

Digital Repository Service (DRS) at National Institute of Oceanography: A case study

Satya Ranjan Sahu

National Institute of Oceanography, Dona Paula, Goa- 403004

Abstract:

Open access to the publications over Internet has become a reality. The software likes DSpace help in creating institutional repositories to place the publications on Internet. Implementation of an institutional repository at National Institute of Oceanography (NIO) is described. The paper shares our experiences for sound footing and progress. Our experience indicates that DSpace is a useful software in promoting institutional repositories and open archives help to increase visibility of research work.

Keywords: Institutional Repository, Open Access Initiative, Digital Repository, Research Impact, NIO

Corresponding Email: satya@nio.org

Tel: +91-832-2450 370; Fax: +91-832-2450606

Introduction:

The Scientific and technical publications bring credibility to the research organizations and the quality impact in the researchers community, especially, the peers in specific subject. Easy access to such publications, especially from and to developing countries, is still a challenge. Rising costs of journals and depleting library budgets make it difficult for libraries to subscribe them on a continual basis. This means, the potential research impact of this research literature is lost.

Lawrence¹ accounted a substantial increase in citation rate for online articles in the computer sciences discipline. Similar studies by Antelman² for four other subjects – mathematics, electrical and electronic engineering, political science and philosophy and Eysenbach³ for same journal PNAS (*Proceedings of the National Academy of Sciences*) have confirmed Lawrence findings.

To promote wider access to and visibility of the publications of the National Institute of Oceanography (NIO), an online reprint request system was developed in mid of 2004 using the existing bibliographic database of the institute publications. This resulted in over 3000 reprint requests per year – all supplied as scanned copies over mail. The service is still being continued. Understanding the limitation of the service like:

1. Digital content is still not downloadable online,
2. Users require to wait to receive the E-prints over email, and
3. No provision to archive and preserve digital assets as well as no permanent and persistent URL,

An open access self-archiving Institutional Repository at NIO was established in 2006

Implementation of OAI-PMH compliant Digital Repository Service (DRS) at NIO:

Open Access self-archiving movement offers open source software to develop OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) compliant archives. DSpace (<http://www.dspace.org/>) is one of the widely used open source software jointly developed by MIT libraries and Hewlett-Packard labs. Studies^{6,12} were proved that DSpace is easy to install, requires common/basic software, require minimum effort to work on code and has better security features. A strong backup of help from developers as well as users community is an added advantage.

Considering above features DSpace was the choice for an institutional repository at National Institute of Oceanography. The same was installed during February 2006. The URL: <http://drs.nio.org> (a Digital Repository Service at NIO).

DRS Operational issues and their solutions:

Installation of software is just the first step towards a successful implementation of an IR. Several others issues associated with implementation include:

- Content organization and retrieval
- Content policies
- Content collection process
- Copyright
- Administration and workflow
- Content contribution to harvesters
- Evaluation and analysis

Content Organization and Retrieval:

Success of any IR depends on the size and organization of content. An interface of DSpace provides flexible well-developed features to organize the content as well as its retrieval. In a research institute like NIO, where the research activity is interdisciplinary and organizational setup is not large enough, it was a tricky to

classify the content into different communities and collections. Following broad categories decided with single community- Institute's Contribution:

- Biological Sciences
- Chemical Sciences
- Engineering and Instrumentation
- Geology and Geophysical Sciences
- Interdisciplinary Oceanography
- Marine Archaeology
- Material Sciences
- Miscellaneous
- Physical Sciences

To avoid an extra step browsing to collections in DRS, the link to communities' page was replaced with collection page.

Content policies:

IR implies to an online locus for collecting and preserving the intellectual outputs of an institution. It was decided to cover journal articles, conference proceedings articles, Technical reports, thesis, dissertations, etc in NIO's repository (DRS). Depending upon permissions the documents upload in DRS of source publications varies from pre-print, post-print to original reprints. The PDF, so far is the preferred format of archiving in DRS although Dspace supports many other formats.

Content collection process:

The pre-prints or post-prints of yesteryears with the authors didn't exist though the 'green' publishers had permission to place same on the IR. Moreover, the benefits of increase in citations to older publications were not seen from Institute's point-of-view. Therefore it was decided to concentrate efforts to obtainable publications from the year 2000 onwards for DRS. We explained the authors that DRS will benefit them to...

- Promote wider access and visibility to their publications.
- Increase in citation.
- Enhancement in rapid communication.
- Chances for collaborative research.
- Establishment of priority to research findings.

On establishment of DRS, we even organized a seminar from an expert researcher from outside NIO to stress on the above benefits and developments that are taking place in this direction in India.

Follow-up with the user begins when we identify the article. This includes initial request for self-archiving followed by reminders with request to self-archive or provide their pre or post-refereed author version of the text (especially of the 'green' publishers). At times personal visits to the researchers for the content are needed. While utmost efforts are made to archive current publications in DRS, permissions were obtained to archive scanned articles/publisher's PDF files from Indian publications of earlier years. In certain cases, a confirmation was obtained from the authors that they didn't transfer copyright to the publisher. Table1 gives year wise distribution of documents archived in DRS as on 10 August 10, 2007. Note that the copyrighted book chapters, reports of restricted distribution are out of limits of DRS.

While authors find the input form and metadata fields simple only 10-25% of total archived items are self-archived by them. This is so because we are yet to prove visible benefits of self-archiving in DRS. The default Dspace input form is found to be self-explanatory covering all metadata. Only one extra field for 'guide' was added for document type 'thesis' and input hits were elaborated.

Table-1: Number of documents archived in DRS

Year of publication	Numbers of Institute publications	Number of items in DRS	Percent input
2007	113 till date	74 till date	65%
2006	230	173	75%
2005	213	102	47%
2004	160	49	30%
2003	152	51	33%
2002	151	43	28%
2001	138	37	27%
2000	165	52	31%

Copyright issues:

We respect copyrights. It is a sensitive issue and requires careful handling. Some journals have confusing copyright policy and with a few publishers, they are not clear. Most of the international journals have clear copyright policy. As a preliminary step we consult Sherpa/Romeo site- <http://www.sherpa.ac.uk/romeo.php> to get brief policy on self-archiving. We also maintain printed copyright policies of each green publisher whose articles are added to DRS for the first time. Recently a change in archival policy of ‘green’ publishers for some journals is observed. To avoid any future complications it felt to keep the printed copyright policy as best practice.

In case of Indian journals, clear copyright policies for self-archiving rarely exist. At times, authors are unable to recall if any copyright agreement is signed for a particular article.

Administration and Workflow:

Handling administrative, workflow and promotional aspect of IR are well defined and sharing of activities is more pragmatic in DSpace. But optimum use of inadequate skilled human resource necessitates united responsibility in an R & D like NIO. In other hand it also helps in learning from the practical difficulties of administering and maintaining.

Content contribution to harvesters:

An OAI-PMH supported IR is with ‘harvest ready’ feature but testing and registration require. DRS is registered with a global aggregators -OAIster, Google Scholar; subject specific aggregator – Avano (Marine and aquatic science OAI harvester) and Indian aggregator – CASSIR (Cross Archive Search Services for Indian Repositories).

Evaluation and analysis:

Regular evaluation and analysis of hits to DRS helps in improving the features and provides insight into the usage of this facility. DRS is based on DSpace-1.3.2. The statistics in DSpace donot give a correct picture of usage of the resources. Apache access log is therefore analyzed further to obtain correct usage. Item wise usage also gives insight into the interest in a particular article. The data is maintained and updated. It is too early to comment on this aspect.

DRS impact on marine science community:

Marine science/oceanography is comparatively young discipline. Analysis of union catalogue of International Association of Aquatic and Marine Science Libraries and Information Centres (IAMSLIC – NIO also a member and in total 55 members) shows, not more than 25% members library subscribe all journals in which NIO articles published. To bring “every article published by researchers at NIO to its deserving users[†]” in the field of oceanography DRS is proving to be suitable solution.

625 items on DRS whose addition is spread over 16 months as on today, we experience a download of 45,000 times (average 3000 times per month). In comparison to reprint request facility, DRS is therefore most impressive. Access analysis indicates that the full-text downloads are from over 100 countries, mostly on the coastline (Figure1). 30-40% downloads of these are from Academic and R&D organizations network addresses. Rest downloads are from IPs without any domain name and registered to service providers.

[†] “Every book its reader – Third law of LIS”: The Five Laws of Library Science were enunciated in 1928 by the late Dr. S.R. Ranganathan, considered as the Father of Library Science in India.

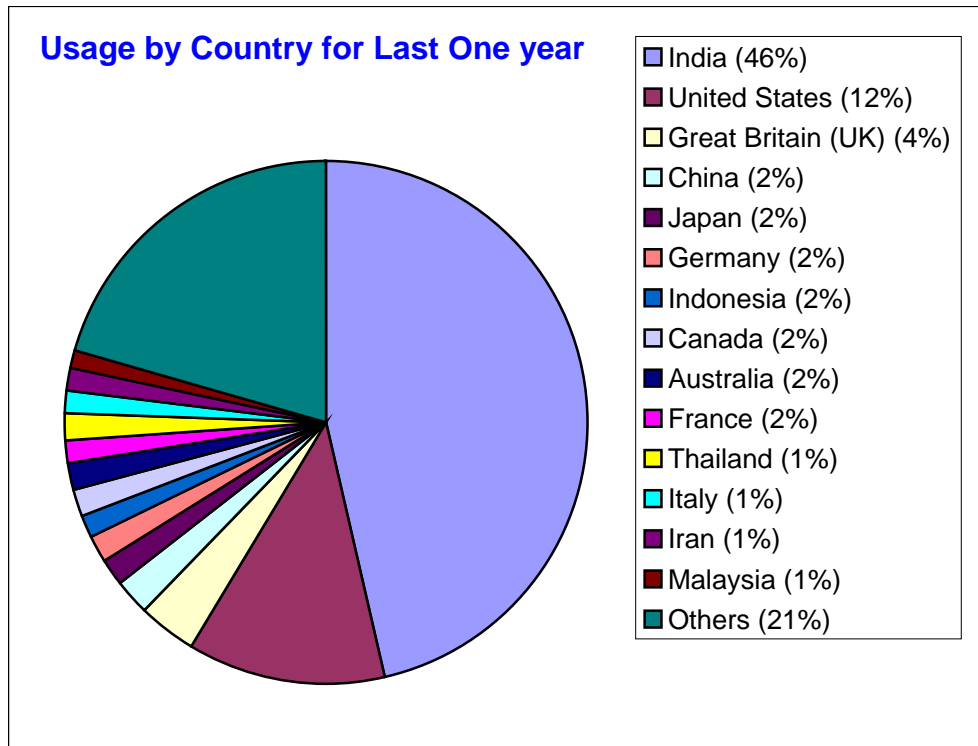


Fig 1: Geographical distribution of article downloads from DRS

Conclusion:

Open access institutional repositories help to increase visibility of research work. This is not only need/useful for developing countries but also for developed countries. DSpace is a useful software in promoting institutional repositories. Although it is better suited to large institutional setting to manage a repository but there is no technical difficulties to implement in small institutional setting. To provide wider access and visibility to institutional literary output academic and research institutes needs to adopt repositories on priority basis.

Acknowledgement:

I am thankful to Dr. S. R. Shetye the Director, NIO, for all facilities and support and grateful to Dr. M. P. Tapaswi for the inputs and encouragement to implement IR at NIO and to present the paper.

Reference

1. Lawrence, S. (2001) Online or Invisible? *Nature* 411 (6837), 521.
<http://citeseer.ist.psu.edu/online-nature01/>
2. Antelman, K (2005) Do open-access articles have a greater research impact? *College & Research Libraries*, 65 (5), 372-282.
<http://www.la-press.com/include/Antelman.pdf>
3. Eysenbach G. (2006a) Citation Advantage of Open Access Articles. *PLoS Biol.* 4(5) p. 157. <http://dx.doi.org/10.1371/journal.pbio.0040157>
4. <http://www.openarchives.org/>
5. Suber, P. (2006) Six things that researchers need to know about open access. *SPARC Open Access Newsletter*, 94.
<http://www.earlham.edu/~peters/fos/newsletter/02-02-06.htm#know>
6. Harnad, S. (2001) The Self-Archiving Initiative. *Nature* 410, 1024-1025.
<http://www.nature.com/nature/debates/e-access/Articles/harnad.html>
7. Brody, T., Harnad, S. and Carr, L. (2005) Earlier Web Usage Statistics as Predictors of Later Citation Impact. *Journal of the American Association for Information Science and Technology (JASIST)* 57(8), 1060-1072.
<http://eprints.ecs.soton.ac.uk/10713/>
8. Chan, L., Kirsop, B. and Arunachalam, S. (2005) Open Access Archiving: the fast track to building research capacity in developing countries. *SciDevNet*, November 2005. <http://www.scidev.net/ms/openaccess/>
9. Arunachalam, S. (2004) India's march towards open access. *SciDevNet*, March 2004.
<http://www.scidev.net/Opinions/index.cfm?fuseaction=readOpinions&itemid=243&language=1>
10. Ranganathan, S. R. (1931) *The Five Laws of Library Science*. Madras Library Association (Madras, India) and Edward Goldston (London, UK).
<http://dlist.sir.arizona.edu/1220/>
11. Foster, N.F. and Gibbons, S. (2005) Understanding Faculty to Improve Content Recruitment for Institutional Repositories. *D-Lib Magazine* 11(1).
<http://www.dlib.org/dlib/january05/foster/01foster.html>
12. CPIT Technical Evaluation of selected Open Source Repository Solutions (2006).
<https://eduforge.org/docman/view.php/131/1062/Repository%20Evaluation%20Document.pdf>