

## Setting up an open access digital repository: A case study

Francis Jayakanth<sup>1</sup>, Filbert Minj<sup>2</sup> and Prabir G. Dastidar<sup>3</sup>

<sup>1,2</sup>National Centre for Science Information, Indian Institute of Science, Bangalore – 560012,  
Email: franc@ncsi.iisc.ernet.in, filbert@ncsi.iisc.ernet.in

<sup>3</sup>Ministry of Earth Science, Government of India, 'Mahasagar Bhavan', Block-12, C.G.O. Complex, Lodhi Road, New Delhi – 110003,  
Email:prabirgd11@gmail.com

*Received 03 August 2011, revised 06 March 2012*

Setting up of institutional repositories has been gathering momentum in India and many academic and R&D establishments have made it mandatory to set up institutional repositories. This paper briefly details the work that has gone into setting up and configuring the digital repository of the Ministry of Earth Sciences (MOES). The repository has been set up using the free and open source software, GNU Eprints.org (<http://eprints.org>). Such a repository will not only help in the wider dissemination of the publications that emerge from the projects and programmes supported by the MOES, but it will also serve as an information management system for the ministry.

### Introduction

An old tradition and a new technology have converged to make possible an unprecedented public good. The old tradition is the willingness of scientists and scholars to publish the fruits of their research in scholarly journals without payment for the sake of inquiry and knowledge. The new technology is the internet. The public good they make possible is the worldwide electronic distribution of the peer reviewed journal literature, completely free and unrestricted access to it by all scientists, scholars, teachers, students, and other curious minds. Removing access barriers to peer reviewed journal literature will accelerate research, enrich education, share the learning of the rich with the poor and the poor with the rich, make this literature as useful as it can be, and lay the foundation for uniting humanity in a common intellectual conversation and quest for knowledge. This kind of free and unrestricted online availability of peer reviewed journal literature is referred to as open access (OA) (<http://www.soros.org/kopenaccess/read.shtml>).

There are two complementary strategies to facilitate open access to scholarly literature –through Open Access Journals (OAJ) and Open Access Archives (OAA).

Open access journals are scholarly journals that are available online to the reader "without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself." Some are subsidized, and some require payment on behalf of the author. Subsidized journals are financed by an academic institution or a government information centre; those requiring payment are typically financed by money made available to researchers for the purpose from a public or private funding agency, as part of a research grant. There have also been several modifications of open access journals that have considerably different natures: hybrid open access journals and delayed open access journals ([http://en.wikipedia.org/wiki/Open\\_access\\_journal](http://en.wikipedia.org/wiki/Open_access_journal)).

The directory of open access journals (<http://www.doaj.org/>) lists more than 7500 OAJs as on date.

The second and the faster route of facilitating open access to scholarly literature is through the OAA. OAA are electronic repositories that may hold the authors' final versions of peer reviewed publications (post-prints), pre-published articles (pre-prints), theses, manuals, teaching materials, and any other material that the authors or their institutes wish to make publicly available without financial or other

access barriers. The only requirement for the end-users to access such archives is a computer system with internet connectivity.

An OAA can be either subject-based or university/institution-based. OAA administered by research institutes or universities for members of their community is referred to as 'Institutional Repositories' (IRs). An institutional repository is a set of services that a university or institution offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution<sup>1</sup>.

Many research institutions and universities across the world have already established their IRs. The Directory of Open Access Repositories (DOAR) (<http://www.opendoar.org>) and Registry of Open Access Repositories (ROAR) (<http://archives.eprints.org>) maintain comprehensive listings of all the IRs. Research has demonstrated that, with appropriate indexing and search mechanisms in place, open access online articles have appreciably higher citation rates than traditionally published articles<sup>2</sup>.

Indian scientists face two problems common to scientists everywhere, but acutely felt by scientists in poorer countries: access and visibility<sup>3</sup>. They find it difficult to access what has already been published because of the high costs involved in subscribing to journals and databases. Researchers in the rest of the world are unable to access what Indian researchers are doing. This is because much of the research work done in India is published in the national journals, which are not indexed by the leading abstracting and indexing databases. This leads to low visibility and low use of their work.

The visibility problem can be easily overcome by adopting the OA strategies – publishing in OA journals and/or self-archiving copies of accepted papers either in a subject-based central repository or in an institutional repository.

Government ministries and agencies are the major source of funding for research and development activities in the country. Amongst the various ministries that support research funding in the

country, the Ministry of Earth Sciences (MOES) is the major funding body in the field of earth sciences.

### **Ministry of Earth Sciences**

The Ministry of Earth Sciences (MOES) is mandated to provide the nation with best possible services in forecasting the monsoons and other weather/climate parameters, ocean state, earthquakes, tsunamis and other phenomena related to earth systems through well integrated programmes. The Ministry also deals with science and technology for exploration and exploitation of ocean resources (living and non-living), and play nodal role for Antarctic/Arctic and Southern Ocean research. The Ministry's mandate is to look after Atmospheric Sciences, Ocean Science & Technology and Seismology in an integrated manner.

The various units under the Ministry of Earth Sciences are: India Meteorological Department (IMD), National Centre for Medium Range Weather Forecasting (NCMRWF), Indian Institute of Tropical Meteorology (IITM) Pune, and Earthquake Risk Evaluation Centre (EREC) under the Atmospheric Sciences and Seismology sector; National Institute of Ocean Technology (NIOT) Chennai, National Centre for Antarctic & Ocean Research (NCAOR) Goa, Indian National Centre for Ocean Information Services (INCOIS) Hyderabad, Integrated Coastal and Marine Area Management Project Directorate (ICMAM-PD) Chennai, and Centre for Marine Living Resources & Ecology (CMLRE) Kochi under the ocean science & technology sector.

The ministry also supports various other programmes and has also established the Ocean Atmospheric Science Technology Cells (OASTC) as given in Table 1.

Many publications in the form of journal articles, conference papers, conference posters, book chapters, books, project reports, status reports and annual reports produced by the institutes based on the research programmes and projects supported by the MOES. Till recently, there has been no mechanism to manage these publications and, over a period of time, it has become difficult to retrieve such published works. Moreover, as the Government ministry is supporting the programmes and the projects, people at large should have free access to the publications that are produced by the institutes/programmes funded by the ministry. However, the present state of

Table 1—Ocean atmospheric science technology cells in universities

Sl. no.	University name	Subject
1	Andhra University	Coastal Marine Culture Systems
2	Annamalai University	Marine Biology
3	Berhampur University	Marine Coastal Ecology of East Coast
4	Cochin University of Science And Technology	Marine Benthos
5	Goa University	Marine Microbiology
6	Mangalore University	Marine Geology & Geophysics
7	Tamil University	Beach Placers
8	Bhavnagar University	Marine Coastal Ecology of West Coast

subscription-based access model to the majority of scholarly literature creates an access barrier to those who do not have access to subscribed content or to those who cannot afford subscription.

It is under the above mentioned scenario that the Ministry had approached National Centre for Science Information (NCSI) to carry out a project for it whereby an open-access, interoperable, digital repository would be set up by NCSI to manage and disseminate the publications of the ministry and its affiliate institutes and programmes.

### National Centre for Science Information

NCSI has a rich experience in setting up interoperable, open access, institutional repositories. India's first and one of the fastest growing IRs, ePrints@IISc (<http://eprints.iisc.ernet.in>), the institutional repository of the Indian Institute of Science was set up by NCSI in 2002<sup>4</sup>. To date, this repository holds more than thirty three thousand research publications of IISc. NCSI is also maintaining a separate repository for archiving and facilitates open access to the theses produced by IISc, etd@IISc (<http://etd.ncsi.iisc.ernet.in>)<sup>5</sup>. NCSI has also been creating and spreading the awareness and the importance of open access by way of undertaking sponsored projects and conducting seminars and workshops on the topic. NCSI has rendered technical support to various institutions and organizations in setting up their interoperable institutional repositories, online journal publishing, OAI protocol based harvesting service, web-enabling of legacy information systems, and library automation solutions. A sample list of institutions / organizations where NCSI has rendered its technical support in setting institutional repositories is available at <http://casin.ncsi.iisc.ernet.in/ncsi-ir.htm#irs>

### Application software chosen

As per the ROAR statistics, the three most popular software that are being used for IRs are DSpace (<http://www.dspace.org>), GNU EPrints (<http://www.eprints.org>) and FedoraCommons (<http://www.fedora-commons.org/>). All the three are free and open source software (FOSS).

GNU EPrints, developed by the University of Southampton's School of Electronics and Computer Science, is the world's first, most widely used, and by far the most functional software that made it possible for the institutions to create repositories to enable its authors to self-archive their research publications. It celebrated its 10<sup>th</sup> year of existence during the open access week in October 2010 (<http://www.ecs.soton.ac.uk/about/news/3445>).

ePrints@IISc has been using the same software since its inception. It's tailor-made software for IRs. The software is easy to install, maintain, and customize as per the local needs. Newer versions of the software are being released periodically. Technical support for the software both from the developers and also from the fellow users is invaluable. The technical documentation through the manuals and wiki pages (<http://wiki.eprints.org>) are also of immense help to people implementing and administering the software.

### Technical requirements for GNU EPrints

Any computer capable of running UNIX or UNIX-like operating system (OS) such as GNU Linux is recommended. It needs a backend relational database management system (RDBMS) to store metadata. GNU EPrints supports MySQL (<http://www.mysql.com>) and Oracle (<http://www.oracle.com>) RDBMS. All the dependent software for GNU EPrints like the RDBMS, PERL interpreter, Web Server (Apache), and PERL packages are available

for free and the best thing is that they all can be installed along with the OS if Red Hat Fedora or the Enterprise Linux is used. This is what makes GNU EPrints an out-of-the-box solution for setting up IRs.

As far as the hardware configuration is concerned, a present-day, entry-level desktop system is more than adequate to get started with. Depending on the growth of the repository in terms of records and access, the hardware configuration can be upgraded at a much later point of time.

### Prototype ePrints@MoES Repository

Initially, we set up a prototype repository service at the NCSI to facilitate formulation of requirements for the production service and also to address technical issues. The prototype development specifically focused on the following aspects of the repository:

- User interface design
- Document types and document formats to be supported
- Browse views to be created
- Repository policies needed for submission, approval, and copyright and access management
- Creation FAQs for the prospective submitters

Periodic discussion meetings were also held with the ministry's scientists. This helped us in understanding the needs and requirements of the ministry. Experiences gained from the prototype repository helped us to incorporate the needs and the requirements of the ministry in the production service.

### Implementation of production service

The production service was formally launched by Dr. Shailesh Nayak, Secretary MoES and Chairman, Earth Commission during the Steering Committee Meeting of OASTC held on 26th February 2010 at Goa University. Based on the recommendation of the ministry, the production server was set up at the Indian National Centre for Ocean Information Services (INCOIS). INCOIS is an autonomous body under the MoES. The production service became operational on the 15<sup>th</sup> May 2010.

The technical details of ePrints@MoES production service are as follows:

Operating System: Red Hat Fedora 12

CPU: Intel(R) Xeon(R) CPU E5462 @ 2.80GHz

RAM: 8 GB

Hard Disk Drive: 500 GB

GNU EPrints Version: 3.1.3

RDBMS: MySQL 5.1.39

Web Server: Apache 2.2

DVD r/w

The above hardware configuration is much higher than the minimum configuration that is needed to implement GNU EPrints software.

Some of the value additions that have been implemented in the production service are detailed below.

### User interface design

The look and feel of ePrints@MoES has been designed as per the specifications of MOES. Figure 1 shows the screenshot of ePrints@MoES homepage. Several useful features have been taken into consideration while designing the homepage. These include listing of total number of records in the repository as on date, listing of latest records uploaded to the repository, submission guidelines for those who intend to self-archive their publications, FAQs about the repository, guidelines for copyright conformance, provision for depositing articles by email, access to browse views, and links to other repositories that are related to oceanography and geosciences.

### Document types and file formats

By default, GNU EPrints supports various document types such as journal articles, conference papers, monographs, books, book sections, etc. The default document types can be customized as per the local needs. Figure 2 shows screenshot of document types being used in the ePrints@MoES repository. As far as the document formats are concerned, the GNU ePrints software does not impose any restrictions on them. It is up to the repository policy as to what file formats are to be allowed or what to be disallowed. And, when it comes to viewing of files, the end-user computer system should have the requisite software to handle the file formats. Here again, GNU Software has nothing to do with the handling of file formats. To render audio and video content, the repository service

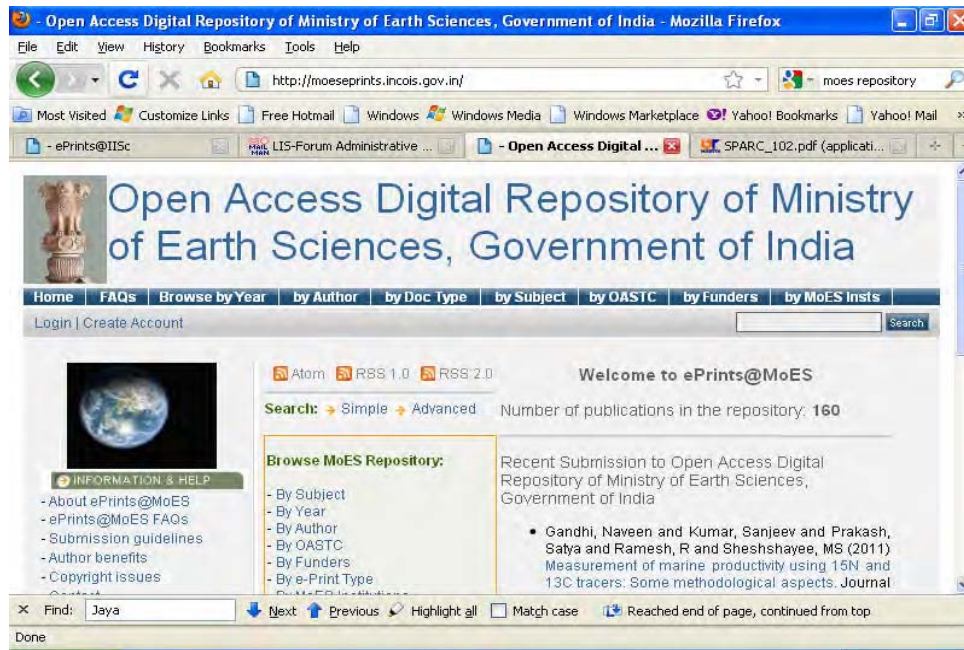


Fig. 1—ePrints@MoES Homepage

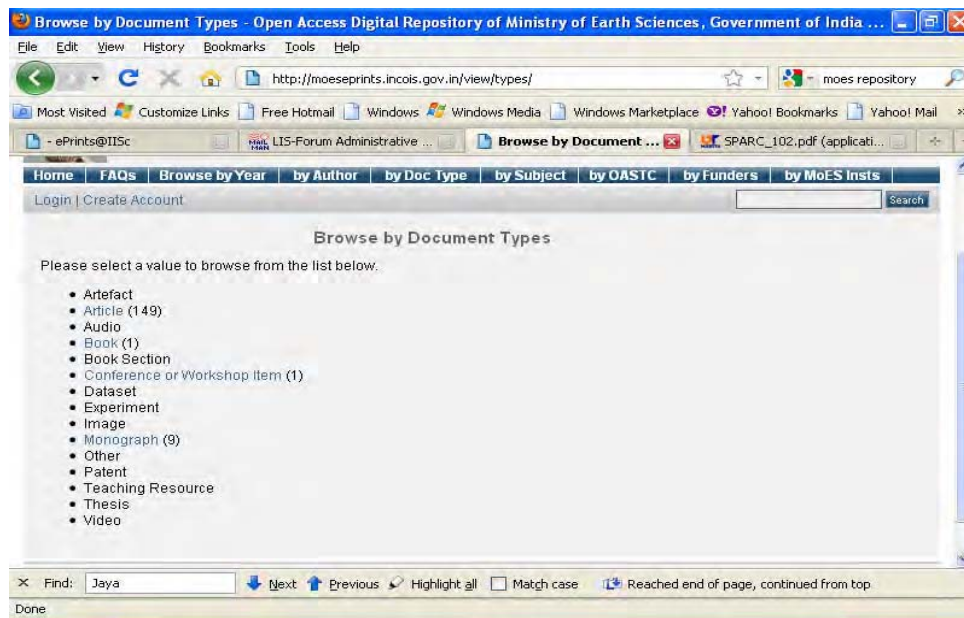


Fig. 2—ePrints@MoES Document Types

provider should facilitate streaming of such content. Streaming feature is not directly supported by GNU software. For the present, MOES repository does not support streaming.

Retrieval of records from the repository is either through searching or browsing. Both simple and advanced searches are supported by the software. The

advanced search facilitates in limiting a search to a specific field.

### **Browse views**

The browsing functionality facilitates in browsing the records in the repository. For ePrints@MoES, several browse views have been created. These include



browse by authors, document type, year of publication, funding agency, institutions etc. Figures 3 and 4 are screenshots of two such browse views, one for the OASTC, a programme supported by the MoES and the other by authors.

**Subject categories**

Categorizing the records that are uploaded in a repository is always a helpful feature. It helps the users to easily locate records pertaining to specific categories. To be able to categorize the records based on a subject classification scheme, the application software should support such functionality. GNU EPrints supports in implementing a subject

classification scheme for a repository. By default, the software supports Library of Congress subject headings. For the MOES repository, a broad-level NASA Geosciences Subject Categories (<http://www.sti.nasa.gov/sscg/42.html>) are being used. Figure 5 shows screen shot of browse by ‘Subject’.

**Copyright compliance**

Ideally, an IR should hold the final accepted versions of all publications produced in an institute. In reality, this does not happen. If it has to happen then the institutes and the funding agencies will have to mandate open access. Worldwide, more than 100

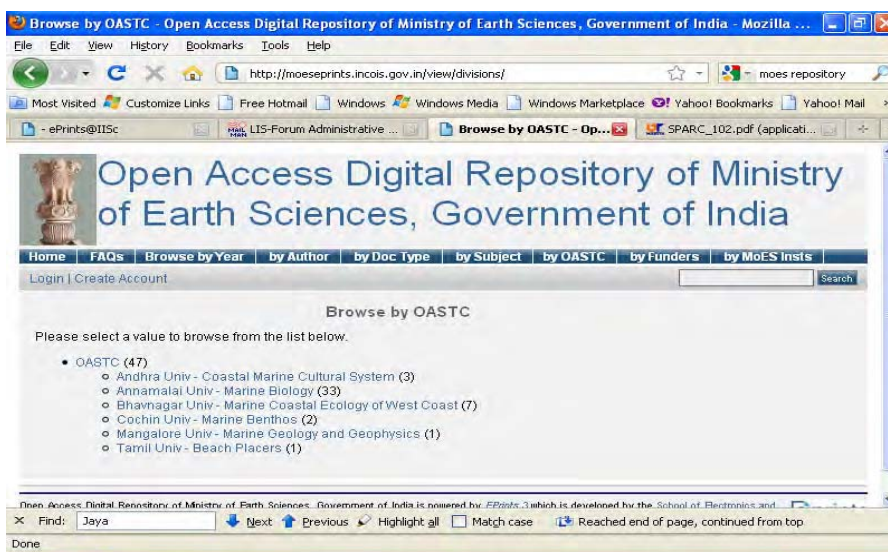


Fig. 3—Browse by OASTC



Fig. 4—Browse by authors

institutions and close to 50 funding agencies have mandated open access (<http://roarmap.eprints.org/>).

Publishers' policies towards depositing research papers in the OAA are being maintained by the SHERPA/RoMEO project (<http://www.sherpa.ac.uk/>).

For the present, most of the full-texts in the repository are the publishers' version. However, the accesses to the full-texts that are not in the OA domain are restricted to the registered users only. The registration itself is restricted to the users from MOES and its affiliate institutions. Others, who are in need of full-texts for academic purposes can make use of the

'Request a copy' button to get a copy of a paper. Figure 6 shows screenshot of MOES repository page with 'Request a copy' feature. End-users clicking on the 'Request a copy' button will be prompted to enter their email id. Upon submitting the form, an email request is sent to the concerned author. The decision to honour the reprint request rests with the author.

### OAI compliance and interoperability

GNU EPrints is compliant with version 2.0 Open Access Initiative Protocol for Metadata Harvesting (OAI-PMH). This functionality is referred to as OAI-

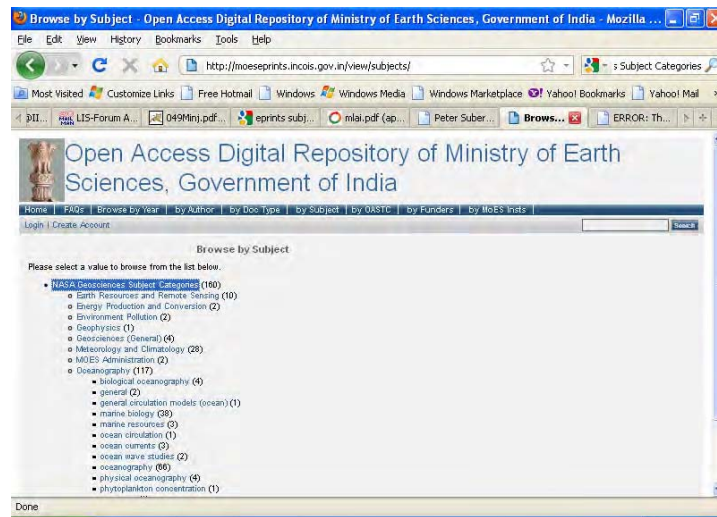


Fig. 5—Browse by subject headings



Fig. 6—Screenshot with 'Request a copy' feature

Table 2—Document types in the MOES repository

Document type	No. of records
Journal Articles	349
Book	1
Book Chapter	10
Monographs	10
Others	3

Table 3—Year-wise distributions of uploads in the MOES repository

Year	No. of records
2011	22
2010	58
2009	43
2008	45
2007	39
2006	36
2005	26
2004	31
2002	17
2001	8
2000	7
1999	2
1997	1
1996	1

compliance. OAI-PMH provides an application-independent interoperability framework based on metadata harvesting<sup>6</sup>. OAI-PMH defines a mechanism for harvesting records containing metadata from OAI-compliant repositories. The OAI-PMH gives a simple technical option for the data (repository service) providers to make their metadata available to services, based on the open standards, HTTP (Hypertext Transport Protocol) and XML. The metadata that is harvested may be in any format that is agreed by a community, although unqualified Dublin Core is specified to provide a basic level of interoperability. Thus, metadata from many sources can be gathered together in one database, and service can be provided based on this centrally harvested or aggregated data. GNU EPrints supports unqualified Dublin Core metadata standard.

MOES repository is OAI-compliant. We are now in the process of registering the repository with suitable OAI service providers. Once the registration process is completed, the metadata of the records in the repository will be periodically harvested by the OAI

service providers. A comprehensive listing of OAI service providers are listed at <http://www.openarchives.org/service/listproviders.html>. The OAIster database (<http://www.oclc.org/oaister/?c=oaister%3bpage=simple>) is one such service provider. It holds millions resources from thousands of contributors (repositories).

### The present status of the MOES repository

As on date, the repository holds three hundred and seventy three articles. Most of these articles have been uploaded by the INCOIS staff. Self-archiving by the MOES scientists and the other stakeholders is very sporadic. To improve the situation, MOES scientists are to be educated about the benefits of self-achieving their publications in the MOES repository. Also, some of them may have misconceptions about issues related to copyright compliance. A great deal of promotional and advocacy about the OA repository needs to be undertaken by the ministry.

Different type of articles included in the repository and the year-wise distributions are shown in Tables 2 and 3 respectively.

### Promotion and advocacy

Setting up an archive is one thing, but getting the researchers to fill the repository with their scholarly publications is quite another. Constant promotion and advocacy are to be pursued to influence upon the researchers to either self-archive their scholarly publications or use some mediation to do it on their behalf. Libraries have a crucial role to play in the promotion and advocacy of OA concepts to their researchers. As far as the MOES repository is concerned, there has not been much of promotional and advocacy campaign to date. However, the ministry is intending to mandate submission of all research publications that emerge from the programmes and projects that are funded by the ministry

### Conclusions

Interoperable, open access archives/repositories can significantly reduce the access barrier to the scholarly literature. It also enhances the visibility and therefore the impact of research works. Setting up and maintaining OAA does not require a great deal of



technical competency. There is no much of financial commitment either. The existing infrastructure in an organization could be effectively made use of to demonstrate the concept and also to set up the production service.

Most of the research projects in India are being financially supported by the Government ministries and departments. Research publications that emanate from such research projects can be published in any of the journals that the authors wish to. The only extra effort that is needed on the part of the authors is that a copy of the final accepted versions of their publications should be placed either in a subject-based repository or in the respective IRs. The present situation is that there may not be a subject-based repository for all the subject domains. Also, there may not be IRs in all the institutions engaged in research work. The need of the hour therefore is, to create more and more IRs or to create a central repository for the entire country as a national facility. Once such facilities are available, the project grantees can be mandated to self-archive the final accepted versions of all the publications emanating from government-funded sponsored projects.

The creation of MOES repository is a step in the right direction. Henceforth, the ministry will mandate all the project grantees to archive all the publications – both published and also the grey literature – in the repository. This will enhance the visibility of such publications. It can also be effectively made use of as an information management system. Over the period of time, such a system can help in the avoidance of duplication of research projects.

Other Government ministries and funding agencies should replicate OAAs and thereby mandate the research grantees to archive copies of research publications and other grey literature like the project reports, presentations, and experimental or survey data emanating from the projects in the respective OAAs. Such a practice would give much needed visibility for research works published in the local

journals that are not indexed by the leading international abstracting journals / databases.

### Acknowledgements

We thank the Ministry of Earth Science, Government of India for approaching and also giving NCSI an opportunity to set up its interoperable, open access digital repository.

The ministry would like to thank Prof. Subbiah Arunachalam, Distinguished Fellow, Centre for Internet and Society, Bangalore, India. He is a tireless advocate of open access movement in India. It was he who impressed upon the ministry to set up the OAA.

### References

1. Lynch C A, Institutional Repositories. Essential infrastructure for scholarship in the digital age, ARL. A Bimonthly Report on Research Library Issues and Actions from ARL, CNI, and SPARC, no. 226 (2003), Available at: <http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>. (Accessed on 2 February 2012).
2. Lawrence S, Free online availability substantially increases a paper's impact, Nature: Web Debates. Available at: <http://www.nature.com/nature/debates/e-access/Articles/lawrence.html> (Accessed on 5 February 2012).
3. Arunachalam S, Open Access in India: Hopes and Frustrations. Proceedings ELPUB 2008 Conference on Electronic Publishing - Toronto, Canada - June 2008. Available at: [http://elpub.scix.net/data/works/att/271\\_elpub2008.content.pdf](http://elpub.scix.net/data/works/att/271_elpub2008.content.pdf) (Accessed on 02 March 2012).
4. Jayakanth F, Minj F, Shiva U, and Jagirdar S, ePrints@IISc: India's first and fastest growing institutional repository, *OCLC Systems and Services*, 24 (1) (2008) 59-70.
5. Minj F, Jobish P, Biswal H, Rajashekar T B, and Venkadesan S, etd@IISc: a DSpace-based ETD-MS and OAI compliant theses repository service of the Indian Institute of Science, In Proceedings of Eighth Symposium on Electronic Theses and Dissertations", Sydney Australia, (2005). Available at: <http://snidely.dlib.vt.edu:8081/dspace/bitstream/2340/229/1/049Minj.pdf> (Accessed on 15 February 2012).
6. Lagoze C, Somplé V D H, Nelson M, Warner S, The Open Archives Initiative Protocol for Metadata Harvesting – Version 2.0, (2002), Available at: [http://www.openarchives.org/OAI\\_protocol/openarchivesprotocol.html](http://www.openarchives.org/OAI_protocol/openarchivesprotocol.html) (Accessed on 17 February 2012).