also some high quality journals that don't and meet all other criteria yet have great impact on their fields of study. D-Lib Magazine is an example of such a journal. Likewise, the question of whether the journal must use a Creative Commons or similar license is another area of dispute. This dispute reflects the deeper, fundamental question of whether "Open access" is just free access or free access plus a set of specified use rights that go significantly beyond normal copyright rights.

3.3.1 Types of Open access Journal Publishers

The major types of Open access journal publishers are: (1) born-OA publishers, (2) conventional publishers, and (3) non-traditional publishers. The same disclaimers apply to this taxonomy as were indicated for the self archiving one. Let’s examine these types of Open access journal publishers in more detail:

3.3.1.1 Born-OA Publishers

With the establishment of the Open access journal publisher BioMed Central in 2000, a new type of journal publishing venture was created “born-OA" publisher. These digital commercial or nonprofit publishers were established for the sole purpose of publishing Open access journals, and they typically utilize the Creative Commons Attribution License for their publications. Authors usually retain their copyrights. Different funding strategies are employed by these publishers, including advertising, author fees, grants to the publishers, library membership fees and supplemental products.

3.3.1.2 Conventional Publishers

As the Open access movement has gained momentum, conventional commercial and nonprofit journal publishers have begun to experiment with Open access publishing programs or to establish permanent Open access programs.

3.3.1.3Non-Traditional Publishers.

During the late 1980s and early 1990s, the Internet had developed to the point that scholars began to publish free digital journals utilizing existing Institutional infrastructure. These journals were not intended to generate income; they were "no-profit" journals. Although many of these journals allowed authors to retain their copyrights and they had liberal copyright statements regarding non commercial use, they preceded by a decade or more the Creative Commons, and, consequently, did not embody that kind of copyright stance. While some of these journals ceased publication and others were transformed into non-profit ventures, they provided a model that others followed, especially after the popularization of the Internet began in the mid-1990s, which followed the earlier introduction of Web browsers. In recent years, the availability of free Open source journal management and publishing systems, such as the Open Journal Systems, further simplified and streamlined

digital journal publishing, fuelling additional growth in this area. Now, a wide variety of academic departments or schools, institutes and research centres, libraries, professional associations, scholars, and others publish digital journals, a subset of which comply with the strictest definition of an Open access journal and a larger subset which comply with the looser definition of an Open access journal as a free journal. Since these diverse "publishers" would have been unlikely to be engaged in this activity without facilitating digital technologies and tools, refers to them as "non-traditional publishers." Many of them are also "no-profit" publishers as well.

3.3 DEVELOPMENT OF OPEN ACCESS: Indian Scenario

In the scholarly publishing scenario, India has its unique position. India does not have a high percentage of its scientific journals available online. Though it is placed at the 12th position for overall number of journals among the top 25 publishing countries, its position falls down to 18th for journals with online content (Haider, 2005). Surprisingly, its position in the list of Open access (OA) journals is fifth, well ahead of countries such as the Netherlands, China, Germany, Australia, and so on, which are higher in the list of online journals. Among the non high-income countries, India ranks second only to Brazil for the number of OA journals. Almost 50% (48 out of 103) of the online journals from India are OA (Haider 2005). What makes India do so well in the list of OA journals? How are such a large percentage of electronic journals in India able to provide OA without even charging the author or authors' institution for publication of the articles? On the other hand, in the Registry of Open access Repositories, India ranks 11th in the list of countries with registered interoperable archives

3.4.1 OA journal publishing in India

A large number of the journals published from India belong to learned societies and associations, and are published by the association or the editor themselves without the involvement of any commercial publisher. The members of these learned societies receive the print copies of the journals without paying an annual or recurring fee. For continuing their publication activities the associations

depend on non-member subscriptions, which are limited in number and restricted by and large to the Indian universities and colleges, on advertisements in print editions, and on income generated from other sources such as the annual conferences of the associations. Most of the Indian journals suffer from 'low circulation - low visibility, low impact factor' syndrome. With many fewer paid regional or international subscriptions, these journals have limited visibility, restricted mainly to the members of the association. With this limited visibility, these journals are cited less frequently than their western counterparts. The low impact factor inhibits authors from submitting their quality work to the Indian journals. Thus, it is expected that with OA, Indian journals will be able to reach to a wider audience. At the same time, loss, if any, of paid non-member subscriptions is less likely to have a major effect on the economics of these journals.

A number of biomedical journals have been online from late 1990s; Neurology India, Current Science, Indian Paediatrics, and the Indian Journal of Critical Care Medicine have been online since 1998–99. These individual endeavours have now evolved into more organised and collaborative efforts, including those of the Indian Medlars Centre, the Indian Academy of Sciences, and the Indian National Science Academy under the not-for-profit sector, and Medknow Publications in the commercial segment. International players such as Bioline International have also helped many Indian journals to have web presence.

3.4.1.1 Indian Medlars Centre

The Indian Medlars Centre (IMC), set up jointly by the National Informatics Centre and the Indian Council of Medical Research, has taken the pioneering step of putting Indian biomedical journals accessible from a single platform. IMC's first bibliographic database IndMed, established in 1998, provides abstract level information from more than 70 journals. Each of the articles in the database is tagged with Medical Subject Headings. In 2003, IMC launched its full-text database, MedInd, which now hosts full text version of 38 journals in PDF format. The understanding between the

journals and the IMC ensures that the digitised work hosted by IMC will continue to be accessible even if the journal discontinues providing newer issues.

3.4.1.2 Indian Academy of Sciences

The Indian Academy of Sciences (IAS), founded in 1934, publishes 11 journals with the basic philosophy that no journal published by the Academy is in competition with another journal published in the country. Current Science, published by the Current Science Association in collaboration with the IAS, has entire back volumes from 1932 online in PDF format and has been online since 1999. Many other journals including the Journal of Biosciences, Sadhana, and Pramana also have the entire back volumes online.

3.4.1.3 Indian National Science Academy

The Indian National Science Academy (INSA), established in 1935, publishes four journals including the Proceedings of INSA. Under the project 'Building Digital Resources: Creating Facilities at INSA for hosting S & T Journals on Online', INSA launched the Open access version of these journals in December 2003.INSA publishes 4 peer-reviewed journals, organizes scientific discussions and brings out proceedings and monographs. The e-journals@insa is a project of the Indian National Science Academy that was started in July 2002. Initially this was a NISSAT supported project entitled "Building Digital Resources: Creating Facilities at INSA for hosting S&T Journals on Online", to facilitate conversion of INSA journals from print to digital format and host these materials online. The initial objectives of the above project were:

- ✓ To build up a national digital resource base with global visibility and accessibility.
- ✓ To bring out electronic versions of INSA journals and to make them available on a web server.

This web portal provides access to current and back volume full-text literature of INSA journals. This project became one of the much-acclaimed Open access initiatives in India, supporting the concept of free access to scientific literature. All INSA journals are Open access and full-text is available as PDF files from the common journal gateway. The common journal gateway is freely accessible if users register their names in the website. This portal provides a search interface for advanced search, where search query can be given with any or more than one metadata elements, such as keyword, article title, and name of the author, journal volume-issue-initial page number, and journal year. The combination of search elements is made possible by using Boolean operators. The recent journal issues as well as back volumes of INSA journals can be browsed from the main page of this portal. The search result displays a list of articles that match the search query; when a user selects a reference from the list it leads to the particular journal issue that published the particular reference, the PDF version of article appears on the screen and can be downloaded to the user workstation.

3.4.1.4 Indian Journals.com

IndianJournals.com provides single window access to multidisciplinary Indian journals published by different scholarly societies and institutions. It provides access to eleven Open access journals and periodicals. This journal gateway also provides access to subscription-based content. These Open access scholarly journals mainly belong to the science, technology and medicine (STM) areas. These journals are Open Archives Initiative (OAI)-compliant. This common journal gateway is freely accessible if the users register their names in the website. These journals are searchable from the main page of the gateway, which also maintains an archive of back volumes.

3.4.1.5The Kamla-Raj Enterprises

The Kamla-Raj Enterprises is a Delhi-based publisher established in 1933. Kamla-Raj publishes seven print-based peer-review scholarly journals mainly in the areas of social sciences which are also available in electronic format on Open

access. These journals are OAI-compliant. The publisher maintains an archive of each of these Open access journals starting from volume one.

3.4.1.6 Bioline International

Bioline International is a not-for-profit collaborative effort of the University of Toronto Libraries, Canada, and the Reference Center on Environmental Information, Brazil, and Bioline, UK. It provides electronic publishing services to journals published in developing countries. Bioline provides access to 14 Indian journals on their primary site as well as archives these journals at the Bioline EPrints Archive.

3.4.1.7 MEDIND@NIC

The Indian MEDLARS Centre at the National Informatics Centre has initiated two unique projects with support from the Indian Council of Medical Research. The first one is INDMED@NIC that indexes 70 prominent biomedical journals of India from 1985 onwards. This INDMED bibliographic database is available online. Another project, MEDIND@NIC is an Open access initiative from NIC that provides Open access to the full-text content of 38 Indian biomedical journals. MEDIND@NIC aims at providing online access to full-text Indian biomedical periodicals to the users within and outside India. Different publishers, mainly learned societies in the respective specialized areas, publish these journals in print-on-paper format. Initially, MEDIND@NIC started digitization of the contents of back volumes of the journals. But now they are getting born-digital content of current issues from the publishers. Some of these publishers also provide full-text access separately from their own websites. The metadata of these journal articles are also maintained and searchable in the INDMED database. Current issues of all journals are available from this single gateway.

3.4.1.8 Medknow Publications

Medknow Publications Private Limited is a publisher of high-quality peer-reviewed scholarly Open access journals in India. Medknow publishes, maintains

and hosts 48 peer-reviewed scholarly journals, mainly in the biomedical subject areas. Medknow also provides solutions to the scientific societies and scholarly institutions, through Journal-on-Web a web-based manuscript submission and peer review system that handles pre-publication and post publication processes for journal issues. It also accepts online submission of manuscripts through dedicated websites for respective journals. These journals are OAI-compliant. Metadata harvesters, search engines and indexing services index and harvest metadata of the current issues of the journals from the individual journal websites

3.5 IMPACT OF OPEN ACCESS PUBLISHING

OA has certainly helped the Indian journals to reach an international audience, as could be seen by the number and distribution of article downloads. The Journal of Postgraduate Medicine, a quarterly journal with a print circulation of less than 1,000, attracts close to 100,000 visitors with more than 110,000 article downloads per month. The increased accessibility and visibility has also increased the citations received by this journal (Bavdekar and Sahu, 2005). The number of manuscripts submitted to the journals has increased many folds, with increases in the number of articles coming from other countries ranging from 12–44% for various journals. Interestingly, these OA journals have not lost the paid non-member subscriptions but, in fact, have benefited from increased subscriptions, including many international subscriptions. As the cost of putting a website is miniscule compared with the cost of sending printed copies free to the hundreds of members, the online publication has not had any impact on the economics of these journals. In addition, helping hands from MedInd and Bioline have helped these journals to take care of the expenses for the OA version.

3.6 METADATA HARVESTING SERVICES: Indian Scenario

3.6.1 Cross Archive Search Service for Indian Repositories (CASSIR)

Project initiated by the National Centre for Science Information (NCSI) of the Indian Institute of Science (IISc) and supported by DSIR. This web-based

search and browse service is a part of the ongoing project entitled 'Development of OAI-Based Institutional Research Repository Services in India'. In this service, Open Archives Initiative-Protocol for Metadata Harvesting (OAI-PMH) compliant software (PKP Harvester) harvests metadata from the registered Open access Repositories in India. At present, CASSIR covers eighteen Indian Open access Repositories:

- Catalysis Database (ePrints@NCCR IIT Madras)
- DRS at National Institute of Oceanography
- DSpace at ICFAI BUSINESS SCHOOL (IBS), Ahmedabad
- DSpace at IIMK
- DSpace at National Chemical Laboratory, Pune, India
- DSpace at Vidyanidhi
- DSpace@DRTC
- Dspace@NITR
- ePrints@IISc
- ePrints@IIT Delhi
- ETD @ Indian Institute of Science
- Indian Institute of Astrophysics
- DSpace at ISI Library, Bangalore
- NAL Institutional Repository
- OneWorld South Asia Open Archive Initiative
- OpenMED@NIC
- RRI Digital Repository
- DU eprints archive

These Open access Repositories store a number of scholarly publications in digital format. Thus, CASSIR serves a functionality of cross-searching of Repositories. This cross-searching functionality is much needed in a country like India where hundreds of Open access Repositories are expected to be proliferated in near future, if the UGC (Submission of Metadata and Full-text of Doctoral Theses in Electronic Format) Regulations implemented soon.

3.6.2 Open Index Initiative (OII)

The Open Index Initiative (OII) is a collaborative effort of more than sixteen volunteers working in Indian social science libraries across more than five Indian cities, with a basic objective to develop an exhaustive online database with index and abstract to journal articles, book reviews, conference papers, and working papers published in Indian social science journals or published by Indian social science institutions. The Open Index Initiative (OII) portal facilitates information on:

- ✓ Forthcoming Conferences/Seminars/Workshops in India in Social Sciences and related areas
- ✓ Indian Social Science Associations and Professional Bodies
- ✓ Search for Book Reviews
- ✓ Index to Articles of Economic and Political Weekly (EPW, 1992-2005)
- ✓ Index to Journal Articles
- ✓ Open Source Software and Experts
- ✓ Online Directory of Social Science Institutions
- ✓ Theses and Dissertations
- ✓ Union Catalogue of Journals
- ✓ Working Papers Online
- ✓ IGIDR OPAC (Online Public access Catalogue)

The volunteers in this initiative contribute their data on the scholarly resources available with them. The web-based data entry module facilitates recording of data, related to:

- ✓ Experts in Open Source Software
- ✓ Index to Journal Contents
- ✓ Update the Online Directory of Libraries
- ✓ Statistical and Government Publications
- ✓ Theses and Dissertations
- ✓ Union Catalogue of Journals
- ✓ Working Papers

OII is a public-access database that promotes scholarly literature in social sciences, produced by Indian social science researchers/institutions or published in Indian social science journals. It follows the human-driven metadata submission model, in contrast to the software-driven metadata extraction model. Thus, the OII system gets a wide variety of data sets, which is unparallel to the metadata harvesting services established elsewhere.

3.6.3 Open J-Gate

On February 27, 2006, Informatics India Limited launched Open J-Gate, a searchable portal of Open-access journals that provides access to 4,300 Open access journals and million plus records of articles. Open J-Gate is the first corporate initiative in India to promote and support Open access initiatives. The bibliographic database is maintained in a well-index database, having the metadata such as, the title of the article, name of authors, author email-id, and author's affiliation, keyword, abstract and full-text links to each article. It has Quick Search, Advanced Search, and Browse by journal options. In Advanced Search option, search can be limited to a few subject categories or sub-categories, journal type, a few metadata (such as title, keyword, abstract, author, institution/address, or, all). When search results are displayed, each retrieved record provides details of the article and links to full-text article at publisher sites. In this service, full-text links are regularly validated. Open J-Gate follows three-level hierarchical system, where first consists of seven main subject categories, such as:

- Agricultural and Biological Sciences
- Arts and Humanities
- Basic Sciences
- Biomedical Sciences
- Engineering and Technology
- Library and Information Sciences
- Social and Management Sciences

Open J-Gate covers Open access peer-reviewed journals as well as professional and industry journals. Although this service has worldwide coverage, mainly Open access journals published in English language are indexed in this online database.

3.6.4 Scientific Journal Publishing in India (SJPI)

The Scientific Journal Publishing in India (SJPI) is a project funded by international development agencies, having as project goal to improve the accessibility of scientific literature published in Indian journals by introducing an indexing system. This is a kind of demonstration of a system that would be self-sustaining under the existing infrastructure to support Indian scientific journals and academic publishing. The project popularizes to editors and administrators of Indian scientific journals and academic publishing, a Free and Open Source Software (FOSS), named Open Journals Systems (OJS) developed by UBC Public Knowledge Project. OJS assists with every stage of the refereed publishing process, from submissions to online publication and indexing. Additionally, OJS helps OAI-PMH compliant meta data harvesters to index/harvest the contents of Open access journal issues. The SJPI is implemented in two phases, where each phase has some specific objectives, in which major objectives are;

- To help Indian Journals to go online with the Journal Management Software
- To support Hindi language to the OJS system
- To add new features to the OJS system, e.g. Latex support, Classification Scheme
- To conduct another workshop on Open access Journal Publishing

In the first phase, a Prototype Journal System was set up using OJS software, where following journals participated:

- Bulletin of Materials Science
- Current Science
- Journal of Astrophysics and Astronomy
- Journal of Biosciences

- Journal of Chemical Sciences
- Journal of Genetics
- Journal of Indian Institute of Science
- Pramana: Journal of Physics
- Proceedings - Earth and Planetary Sciences
- Proceedings - Mathematical Sciences
- Resonance - Journal of Science Education
- Sadhana (proceedings in engineering sciences)
- SRELS Journal of Information Management (SJIM)

The SJPI Cross Journals Search Service is part of the SJPI project. In this service, OAI-PMH compliant software (PKP Harvester) harvests metadata from the sample full-text contents of participating Indian scholarly journals. This demonstrates a single point search service across multiple scholarly journals. The SJPI Cross Journals Search Service covers Indian journals listed above. Although this SJPI Cross Journals Search Service is a prototype system and not updated regularly, this is still available online.

3.6.5 Search Digital Library (SDL)

The Search Digital Libraries (SDL) is a metadata harvesting service for Open access Repositories in the area of library and information science, a project initiated by the Documentation Research and Training Centre (DRTC) of the Indian Statistical Institute (ISI Bangalore). In this service, OAIPMH compliant software (PKP Harvester) harvests metadata from the registered Open access Repositories in the subject area of library and information science. At present, SDL covers nine Open access Repositories, including one from India:

- Australian Library and Information Association e-prints
- CALTECHLIB
- CCSD: Sciences de l'Information et de la Communication, France
- DLIST, University of Arizona

- E-LIS: E-Prints in Library and Information Science
- Journal of Technology Management & Innovation
- Librarians' Digital Library (LDL)
- OCLC Research Publications
- University of North Carolina, USA

3.6.6 SEED (Search Engine for Engineering Digital-repositories)

The IIT Delhi has developed a number of discipline-specific Research Support Tools (RST), which accompanies individual research studies indexed from e-journal and conference paper websites covering a wide range of disciplines. The RST utilizes the study's metadata to search relevant open-access databases for related studies, theory, news, policies, and other resources, as well as offering access to the study's metadata and citation, to a personal portfolio, and to email and comment options. The software used is Public Knowledge Project (PKP) system. Simple and advanced search facility along with browsing capabilities is available.

3.7 INSTITUTIONAL REPOSITORIES

In developing countries, academic librarians have begun experimenting with Open source software in the creation of Institutional Repository (IR) systems in local libraries, using mainly Greenstone, Fedora, DSpace, or Eprints. These are some of the available Open access Repository software which has stabilized over the past years. (Aschenbrenner et al, 2008). The Directory of Open access Repositories (DOAR, http://www.Opendoar.org) and Registry of Open access Repositories (ROAR, http://roar.eprints.org/) list 21 Asian countries that have initiated Open access in their countries. Among these the most active is Japan with 90 Open access Repositories, followed by India (67), Taiwan (61) and Turkey (31).

3.7.1 Institutional Repository: Definitions and concepts

A simple definition of IR is a Web-based archive (Johnson 2002) or digital collection (Crow, 2002) of scholarly materials produced by members of the

institution. Since the Repository relies on input from members of the institution, the commitment and participation of contributors, users and managers is crucial. In recent years there have been many studies that show that uptake of IR in universities is slower than it was hoped for (Zuber, 2009; McKay, 2007; Kim, 2006; Whitehead, 2005; Callan, 2004). Most authors are still unfamiliar with Open access and making their works available on IR (Suber in Dillon, 2008; Foster & Gibbons, 2005), some are not even aware that their institution has an IR (Kim, 2006). Besides the authors, there are also the information seekers or end-users of IR who access the system to search for information for their teaching, learning and research needs. These users are particularly important as they can be the spokesperson for the IR. Satisfied users, who successfully use the IR, will value the service and help promote it to others within the institution or his own research community. However these end-users of IR are particularly under-studied (McKay 2007).

According to Heery & Anderson (2005) Institutional Repositories:

- ✓ Contains content, deposited by owner, creator, or third party;
- ✓ Repository architecture manages content as well as metadata;
- ✓ Repository offers a minimum set of basic services, e.g. put, get, search, access control;
- ✓ Repository must be sustainable and trusted, well-supported and well-managed;
- ✓ If an Open access Repository, it must also:
- ✓ Provide Open access to its content (notwithstanding legal constraints);
- ✓ Provide Open access to its metadata for harvesting.

3.7.2 Open access Institutional Repositories

The concept of Open access is clearly defined and understood in the literature. It emerged in response to the restrictive access to knowledge in scholarly and scientific journals imposed by commercial publishing houses via subscription fees, license fees or pay-per-view fees (Gideon, 2008). The many definitions revolve

around the fundamental concept that Open access is all about public access without financial, legal or technical barriers (Budapest Open access Initiative, 2002), to information materials including original scientific research results, raw data, metadata and scholarly multimedia (Berlin Declaration on OA Knowledge, 2003), in a cost effective way (ARL, 2004).

3.7.3 Importance of IR

The capture and preservation of a university's intellectual output can serve as a tangible indicator of the institution's quality (Crow, 2002), which in turn contributes to its visibility (Harnad, 2003) prestige and public value (Crow, 2002). Chan (2004) states that as research become more data intensive, a scholar's ability to store, access and share primary data will be crucial to the advancement of scholarship. Universities all over the world have established Institutional Repositories of intellectual materials 'born' within the institution. These IRs not only act as a Repository to published materials but are much more valuable as an access to digital content of local materials. In developing countries mainly, financial resources are a barrier to wider access to published scholarly information (Gideon, 2008). Local intellectual content on the other hand are often found only in printed journals and conference proceedings that are available only at that particular institution where it originates from, either with the author or in the library's collection. This lack of accessibility has somewhat become a barrier to access, use and cite local research. IR has made the often limited dissemination and access of information extremely possible now in developing countries (Ghosh and Anup 2007). Peter Suber, the pioneer of OA expresses his views on Open access movement in an interview with Dillon (2008) cites some of OA benefits as: i) it removes price as an access barrier; ii) it connects authors to readers and iii) for readers, it enlarges the library.

Researchers

- ✓ Journals are the primary research publication channel.
- ✓ Journal publishing is dominated by commercial ventures.
- ✓ Researchers write papers for journals (free or page charges!).

- ✓ Researchers transfer copyright to publishers (free).
- ✓ Researchers on Editorial Board (free).
- ✓ Researchers review papers (free).

Libraries

- Cannot satisfy the information needs of their users.
- Libraries pay huge subscriptions to publishers to access the paper (and electronic) and universities pay more than once: subscription, photocopying license and for study packs.
- Or possibly they cannot afford the subscription.
- Dissatisfaction with the current scholarly communication model.
- Even the wealthiest institution cannot purchase access to all the information that all of its researchers require.
- Site-licenses and consortia deals have helped, but mainly in the richest countries; though good examples of deals for developing countries (INASP).
- Many commercial publishers charge extra for online access – so causing more pressure on budgets.

Authors

- ➢ Their work is not seen by all their peers – they do not get the recognition they desire.
- ➢ Despite subscriptions, they often have to pay page charges, colour figure charges, reprint charges, etc.
- ➢ Often the rights they have given up in exchange for publication mean there are things that they cannot do with their own work.
- ➢ Readers: They cannot view all the research literature they need – they are less effective.
- ➢ Society: We all lose out if the communication channels are not optimal.

3.7.4 Essence of Institutional Repositories

The challenge for digital preservation is not just the volume of data. The hardware and software used to store and access digital information are constantly upgraded and superseded. Technology obsolescence is generally regarded as the greatest technical threat to ensuring continued access to digital material. The speed of changes in technology means that the timeframe during which preservation action must be taken is very much shorter than for paper, often measured in just a few years. Digitization projects in libraries seem ubiquitous as libraries become increasingly involved in the acquisition, development, and management of digital information and libraries typically target archival and special collections materials. Projects to digitize vast collections of books began as early as 1971 with Project Gutenberg and are now getting widespread media attention with the launch of Google Book Search, the Internet Archive, and others (Coyle 2006). The broad view of Institutional Repositories as a means to manage and preserve effectively an Institution's knowledge base and intellectual assets results in the content of Institutional Repositories expanding beyond e-prints to include research data, e-learning materials and other forms of Institutional intellectual outputs, which are generally not published or preserved elsewhere. Researchers, students, staff and institutions will require ongoing availability and confidence in the future accessibility of the content within the Repositories. Those running Institutional Repositories, therefore, naturally have the responsibility to ensure this for the content they are entrusted with managing by their institutions and researchers. It needs to be ensured that content within the Repositories remains accessible and retains its authenticity, reliability and integrity for as long as it is needed. As Lynch (2004) has rightly pointed out, "An Institutional Repository needs to be a service with continuity behind it … Institutions need to recognize that they are making commitments for the long term".

Institutional Repositories are a new but important area within the educational landscape. Through free and unrestricted online availability, they make it easier for researchers to disseminate and share research outputs and thus support the Open

access goal of scholarly communication. As noted by Scholarly Publishing and Academic Resources Coalition (SPARC), Institutional Repositories are becoming a major component of the evolving structure of scholarly communication (Crow, 2002). Institutional Repositories are now being created to manage, preserve, and maintain the digital assets, intellectual output, and histories of institutions. Librarians are taking leadership roles in planning and building these Repositories, fulfilling their roles as experts in collecting, describing, preserving, and providing stewardship for documents and digital information. Repositories provide services to faculty, researchers, and administrators who want to archive research, historic, and creative materials. Thus, development of Institutional Repository has become a necessity to reveal the scientific research output for which Library and Information professionals have to take keen interest and initiation.

3.7.5 Institutional Repository Policy

Repositories now represent potentially rich sources of information, data, images, and valuable research results. The movement is new and the time it takes to plan, formulate policies, and bring Institutional communities to consensus can make it a slow process. Each institution defines its own policies dealing with access to and use of materials in Repositories. Not all materials can be made available freely. Copyrighted materials may carry a variety of restrictions. Nonexclusive publisher licenses would increase availability to these materials and place the publishers in the Open access arena. Some publishers permit authors to self-archive. Other publishers opt for exclusive licenses for a limited time, while still others will not allow any deviation from exclusive copyright. Some materials may be restricted to a small group of researchers or to people associated with the institution because they represent work in progress deemed proprietary or that may entail sponsor restrictions.

The key issues to be considered while developing Institutional Repositories are Institutional culture and consensus has to be discussed

- The scope of the Repository as to type of publications to be part of Institutional Repository.
- The information content of the Repository.
- Access levels as to full text or abstract level/ access to type of users or privileges.
- Legal aspects i.e. copyright acts.
- Standards
- Sustainability and maintenance

The appropriate software and funding has to be taken into account before going for Institutional Repository projects at institution level. Institutional culture depends on how the organization is structured as well as how much collaboration and trust exists within an institution. In academic organizations, faculty belongs to departments, disciplines, and research groups. Academic competition may be fiercer in some universities than in corporations. In an internally competitive environment where cooperation and trust are not nurtured, building a Repository will become more difficult. Faculty will not contribute willingly to a central Repository unless they have been consulted and trust the process. Faculty need to be convinced that contributing to a Repository will enhance their reputations in their disciplines and result in wider dissemination of their work.

3.7.6 Software for Institutional Repositories

The two leading software packages, DSpace (MIT) and EPrints (Southampton) are both available free under Open source licenses, and there are at least half a dozen other possible packages. In theory, commercial document management or knowledge management software packages might also be suitable but are unlikely to be adopted given their costs.

There is a Guide to Institutional Repository Software available at the Budapest Open Archive Initiative website, which includes a feature-set comparison of DSpace, EPrints, DCSware, i-Tor and MyCoRe. There are a handful of Open-source packages for creating and maintaining such archives. The four most important are

eprints (from Southampton University), DSpace (from MIT), CDSWare (from CERN), and FEDORA (from Cornell and U. of Virginia). In addition to this there is few more software, which is being used for archiving. Researcher has listed them with brief description, these are:

3.7.6.1 ARNO

The ARNO project (Academic Research in the Netherlands Online) has developed software to support the implementation of Institutional Repositories and link them to distributed Repositories. The project is funded by IWI (Dutch acronym for: Innovation in Scientific Information Supply). Project participants are the University of Amsterdam, Tilburg University and the University of Twente. The ARNO system was released for public use in December 2003. It has been in use at the universities of Tilburg, Amsterdam, Rotterdam, Twente and Maastricht.

3.7.6.2 CERN Document Server Software (CDSware)

The CERN Document Server Software (CDSware) was developed to support the CERN Document Server. The software is maintained and made publicly available by CERN and supports electronic preprintservers, online library catalogs, and other web-based document depository systems. CERN uses CDSware to manage over 450 collections of data, comprising over 620,000 bibliographic records and250,000 full-text documents, including preprints, journal articles, books, and photographs. CDSwarewas built to handle very large Repositories holding disparate types of materials, including multimedia content catalogs, museum object descriptions, confidential and public sets of documents, etc.

3.7.6.3 DSpace

The DSpace software has been purpose built in collaboration between Hewlett Packard and MIT to offer IR services. It is specifically designed to manage diverse heterogeneous types of digital content. It offers interoperability via OAI-MHP (Open Archive Initiative – Metadata Harvesting Protocol – a software standard that

allows specialised search engines to gather article metadata from compliant websites) and built-in support for Dublin Core metadata (Dublin Core is an agreed metadata standard used in library cataloguing and elsewhere, though other metadata schemes are possible). It uses persistent identifiers via the CNRI Handle system.

3.7.6.4 Eprints

The Eprints software has the largest and most broadly distributed installed base of any of the Repository software systems described here. Developed at the University of Southampton, the first version of the system was publicly released in late 2000. The project was originally sponsored by CogPrints, but is now supported by JISC as part of the Open Citation Project and by NSF. Eprints worldwide installed base affords an extensive support network for new implementations. The size of the installed base for Eprints suggests that an institution can get it up and running relatively quickly and with a minimum of technical expertise. The number of Eprints installations that have augmented the system's baseline capabilities for example, by integrating advanced search, extended metadata and other feature indicates that the system can be readily modified to meet local requirements.

3.7.6.5 Fedora

The Fedora digital object Repository management system is based on the Flexible Extensible Digital Object and Repository Architecture (Fedora). Fedora is one of a number of repository architectures that have been proposed over recent years for use in digital libraries. The Digital Library Research and Development Department at the University of Virginia Library interpreted the architecture and built a prototype that demonstrated the feasibility of the architecture with their large and diverse digital collections.

In September 2001 the University of Virginia received a grant of US$ 1,000,000 from the Andrew W Mellon Foundation to enable the library, in collaboration with Cornell University, to build a sophisticated digital object repository system based on the Flexible Extensible Digital Object and Repository Architecture (FEDORA).

Functionality of FEDORA

- Open Source
- Web Services
- Flexible Digital Object Model
- Data streams
- Default Disseminator
- Extensible Disseminators
- Content Versioning
- XML Ingest and Export
- Digital Object Storage
- access Control and Authentications
- Searching
- OAI Metadata Harvesting Provider
- Fedora Administrator Client
- Migration Utility

Fedora is evolving software, where rapid development is going on. It looks very promising for the access and preservation of the digital content. The emerging like Meatadata Encoding and Transmission Standards (METS) for better visibility, OAIS for preservation and OAI-PMH for the harvesting of metadata are implemented completely. It can be implemented very easily with some modifications in both windows as well as Linux environments.

3.7.6.6 i-TOR

I-Tor (Tools and technologies for Open Repositories) was developed by the Innovative Technology- Applied (IT-A) section of Netherlands Institute for Scientific Information Services (Dutch acronym: NIWI). NIWI calls i-TOR “a web technology by which various types of information can be presented through a web interface,” irrespective of where the data is stored or the format in which it is stored. It aims to implement a “data independent” Repository, where the content and the user-

interface function as two independent parts of the system. In essence, i- Tor acts as both an OAI service provider, able to harvest OAI compatible Repositories and other databases, and an OAI data provider. Because i-Tor is able to publish data from a variety of relational databases, file systems, and websites, the system allows institution considerable latitude in the way it organizes its Repository.

3.7.6.7MyCoRe

MyCoRe grew out of the MILESS Project of the University of Essen. The MyCoRe system is now being developed by a consortium of universities to provide a core bundle of software tools to support digital libraries and archiving solutions. The bundle is designed to be configurable and adaptable to local requirements without the need for local programming efforts. In contrast to MILESS, which provides a hard-coded Qualified Dublin Core data model, the MyCoRe data model is completely configurable. Further, MyCoRe provides a sample application, based upon a "core" of functionality, which shows users how to build their own applications using metadata configuration files.

3.7.6.8 DARE

Digital Academic Repositories (DARE) is a collective initiative by the Dutch universities to make all their research results digitally accessible. It can be seen as a national level, albeit federally structured, Repository. Its programme of research projects is broadly similar to the JISC FAIR programme. The project was awarded 2 million Euros for the period 2003-2006 by the Dutch government. All participating institutions will adopt the same standards, while retaining their own responsibility in setting up and maintaining their own Repositories.

3.7.6.9 Greenstone Digital Library Software (GSDL)

The Greenstone Digital Library Software is internationally renowned Open source software for developing Institutional Repositories. It is promoted by the New Zealand Digital Library project research group at the University of Waikato, headed by Dr. I.H.Witten. The software is issued under the terms of GNU General Public

License. Greenstone is supported by the UNESCO and Human Info NGO Belgium for spreading the benefits of this software to developing countries. It provides a way of building, maintaining and disseminating digital library collections, Opening up new possibilities for organizing information.

Greenstone builds collections using popular and standard digital formats such as HTML, XML, PDF and other formats which include multimedia files. It is provided with effective full text searching and metadata based browsing facilities that are attractive and easy to use. It runs on a wide variety of platforms such as Windows, Linux and provides indexing, searching, browsing, metadata extraction and is OAI compliant.

3.8 INSTITUTIONAL REPOSITOTY PROJECTS

3.8.1 DSpace Project

The DSpace project at MIT was funded by Hewlett Packard to the tune to $1.8m, plus 3 FTE HP staff and $400k in systems equipment. The DSpace Federation is a group of institutions that are using DSpace software to build their own Repositories. The DSpace Federation Project is a one-year study which will begin the process of building a collaborative federation of institutions running DSpace. This group will test the adaptability of the system to a targeted group of institutions with varied needs. Federation Project members include Cambridge, Columbia, Cornell, Massachusetts Institute of Technology, Ohio State, Rochester, Toronto, and Washington.

DSpace was completed in November 2002 through a joint effort between Hewlett-Packard Labs (HP) and the Massachusetts Institute of Technology (MIT), who have released the resulting code under an open-source license, specifically the permissive BSD license (Smith, 2003). This means that end-users can adjust, modify or improve the code as they see fit, and furthermore the project developers do evaluate and reincorporate any improvements made by users into the main distribution (Smith, 2003). As of this writing the software is hosted on the open-source repository

Sourceforge which currently offers version 1.4 of the software, indicating the project is beyond beta testing ready for end-users. DSpace Federation's informal list has over 100 institutions using DSpace.

The project was designed to be a tool for institutions, in MIT's case a university, to implement a central location where faculty, departments, disciplines, labs and research centres could store their published and pre-published research for access by others and long-term archivisation. The developers claim that the software was build to support "every function that a research organization needs to run a production digital repository service, but as simply as possible" (Smith, 2003). Furthermore, the software was designed to be multidisciplinary: it is designed around the idea of the "Community," which designs its own work flows and manages its own deposits, which we will examine under "Usage and Institutional Policy." Communities can be any size, from labs to departments to entire institutes of research (Smith, 2003).

3.8.2 RoMEO Project

The RoMEO Project (Rights MEtadata for Open archiving) was a JISC funded project at the University of Loughborough to investigate the rights issues surrounding the self-archiving of research in the UK community under the Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH). Through surveying the academic community it ascertained how give-away research literature and metadata was used and how it should be protected. From this work, the RoMEO project created a list of publishers' conditions for self-archiving. After negotiation with the RoMEO Project team and JISC, SHERPA developed the outcomes of this project into a database-driven searchable service and knowledge bank of information. SHERPA maintains the SHERPA/RoMEO listing, which details the rights given to authors by the major publishers of peer-reviewed academic journals. It is now possible to search for many publishers and find out what permissions are normally given as part of each copyright transfer agreement.

The majority of publishers support the right of academic authors to mount their own work online: however, some prohibit authors from using their work in this

way as a condition of their copyright transfer agreement which they ask the author to sign. The project made a detailed analysis of journal publishers' copyright assignment terms – a useful database of these is maintained on the RoMEO website. Around 90% of publishers ask for copyright, 6% for exclusive and 4% for non-exclusive licences, and 75% of authors are asked to warrant that their work has not previously been published. The project also surveyed OAI data and service providers, revealing a degree of ignorance and/or unconcern with rights issues. Only 25% of data providers had licence agreements with their depositing authors, and 50% either just trusted the depositors, or simply provided a general warning statement. Regarding metadata protection, 50% of data providers thought (incorrectly) that metadata records were facts and as such had no copyright. Also 68% believed that though there was database right in metadata collections, this was "implicitly waived" within the OAI community.

3.8.3 SHERPA Project

The SHERPA project (Securing a Hybrid Environment for Research Preservation and access) has been set up to encourage change in the scholarly communication process by creating Open-access Institutional e-print Repositories for the dissemination of research findings. The outcomes of the project will be advice on the building and maintenance of IRs, guidelines on IPR and copyright issues, and advocacy material to publicise an institution's Repository (Hubbard 2003; Pinfield 2003). It aims to set up OAI-compliant e-print Repositories at each of the partner sites. The project aims are:

- ✓ To set up Open access e-print Repositories which comply with the Open Archives Initiative (OAI) Protocol for Metadata Harvesting (OAI PMH) using EPrints.org software;
- ✓ To investigate key issues in creating, populating and maintaining e-print collections, including: Intellectual Property Rights (IPR), quality control, collection development policies, business models, scholarly communication cultures, and Institutional strategies;

- ✓ To work with OAI Service Providers to achieve acceptable standards for the effective dissemination of the content;
- ✓ To investigate digital preservation of e-prints using the Open Archival Information System (OAIS) Reference Model (an ISO standard for the long-term preservation of digital information, initially developed by RLG).

3.8.4 TARDis Project

The TARDis project (Targeting Academic Research for Dissemination and Disclosure), run by Southampton University, is planning to develop a multidisciplinary Institutional e-print archive and assess and evaluate the activity within a library-led infrastructure. It is designed to tackle head-on the major problem faced by IRs, namely the lack of participation by faculty: "TARDIS will investigate and report on strategies to overcome the technical, cultural and academic barriers, which currently restrict the development and particularly the acquisition of content of Institutional e-Print archives. It will develop a working model of a multidisciplinary Institutional archive." The project runs from August 2002 until July 2004. The TARDis Project has now come to an end. There will be no further major updates to this site but it will be maintained until at least 2008. All project materials will be deposited into the University of Southampton Research Repository. The TARDis Project has now evolved into a University funded service, maintained by the Southampton University Library in partnership with the Information Systems Service and the School of Electronic and Computer Science.

3.8.5 DAEDALUS

DAEDALUS is a project concerned with the establishment of a range of OAI-PMH-compliant digital collections at the University of Glasgow. These will include e-prints (both published and peer reviewed academic papers, and pre-prints and grey literature), theses, resource-finding aids and Institutional documents. It runs until July 2005.

3.8.6 Caltech

Caltech's CODA (Collection of Open Digital Archives) Repository was established in 2001. It currently consists of some 11 archives with a further six listed as in development. The system allows document counts and a wide range of access statistics to be viewed by any user. The following table lists the active archives, their numbers of records and the total number of document accesses since launch. The vast majority of the content appears to be grey literature: theses & dissertations, technical reports and conference presentations.

3.8.7 eScholarship

The California Digital Library (CDL) eScholarship Repository, announced in April 2002, illustrates the continuum between digital libraries broadly conceived and Institutional Repositories. The CDL launched the eScholarship Repository, a web site and a suite of digital support services, to distribute academic research and working papers of University of California faculty. eScholarship uses the OAI metadata harvesting protocol to provide interoperability.

The CDL initiative includes a suite of digital services to store and disseminate faculty research in digital formats. The CDL system uses the web-based bepress (vendor) system to manage paper submission, processing, and dissemination. Additionally, the system also supports a topical alerting service that alerts users to new content in their specified areas of interest.

There are about 2450 e-prints (both pre-prints and post-prints), mostly in the social sciences, mainly economics and related areas (e.g. transportation). The stated policy is to accept "journals, peer-reviewed series, working papers, discussion papers series, and other electronic forms of scholarship". Most of the documents we viewed were working papers. Content is stored as PDF only (but the system will accept and convert Word, RTF etc. into PDF).

3.8.8 CARL

The CARL Institutional Repositories Pilot Project is an initiative to implement Institutional Repositories at several Canadian research libraries. The project, which is spearheaded by the Canadian Association of Research Libraries, was launched in September 2002 and has 12 libraries participating. The Repositories will be searchable using one interface and freely accessible to anyone with an Internet connection. The ultimate vision is to have a number of robust and interoperable archives containing Canadian scholarly output that will form a part of a larger global system of Repositories.

In the initial phase of the project, participants are sharing best practices and lessons learned in order to assess the feasibility of IRs in the Canadian context. Members are experimenting on a trial basis with a variety of software types, content, and archiving policies, among other things. CARL has published a Position Statement on Institutional Repositories (CARL 2003). It sees the benefits of (Canadian) IRs as:

- ✓ increasing the visibility of Canadian researchers and institutions;
- ✓ increasing the accessibility and impact of Canadian research domestically and internationally;
- ✓ long-term preservation of research output of Canadian academic institutions;
- ✓ increasing the proportion amount and diversity of scholarly output that is collected and preserved (cf. traditional collections policies focussed on published materials);
- ✓ Facilitating more timely access to research and scholarship.

3.8.9 Utrecht

DISPUTE, the Institutional Repository of the University of Utrecht was originally scheduled for release at the end of 2002. The site is currently (in early 2004) in a preliminary state and states variously that the final site will be available in October 2003 (sic) or by the end of 2003 (sic). All this suggests the going has not

been as easy as was originally envisaged. At present there appears to be about 400-500 maths and 40-50 physics e-prints, plus some other documents and theses. Several of the e-prints we viewed were scanned from published literature (e.g. *Physical Review*). The site also appears more a portal to a disaggregated collection of resources than a single IR.

3.8.10 ARNO

The Academic Research in the Netherlands Online (ARNO) project, run from September 2000 until September 2002, sought to design and implement university digital archive servers to preserve the academic output (including research reports, pre-prints, theses and dissertations, and articles published in regular scholarly journals) of member institutions. The project's goal was to make the Repository freely accessible via OAI interoperability standards. The project was being implemented by the library staffs of the University of Twente, the University of Amsterdam, and Tilburg University.

Specific project goals included:

- ✓ Connecting the document servers to international distributed digital archives and to the Dutch national information infrastructure;
- ✓ Developing an infrastructure that will couple with the production processes of scientific publishers and offer a good basis for handling peer review.
- ✓ Connecting seamlessly to digital learning environments.

3.9 INSTITUTIONAL REPOSITORIES IN INDIA

Institutional Repositories have been developed by some of the major institutes of India with a purpose for

- ➢ Self-archiving: A researcher wants a place to put working papers and similar documents.

- Developing national level, domain-specific or omnibus harvesting services
- Preservation of information Repository content for perpetual access
- Integration of Institutional Repositories with other Institutional e-resources and e-services
- Learning objects includes instructional materials.
- E-learning: An instructor wants to store problems for a web based home work system ion a Repository. The problems themselves are expressed in a HTML language that requires external application to render. The instructor would like to be able to efficiently search for problems render appropriately when delivered to students' browsers.
- Repository management: Operators of the Repositories listed in these use case need to be able to manage these facilities and the content. The Repository manager then needs to develop tools to deal with a large-scale format migration. Repositories need to provide reporting facilities and interfaces that will support these activities.

Presently, there are 67 IR in India from different academic institutions such as research & development, management, universities etc. It indicates that India is the 2nd position in Asia after Japan (90). So, India is competing with other developed countries across the world regarding this matter. India is moving towards Open access movement in the developing countries since last decade by establishing a number of Open access Repositories, embracing Free and Open Source Software (FOSS). The Indian information professionals experiment with the Open source software for establishing Institutional Repository (IR) in a local library, such as Greenstone, DSpace and EPrints. If an IR is successfully implemented in the local library setup, it then scales up to the institution-wide application through campus-wide network or intranet. Likely, it turns Open to the wider audience with the implementation of Open access Institutional Repository, when the authority of institution convinced. With the availability of the dedicated information infrastructure combined with broadband connectivity and national educational grid, some national institutions and universities implemented Institutional Repositories for wide

dissemination of scholarly literature emanated from the respective institutions..(http://www.doar.org)

Some Institutional Repositories in India are specially established to diffuse intellectual outputs of the country in the form of electronic theses. Vidyanidhi and ETD@IISc are examples of such kind. Other Institutional Repositories provide all kind of scholarly materials such as research papers, conference papers, presentations, photographs, along with e-theses. The OpenMED and Librarians' Digital Library are examples of such kind.

Table 3.1 List of IRs in India (as on 20.10.2011)

Repository name	Number of Records	publications	Conference Proceedings	Theses	Unpublished Documents	Other Materials	Software Used
Catalysis Database	1478		Yes	Yes			EPrints
CMFRI Digital Repository	3633	Yes	Yes	Yes	Yes	Yes	EPrints
Delhi College of Engineering Repository	326				Yes	Yes	DSpace
Dhananjayarao Gadigil Library	1539	Yes			Yes	Yes	DSpace
Digital Knowledge Repository of Central Drug Research Institute	135				Yes		DSpace
Digital Library at Indian Statistical Institute, Bangalore	191						DSpace
Digital Repository of Cochin University of Science & Technology	995			Yes		Yes	DSpace
DRS at National Institute Of Oceanography	3763		Yes	Yes			DSpace
DSpace @ GGSIPU	133			Yes		Yes	DSpace
dspace @ sdmcet	67	Yes		Yes	Yes	Yes	DSpace
DSpace at CUSAT	3199	Yes	Yes	Yes		Yes	DSpace
DSpace at IBS Ahmedabad	171	Yes	Yes		Yes	Yes	DSpace
DSpace at Indian Institute of Management Kozhikode	290		Yes	Yes	Yes		DSpace
DSpace at National Chemical Laboratory	407			Yes	Yes	Yes	DSpace
DSpace at NCRA	84			Yes	Yes	Yes	DSpace
DSpace at Vidyanidhi	5480			Yes			DSpace
DSpace@IITB	1657		Yes				DSpace
DSpace@INFLIBNET	1144		Yes			Yes	DSpace
Dspace@NITR	653	Yes	Yes	Yes			DSpace
DSpace@PDPU	64				Yes		DSpace

DSpace@TU	938		Yes	Yes			DSpace
DU Eprint Archive	170	Yes	Yes	Yes		Yes	EPrints
eGyankosh	6190					Yes	DSpace
ETD@ Indian Institute of Science	206			Yes			DSpace
Eprint@NML	1419	Yes	Yes	Yes		Yes	EPrints
Eprints@IARI	82		Yes	Yes	Yes		EPrints
EPrints@IITD	2141			Yes			DSpace
ePrints@NII	10				Yes		EPrints
Eprints@SBT MKU	21						EPrints
ICRISAT Open access Repository	3432		Yes			Yes	DSpace
IIT Roorkee Repository	823		Yes			Yes	DSpace
IMSc Eprint Archive	41		Yes			Yes	EPrints
Indian Academy of Sciences: Publications of Fellows	39024						EPrints
Indian Institute of Astrophysics Repository	4211			Yes		Yes	DSpace
Indian Institute of Management Kozhikode Digital Library		Yes				Yes	
Indian Institute of Petroleum Institutional Repository	439						DSpace
Institutional Repository of Intectual Contributions of Delhi Technological University	841			Yes		Yes	DSpace
Kautilya Digital Repository at IGIDR	193		Yes	Yes	Yes		DSpace
Knowledge Repository of Indian Institute of Horticultural Research	160				Yes	Yes	DSpace
Knowledge Repository Open Network	458		Yes	Yes			DSpace
Librarians' Digital Library	249		Yes	Yes		Yes	DSpace
Mahatma Gandhi University Theses Online	913			Yes			Nitya
Management Development Institute - Open access Repository	325	Yes	Yes			Yes	DSpace
National Aerospace Laboratories Institutional Repository	1320		Yes	Yes	Yes	Yes	EPrints
National Science Digital Library	504	Yes					DSpace
NISCAIR Online Periodical Repository	2020					Yes	DSpace
Open access Agricultural Research Repository		Yes	Yes				
Open access Repository of IISc Research Publications	18949	Yes	Yes		Yes	Yes	EPrints
OpenMED@NIC	2412						EPrints
Raman Research Institute Digital Repository	3568				Yes	Yes	DSpace
Sardar Vallabhbai National							