

Opportunities and Challenges of establishing Open Access Repositories: A case study of OpenMED@NIC

Sukhdev Singh, Naina Pandita and Shefali S. Dash.

**Bibliographic Informatics Division,
(Indian Medlars Centre)
National Informatics Centre,
A - Block, CGO Complex, Lodhi Road,
New Delhi-110003. (India).
sukhi@nic.in
Phone: 91-11-24362359**

**Key Words: Open Access; Institutional Repositories; Subject Repositories;
Opportunities and Challenges; Medical and Allied Sciences; Emerging Roles for
Library and Information Science Professionals**

Abstract:

National Informatics Centre had established a subject repository in May 2005. It is meant for Medical and Allied Sciences and named as OpenMED@NIC <<http://openmed.nic.in>>. It has MeSH® based subject categorization and this makes it one of its own kind. Taking OpenMED@NIC as a case – this paper discusses key issues in establishing and maintaining an open access repository. Librarians and information science professionals can play active role in providing access and exposure to quality research and academic content generated in their institutions. Mature and standard open sources softwares are now available for setting up repositories. Libraries can install one of these on existing institutional or library servers to setup repositories. However to ensure better access and faster response time dedicated hardware and reliable connectivity would be required. Librarians and information science professional can play important role in exposing intellectual content produced by their organizations. They can take of various roles like – generating awareness among staff, researchers and students about benefits of self archiving in institutional or subject repositories; training them in uploading their articles and other documents in such repositories; acting as meta-data editors and repositories managers. Establishing a repository, administrating and inviting authors to deposit their articles and other works in it is golden opportunity available to librarians and information science professionals. This opportunity should be grabbed with open hands.

Introduction:

The academic and scientific environment all over the world is changing fast. Internet connectivity in educational and research institutions is more of a norm rather than an exception. In India, Internet connectivity at even at homes is now affordable and has become a symbol social prestige. The work place is now changing with PCs on working tables. At least, libraries now provide Internet workstations to their users. Many professionals might be anxious – will the readers ditch the libraries? The answer might be difficult to get, but as we know ‘Library is a growing organism’. It is evolving and would become more and more paperless in future. It appears that there would more opportunities to librarians as compared to threats in near future. In the online environment of future, libraries will not only ensure access to collective wisdom of human race; they would also ensure that knowledge creation is facilitated, collaborated and disseminated in the best interest of global society. To start with, librarians would have start with their own institutes and organizations. They will have to play proactive role in exposing intellectual wealth of their institutions and organizations. This could be done in number of ways. Building and managing institutional repositories is the most effective and simplest to project their organization's intellectual output. Having such repositories at institutional level would automatically ensure that every document is ‘indexed’ and discoverable. This is due to the fact that most repository softwares are OAI-PMH compliant which would ensure that structured meta data would be available to numerous search engines.

Open Access Repositories:

Open Access (OA) means free and online access to scholarly literature that can be freely disseminated further with proper author attribution. It brings down barriers to the scientific communication by using Internet (Suber, Peter. 2007). Open Access is manifested in two forms – OA Publishing and OA Self-Archiving. OA Publishing is like conventional scholarly publishing involving peer reviewing of submitted articles by authors. The difference being that published content is freely accessible over Internet. Various business models sustain such open access publishing. It could be Government supported or by reimbursement of publication charges by funding or author's employer. Sponsorships and advertisement revenues are also prevalent models. OA Self-Archiving refers to uploading published or pre-published documents in publicly accessible digital repositories. These repositories provide

easy access to its collection and allow other systems to harvest their metadata associated with documents. The exchange of such metadata is in accordance to now well-established “Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH)” protocol (Open Archives Initiative. 2002). Repositories are of two types – institutional and subject oriented. Institutional Repositories hold documents authored by its staff members and students. Subject repositories hold documents pertaining to a particular subject area.

Key Considerations for setting up a Repository:

Interoperability:

Adoption of an interoperable protocol is necessary to expose metadata associated with repository's collection to external systems and search engines. “Open Archives Initiative” has developed such a protocol to facilitate efficient dissemination of repository metadata. This protocol is known as Open Access Initiative Protocol for Metadata Harvesting or simply - OAI-PMH (Open Archives Initiative. 2002). Under this model, metadata is harvested (extracted) from Data Providers (Repositories) by Service Providers (Search Engines).

Categorization Scheme:

Browsing a directory-type structure is a useful arrangement especially for users not looking for a particular item. It groups related items and provide easy navigational facility. Subject categorization is considered the most helpful arrangement for any repository. It would be better if a standard classification scheme is adopted. An institutional repository would also like to have a departmental / school-wise categorization separately or sandwiched into the subject classification. Arrangement by authors and years are some other arrangements that might be considered for repository designers.

Reliability:

Uploading articles in any repository requires extra efforts from depositors. Thus they must be convinced that there are definite benefits in taking the trouble of uploading. They at least want to ensure that repository is fairly reliable and trusted. To entice trust among depositors, servers should be up round the clock with persistent IDs or URLs. They need to be convinced that repositories can play crucial role in exposing their works to their peers. Institutional commitment and self-archiving policies are important in cultivating sense of

trust towards institutional repositories. Reliability of links encourages other authors to cite works from repositories.

User Friendly:

The documents would be most of the time uploaded remotely without any assistance from the repository staff. They need to input metadata along with proper subject headings. This requires an intuitive and user-friendly interface.

Application Software:

Fortunately there are number of softwares available for developing and maintaining repositories. Open Society Institute (2004) has produced a guide for open source repository softwares. It lists Archimede, ARNO, CDSware, DSpace, Eprints, Fedora i-Tor, MyCoRe, and OPUS . All these supports OAI-PMH. According to data culled from ROAR (2008) on 7th February 2008, number of repositories powered by various softwares are given below:

<i>Repository Software</i>	<i>No. of Repositories</i>
DSpace	271
EPrints	243
Bepress	58
OPUS	26
ETD-db	24
DiVA	16
CDSWare	9
Fedora	9
Open Repository	9
HAL	7
ARNO	5
DoKs	5
Fez/Fedora	4
MyCoRe	4
SciX	2
OJS	1
Others	300

It is clear that EPrints and DSpace enjoy majority of the installation base. EPrints has been the pioneer and had largest installation base. Recent years have seen DSpace taking over EPrints in terms of installation base. It could become a difficult decision to pick one of the

these two softwares. Knowledge of back-end technologies, training and availability of support services would ultimately decide would software fits well for a particular repository.

Development and Deployment of OpenMED@NIC:

Bibliographic Informatics Division of National Informatics Centre has vast experience of creating and maintaining medical databases. Some of the well known databases are : IndMED (Bibliographic Database indexing about 77 Indian Biomedical Journals – <http://indmed.nic.in>); medIND (Full-text of 38 Indian Biomedical Journals – <http://medind.nic.in>) and UNcat (Union Catalogue of Journal Holding of Indian Medical Libraries – <http://uncat.nic.in>) (Singh, Sukhdev; Gaba, Surinder Kumar and Pandita, Naina, 2004) In the year 2004 an Open Access Repository for Medical and Allied Sciences (OpenMED@NIC, 2008) was conceptualized. During that time, there were around 300 repositories all over the world as compared to today's one thousand (ROAR, 2008a) known repositories. In medical field there was only one repository - Bioline International. The same however has been decommissioned about two years back (Registry of Open Access Repositories, 2008b). Even Bioline International lacked an in-depth medical subject classification. Thus need for an international repository with in-depth medical classification was always there. In May 2005, NIC developed and deployed [OpenMED@NIC](http://openmed.nic.in) at <http://openmed.nic.in>.

For developing [OpenMED@NIC](http://openmed.nic.in), prototyping model was adopted. An old retired P-II was selected for developing the prototype. It was formatted and RedHat Linux 9 was installed. However, P-II offered limited hardware resources for smooth running of the system. It was reformatted and downgraded to RedHat Version 7.3. Other major softwares installed were Apache 1.3.31, Mod_Perl 1.25 and MySQL 3.23.49. Finally EPrints 2.3.4 was installed after number of attempts of matching and installing correct versions of various required modules. OAI-PMH Version 2 support is inbuilt in EPrints. In EPrints the default subject categorization is based on Library of Congress (LC). However the same can be replaced by a colon separated text file “subjects”. EPrints imports subject scheme during installation with “import_subjects” command. The software generates static pages from its back-end database. These static pages reduce the response time for end users but takes server resources in generating them. It was found that time required for generating pages was directly proportionate to number of subject terms. MEDical Subject Headings(MeSH) was most appropriate classification scheme for OpenMED@NIC. It had around 23,000 terms. Building

such a huge categorization was not justified to start with. It was decided to have just representative broader categorization based on MeSH. Which can be enhanced later on by the administrator tool meant for the purpose. Manually creating EPrints “subjects” file even for broader representative classes was difficult. So, a PERL script was used to extract such a representative scheme from MeSH Tree file. The script created the “subjects” file based on statistical sampling of subject tree depths.

Once prototype was ready; it was used to demonstrate the core features and functions of the proposed archive. Security aspects were also checked. Prototype provided insight in determining the exact technical requirements. Few dummy documents were loaded to test the prototype. Its look and feel was also changed. Support was provided for few additional document types like PPT and PPS.

For production server (accessible to public), a Rack Mountable Server – RS2 (1 U) with 4 GB RAM and dual processors was procured. This was loaded with RedHat Advanced Server (AS 3). For sake of taking periodic back-ups it was deployed under “Storage Area Network (SAN)” in NIC. Some of the installation procedures used in prototype required change due to the change in version of web server from apache 1.3 to 2.0. SMTP gateway was set up for sending emails which was not done in prototype. Cron procedures were set up to perform routine functions like generation of static pages and taking periodic backups automatically. Relevant DNS entry and firewall rules were added in NIC Network to make the repository accessible to all over Internet. Finally OpenMED@NIC was made public in May 2005.

Present Status of OpenMED@NIC:

OpenMED@NIC is a discipline based International Archive. It accepts peer-reviewed documents having relevance to research in Medical and Allied Sciences including Bio-Medical, Medical Informatics, Dental, Nursing and Pharmaceutical Sciences. These could be peer-reviewed preprints, postprints (refereed journal paper) and accepted theses. In case of non-English documents, descriptive data [Author, Title, Source etc.], abstract and keywords must be in English. According to revised (dated 28th March 2006) acceptance policy, publishers and authors may deposit peer reviewed pre-published and post-published work in the OpenMED@NIC archive provided that:

- *PowerPoint slides of conference presentation would be included only if they are of research content/relevance and have been presented at conferences/workshops.*
- *It is authors' responsibility to ensure that the deposited material is written by them. Third party submission would only be accepted after the authors/publishers consent has been taken by the submitter.*
- *Authors are encouraged to use their own final, refereed drafts and not the publisher's PDF (unless the publisher agrees).*
- *There is an option of depositing the full-text and metadata, and setting access to the full-text as "Restricted Access" instead of "Open Access" in case the author has any doubts about his right to place it in Open Access.*
- *Requests for removing a publication, from anyone other than the author, would be re-directed to the authors.*

At present OpenMED@NIC has about 1100 registered user with about 1791 full text document (OpenMED@NIC, 2008). It provides searching in both simple and advanced modes. It can be also be browsed by Year of Publication and Categories Wise. The categories are largely based on MeSH but also includes special categories like Conferences, Institutional and Journals Repositories.

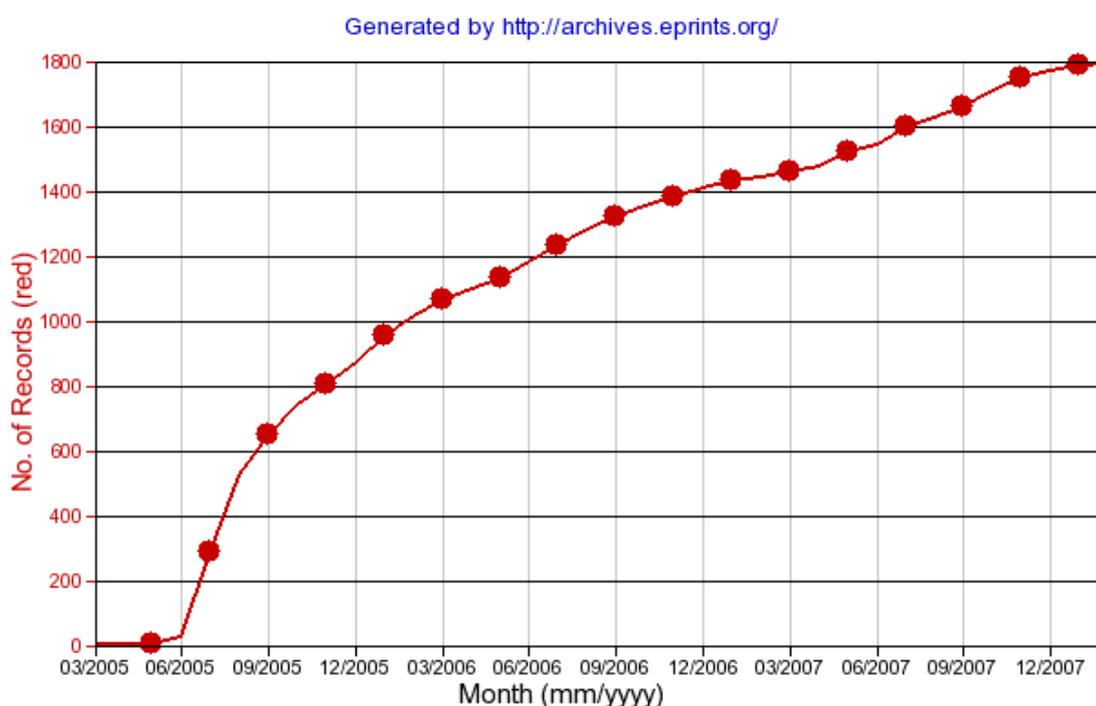
Attracting Submissions to OpenMED@NIC:

Deploying and maintaining a repository is much easier as compared to attracting article in it. It takes lot of efforts in making the content owners and authors aware of Open Access and virtues of self-archiving. They are reluctant to deposit their documents in Open Access repositories (Westrienen, Gerard van and Lynch, Clifford A. 2005). The prime reasons for this could be - confusion, uncertainty and fear on copyright issues; doubts regarding how the material would be used; doubts on getting proper attribution, impact and scholarly credit; myth of low quality material in institutional repositories; unfriendly submission procedures; lack of mandatory provisions to deposit and lack of Internet connectivity.

A series of attempts were made at NIC to spread awareness about Open Access among the bio-medical community. These included writing letters and emails to scientists working in all

various research Institutions and other eminent scientists. Number of emails was sent to various discussion groups. Open Access topic was introduced to participants of various training programmes. Online tutorials (Naina, Pandita and Singh, Sukhdev. 2005) were prepared and archived in the [OpenMED@NIC](#) itself. Open Access topic was included in various NIC's training programmes (Singh, Sukhdev and Pandita, Naina, 2008) related to biomedical information retrieval.

A graph of document submission activity is given below: (Registry of Open Access Repositories, 2008c).



Repositories and Libraries:

It is a common practice to archive a copy of research reports, thesis and dissertations produced by students, faculty and employees in institutional library. Obviously, it is expected from libraries to be the gateway of intellectual content produced by their parent institutions. Thus a library is the most suitable candidate for any organization to develop and maintain an institutional repository. In the initial stages of design and development of repositories, librarians should actively involve themselves at least in the development of categorization scheme and user interface. Once the repository is place, its management and maintenance can

be handled by library staff. Being experts in metadata and bibliographic description, librarians would be best to handle the editing of meta data submitted along with documents. Another very important aspect is generating awareness among faculty, research scholar and student about repositories. Periodical seminars and training programmes could be organized by libraries to make authors self-reliant in uploading their articles and other documents in repositories. At times, librarians depending upon institutional policies, may deposit on behalf of authors. Promoting open access and generating awareness about its benefits among scientists, researchers and scholarly authors remains a major challenge. OpenMED@NIC along with other repositories can helping in better access and dissemination of scholarly content. Such repositories are playing important role in the emergence of new culture of conducting and reporting research among academic and scientific community.

Conclusion:

Building up a repository is a daunting task. It requires meticulous planning of various processes and resources along with dedicated hardware, software, competent human resources and Internet connectivity with high bandwidth around the clock. Once the repository is established, winning trust of content owners for populating repository is another major challenge. It requires spreading awareness among academic and scientific community about various benefits of open access self-archiving. Collaborative efforts are needed to develop a new culture of conducting research and disseminating its results. Libraries, librarians and other information science professionals have an important role to play in this. They can establish, manage, promote and populate open access repositories.

References:

- Bioline International (2008). Bioline International official site. <http://www.bioline.org.br/> (Visited – 06 Feb 2008)
- Bioline International (2006). Bioline Eprints Archive. <http://bioline.utsc.utoronto.ca/> (Visited – 06 Feb 2008 and found repository decommissioned)
- Budapest Open Access Initiative (2008). Self-Archiving FAQ: What can libraries do to facilitate self-archiving? <http://www.eprints.org/openaccess/self-faq/#libraries-do> (Visited – 06 Feb 2008)
- Kwasik, Hanna (2005) Open Access and Scholarly Communication -- A Selection of Key Web Sites. Issues in Science and Technology Librarianship. Summer 2005. <http://www.istl.org/05-summer/internet.html> (Visited – 04 Feb 2008)
- National Informatics Centre (2008). <http://home.nic.in/> (Visited – 04 Feb 2008)

- Open Archives Initiative (2002). The Open Archives Initiative Protocol for Metadata Harvesting. <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>
(Visited – 07 Feb 2008)
- Open Society Institute (2004). A Guide to Institutional Repository Software v 3.0, 3rd Edition. New York, Open Society Institute. <http://www.soros.org/openaccess/software/>
(Visited – 07 Feb 2008)
- OpenMED@NIC (2008). OpenMED@NIC. <http://openmed.nic.in> (Visited – 07 Feb 2008)
- Pandita, Naina and Singh, Sukhdev (2005). OpenMED self help tutorial. IMC. <http://openmed.nic.in/135/>
(Visited – 07 Feb 2008)
- Registry of Open Access Repositories (2008a). <http://roar.eprints.org> (Visited – 07 Feb 2008)
- Registry of Open Access Repositories (2008b). Search for a Repository: bioline. <http://roar.eprints.org/?action=search&query=bioline&sa=Search> (Visited – 07 Feb 2008)
- Registry of Open Access Repositories (2008c). Search for a Repository: bioline. <http://roar.eprints.org/?action=search&query=openmed&sa=Search> (Visited – 07 Feb 2008)
- Singh, Sukhdev and Pandita, Naina (2008). *Group discussion on “Open Access: impact on biomedical research” in World Café format*. In: Workshop on Biomedical Information Retrieval, 15-17 Jan 2008, New Delhi, India. <http://openmed.nic.in/2544/> (Visited – 07 Feb 2008)
- Singh, Sukhdev; Gaba, Surinder Kumar and Pandita, Naina (2004). Architecture and building of medical digital library at National Informatics Centre: what exists and what is required for MeDLib@NIC? In: International Conference on Digital Libraries, 24-27 Feb 2004, New Delhi. <http://openmed.nic.in/12/>
(Visited – 07 Feb 2008)
- Suber, Peter (2007). Open Access Overview. <http://www.earlham.edu/~peters/fos/overview.htm>
(Visited – 06 Feb 2008)
- Westrienen, Gerard van and Lynch, Clifford A. (2005). Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005. D-Lib Magazine, 11(9), <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>