

OPEN MANTRA for Open Access to Information*

A.R.D. Prasad, Devika P. Madalli

Documentation Research and Training Centre,
Indian Statistical Institute
Bangalore 560059, Karnataka, INDIA
{ard, devika}@drtc.isibang.ac.in

Abstract

The importance of Open Access (OA) to information cannot be over emphasized. If 'Knowledge for all' has to be a reality then 'Information for all' must be ensured. Though the concept of OA is simple and straight forward enough, it encompasses many challenges such as copyright, government policies, awareness, technology among several other issues. This paper discusses a few of these issues and possible solutions towards OA. The paper proposes the OPEN MANTRA¹ – Open SOURCE tools based upon Open STANDARDS to achieve OPEN ACCESS to information.

I Introduction

In the scholarly world, access to information has a significant role in the production of information. Not all libraries can afford to acquire all resources needed by their patrons given the rising costs of publications. Further, unless we get access to the research output of other researchers, research productivity gets hampered; additionally, the chances of duplication of efforts would be high. To ensure that information reaches its patrons without barriers of cost or copyright, Open Access (OA) to information is essential. Internet and e-publishing have added a new dimension to the user's expectation towards the open access to information. Publication of scholarly literature on Internet is simpler than traditional modes of publishing. Ease

* Published in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (eds.) (2007). Open Innovation. Proc. 10th International Symposium for Information Science. Constance: UVK, 97-108.

¹ Mantra in Sanskrit means chant or spell.

of use of HTML and availability of open source tools for e-publishing, helped in the emergence of more than 2500 open access journals and magazines.

As the OA movement is gaining momentum, more tools and technologies are coming up which make practical OA system possible. Open source digital library software (e. g. DSpace, EPrints) are the key enabling software to set up OA repositories. Software like Open Journal System (OJS) provide open access platform to publish e-journals incorporating many of the methodologies of print journals with regard to peer reviewing. Open Archive Initiative – Protocol for Metadata Harvesting (OAI-PMH) is an excellent example of a facilitating technology enabling accessing of information across open access digital repositories and OAI – PMH compliant journals.

However, there are several serious challenges in the way of making the scholarly information open to all. In this paper we discuss some of the major challenges in achieving open access to information.

1.1 Peer-reviewing and its alternatives

In general, the journal ranking is established on the basis of the peer review process. Though some of the OA journals are peer-reviewed, there is a common opinion that OA journals are NOT peer-reviewed and hence they are of low quality compared to commercially published journals. The underlying strength of peer review is “...the concerted effort by large numbers of researchers and scholars who work to assure that valid and valuable works are published and conversely to assure that invalid or non-valuable works are not published [...]”. (Weller, 2001).

Though many of us agree with the views of Weller about peer reviewing, it should be pointed out that peer reviewing is not without its pitfalls as expressed in literature and discussions.

- Some editors reject without sending the paper to referees
- Sometimes editors may deliberately choose a harsh referee for a paper the editor wishes to see rejected
- Bias against authors depending on nationality, gender, institution etc.
- Author and referee belonging to opposing school of thought
- Reviewers can reject a good paper and write a paper using the ideas
- Delaying publication of potentially competing research
- Reviewers may be good in the subject and may not have skills in reviewing
- The limitations of the subject knowledge of the reviewer himself may make him sceptical of the publication under review.

It should also be pointed that there are alternatives to peer-reviewing which cover some of the pitfalls of peer reviewing. After all, it is finally the reader who decides what to accept and what not to accept when he reads a publication. The following are some of the alternatives to peer reviewing though they too are not without pitfalls. However, the institution which intends to publish may look at the appropriateness of these alternatives.

- Certification-based (reviewers are trained)
- Open Peer Review (reviewers sign)
- Commentary-based (readers can comment before or after actual publication)
- Collaboratively-filtered (guiding readers on what to read)
- Institution-based (institutional repositories)
- Citation-based
- NO Peer Review

1.2 Mandate

Though the philosophy to OA appeals to most academicians and researchers who are most content generators, OA repositories and journals are not populated easily. Many a project in OA suffers from paucity of volunteered content. The authors have been active in campaigning OA in Indian scientific and academic community. It is often found that there is not enough motivation for authors that want to go Open Access. Performance assessment and career development still largely depends on how many published articles a person has in 'ranked' journals. So the need of the hour for sustained OA systems is MANDATE. Each institute should mandate that its research output should be deposited in their own Open Access Repositories.

The argument of OA activists world over has been that "output of publicly funded research must be made freely, publicly accessible". This is in general the message brought by the Berlin declaration (Berlin Declaration, 2003), and supported by initiatives like the Budapest Open Initiative (Budapest Open Access Initiatives, 2001). The bills such as the *American Center for CURES Act* of 2005 and Federal Research Public Access Act of 2006 (FRPAA) (Peek, 2006) (English & Suber, 2006), tabled in US senate in support of making free access to publically funded research results, if passed, will greatly pave the way of OA to information.

2 Open Source tools and technology for OA

Open source movement has indeed contributed valuable tools for different applications and fortunately a good number of them can be used to offer products as OA resource repositories and archives. Digital Repositories are popular means of making available, institutional, domain based and community resources in an organized manner as openly accessible. There are quite a few very good tools that help build and manage digital repositories. We can broadly categorize the software in the following groups: (i) software for OA repositories; (ii) software for publishing and managing the journals; (iii) Metadata Harvesting software and (iv) Software for Digital Preservation. In the following paragraphs, we have discussed a few of the above mentioned software.

2.1 Digital Repositories

Open source software have facilitated the OA repositories in many ways. Popular open source digital repository tools include DSpace (<http://www.dspace.org>), EPrints (<http://www.eprints.org>), Fedora, GSDL (www.greenstone.org), Fedora, CDSware etc. However, it is important that digital library managers make informed decisions of what tool would be suitable for their collections, patrons and services they plan to offer. Some of these softwares are broadly evaluated by us and we have chosen DSpace for our production system, Librarian's Digital Library (LDL – <https://drtc.isibang.ac.in>).

LDL is India's first DSpace based repository and only second in world in the domain of Library and Information Science (LIS). It has community collection building modus with authors from about 14 countries and has over 300 members. The activities are supported by Documentation Research and Training Centre (DRTC), Indian Statistical Institute (ISI) and through an associated discussion group, Digital Library Research Group, (DLRG). Initially, we have uploaded many articles published in the DRTC's annual and refresher seminar volumes of the recent past with born digital versions of the articles. Digitization of the older seminar volumes has been undertaken and many more publications will be uploaded soon. Though, we have invited many authors to upload their publications, if copleft, we only had a lukewarm response. We believe, the reasons are varied, viz. fear of copyright laws mostly out of unawareness of copyright issues and also lack of awareness of open access movement, misplaced feeling that it requires knowledge of web to upload a document etc.

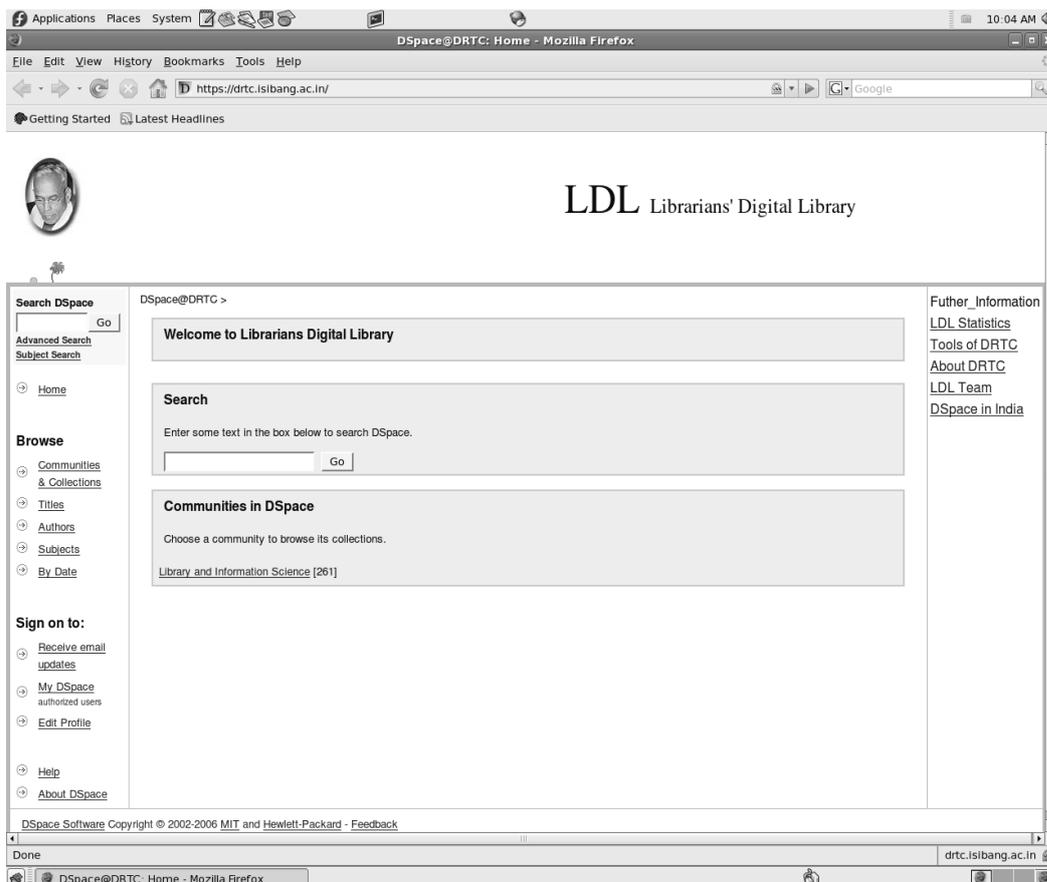


Figure 1: DRTC/ISI LDL Repository Service in LIS

2.2 Directory Services

OpenDOAR (The Directory of Open Access Repositories): OpenDOAR (<http://opendoar.org>) categorizes and lists wide variety of Open Access research archives. The project is a joint collaboration between the University of Nottingham in the UK and the University of Lund in Sweden. Both institutions are active in supporting Open Access to Information. We estimate about 1900 repositories and openDOAR covers about 836. OpenDOAR has an excellent search facility.

Directory of Open Access Journals (DOAJ): There are at present more than 2526 journals in the directory. Currently 753 journals are searchable at article level. As of today 125067 articles are included in the DOAJ (<http://www.doaj.org>) service. All the OA journals are categorized under broad disciplines, in addition to browsing and searching for a specific journal.

Harvester Services

OAI-PMH ensures interoperability among complaint Digital repositories. Using OAI Harvester software many information services have come into existence. One of the best known such services is OAIster (<http://oaister.umdl.umich.edu/o/oaister/>).

Arxiv is yet another service popular in the field of physics, mathematics, computer science etc. At Documentation Research and Training Centre (DRTC), Indian Statistical Institute (ISI), we are providing a harvester server called Search Digital Libraries (SDL – <http://drtc.isibang.ac.in/sdl>). SDL harvests metadata from various digital repositories in the field of Library and Information Science and presently provides access to more than 20,000 full text documents from 17 repositories.

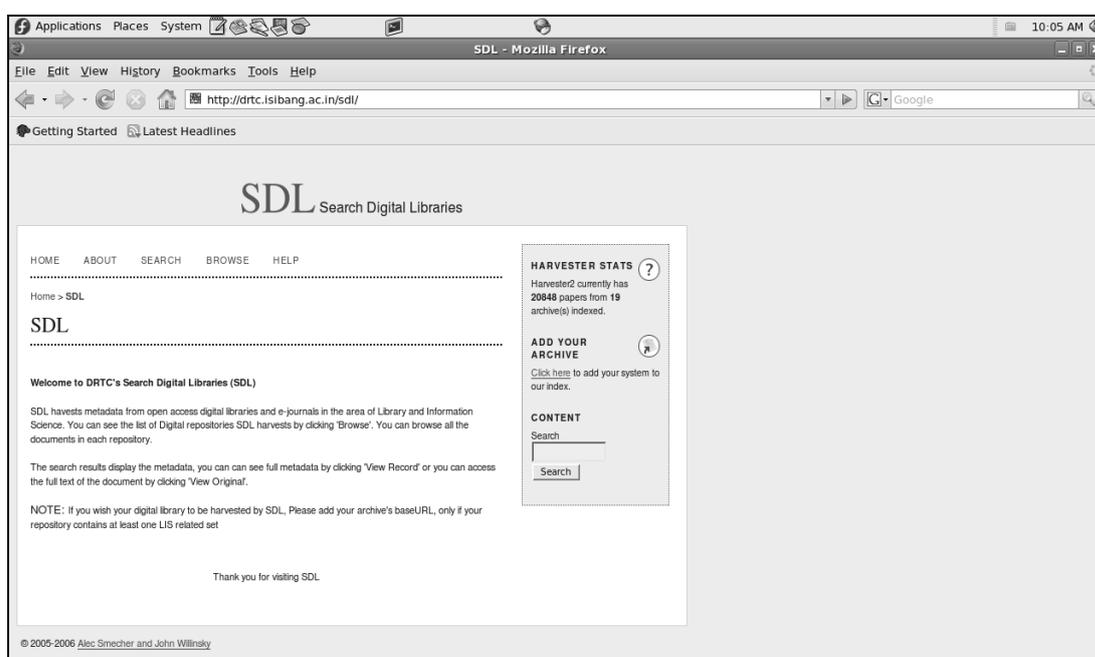


Figure 2: DRTC/ISI SDL Service in LIS

Though we know that there are more than 17 repositories having OA content in Library and Information Science (LIS), we have serious problems in harvesting those repositories. The methodology, we have adopted to identify repositories dealing with LIS content, is to go through the directory services like openDOAR, website of OpenArchives (<http://www.openarchives.org>). In addition, the software websites provide list of repositories that are using their software. For example, EPrints and DSpace sites do have list of organizations and the URLs of their repositories. However, it should be noted that these directories are not exhaustive as many people do not post information about their repositories in these directories. That means there is an urgent need to develop some mechanism or at least a convention to identify digital repositories and their URLs. Even when we identify a repository, identifying the baseURL in order to harvest metadata is yet another challenge. If the repository is using popular software like DSpace or EPrints one can guess the baseURL, even if a repository and its baseURL is not available in any of the directory services. However, in case of a significant number of repositories guessing the baseURL turns out to be a tedious and sometimes impossible effort. After all, one

can expose the metadata of repository using OAI-PMH in various ways using varied URL syntax using a PERL and PHP etc. programs.

In some cases, it is easier to identify the repositories dealing with LIS, especially if they deal exclusively with LIS, as in the case of DLIST, E-LIS, LDL, Australian Library and Information Association. However, it should be noted that many repositories, especially in the case of repositories host by universities, they may host many other subjects in addition to library and information science. The OAI-PMH protocol allows selective harvesting by 'set' or by 'date'. In case of some repositories, the LIS related content is normally available under a set. Using selective harvesting facility of OAI-PMH, one can get the list of sets in particular repository and harvest only the sets that deal with the subject or discipline of interest. The difficulty with selective harvesting is that many repositories might not have organized the content by subject. For example, if they create a set of 'Theses' and upload all theses from various subject under the set 'theses', the selective harvesting approach will not be of any use, as we can not filter out theses of other disciplines.

As there is no provision for selective harvesting by subject using OAI-PMH verbs, one can filter the metadata records after harvesting, that is at database level, as the harvested metadata is normally stored in a database. For example, PKP harvester has provision to store metadata records either in MySQL or PostgreSQL. The problem with this approach is that many people enter keywords following either a thesaurus or the keywords given in the publication or adhoc keywords. Even in case of controlled vocabulary they may follow different thesauri to enter the keywords and this does not ensure uniformity in the choice of keywords.

One way, we believe, to overcome this problem is to add a new element to Dublin Core metadata schema viz. dc:discipline and to mandate that one should choose the discipline from Library of Congress subject heading or Sears list either up to 1st order or 2nd or 3rd order array in the hierarchy of subjects, along with the broader terms. In addition, OAI-PMH protocol should have a selective harvesting by 'discipline'. Then it would be easier to build harvester for various disciplines. With the present OAI-PMH and Dublin Core versions, it is difficult to build satisfactory discipline based harvester services and DRTC's SDL suffers from these lacunae, as do any other domain-based harvester service.

2.3 Journal Publishing

Open Journal System (OJS), HyperJournal, DPubs are some examples of OSS tools for publishing and managing the OA scholarly journals and e-monographs. Over

800 OA journal titles (as of June 2006) in ten languages are already using OJS as OA (OJS-Journals). Many of the journal publishing software support OAI-PMH protocol. Though DOAJ lists about 71 journals in LIS, we have identified more than 80 electronic journals in the field of library and information science. Unfortunately, DOAJ has not tackled issue of baseURLs and we could not get the baseURL of these journals to harvest the contents of the journals in our SDL service. When we wrote to these journals, some did not respond and some of them wrote back their site is not OAI-PMH compliant.

The problem, we believe, is that many journals were initiated much before the OAI protocol came into existence. Though, there are tools available in open access like OCLC's OAIcat to expose the metadata and turn the site into a data provider, still the task of entering metadata for every article published by these journals is a problem to reckon with. In addition to OAI protocol, the RSS feed is yet another way of providing the contents list of journals which a significant number of journals are providing. However, harvesters using OAI protocol do not have mechanism to capture metadata from these journals. An interesting problem worth investigating is possible to have the harvester software to embed RSS facility into their software.

DOAJ site has been made OAI compliant. Of the 2549 journals listed in DOAJ, 757 journals' articles can be searched at article level and total of 125716 articles metadata is available. Unfortunately, the problem is that when we use OAI verb to harvest metadata of these, it gets metadata of all the articles and there is no way to do selective harvesting by journal title or ISSN. We hope in future, these issues will be address for better retrieval of information available in open access journals.

2.4 Open Standards

The commercial competitive world of computer software products and tools often push the standard which in turn makes certain tool inevitable. Ideally, OA models should be able to develop, grow, change and sustain over a period of time. The direct implication is that the solutions over a period of time should be compatible through the versions of software while being cost effective. Also the changes and migration should not entail buying new software or entering into more agreements and licensing for their use. OA systems are often made available through collaborative development online across communities and probably to as many patrons as are interested in the content. Given this situation, Open Source Software seem the ideal which while being free for download and use, provide autonomy in operations of OA collaborative projects. But there is need for caution while choosing the OSS. It is not only enough if it is FREE but it is equally important that it is NOT based on

or using at any stage proprietary standards. To implement OA successfully it is important the OSS be only based on OPEN STANDARDS.

There is no universally accepted definition available for open standards. Many people have different views about it. In general, open standards are publicly available, independent of any single institution or manufacturer. The key characteristics of open standards as identified by Coyle (Coyle, 2002) are: 1) that anyone can use the standards to develop software, 2) anyone can acquire the standards for free or without a significant cost, and 3) the standard has been developed in a way in which anyone can participate. The Danish Government defines (eGovernment News, 2004, 28 June 04) open standards as, it should,

- Be accessible to everyone free of charge: no discrimination between users, and no payment or other considerations should be required as a condition to use the standard,
- Remain accessible to everyone free of charge: owners should renounce their options, if any, to limit access to the standard at a later date.
- Be documented in all its details: all aspects of the standard should be transparent and documented, and both access to and use of the documentation should be free.

It is important for libraries and other cultural institutions to ensure long-term access to digital information. The rapid growth in digital technologies has led to new and improved applications for digital preservation. However at the same time it has also led to some problems as well. Two of these problems are obsolescence and dependency issues. The obsolescence problem is caused by the advances in hardware and software making many computers obsolete within three to five years (Vilbrandt, et al. 2004). Dependency problems can arise if tools that are needed to communicate between systems or read file formats become unavailable. In order to account for obsolescence and dependency problems organizations must be able to migrate data into new systems. Data migration, however, cannot occur without access to data file formats.

Some important characteristics of open standard can be identified as,

- It should not be encumbered by a patent
- It should not require proprietary software
- Can be utilized by anyone without cost
- It should be available for all to read and implement
- Implementations of Open Standards may be extended, or offered in subset form

- It should create a fair, competitive market for implementations of the standard. Should not lock the customer in to a particular vendor or group.

The categories of standards that directly having a bearing on OA broadly are:

1. File formats
2. Digital Preservation Standards
3. Metadata and data standards
4. Technical and Information Architecture standards
5. Interoperability standards
6. Encoding Standards
7. Network Protocols and standards

Examples of some open standards are: SQL (a specification approved by ANSI and ISO), HTML/XHTML (specifications of the W₃C for structured hyperlinked document formatting), OpenDocument (a specification by OASIS for office documents, approved by ISO as ISO/IEC 26300), etc.

3.5 Changing Copyright Scenario

SHERPA's RoMEO project (<http://www.sherpa.ac.uk/romeo.php>) categorizes publishers into different color categories according to their policies regarding copyright and self-archiving. Green category lists those publishers which allow authors to submit the preprints and post prints to self archives. Elsevier, Emerald, Haworth Press, IEEE, John Wiley, Sage, and Springer are among the major commercial publishers who have accepted the self-archiving policies applicable of RoMEO Green category with certain conditions, such as (SHERPA – RoMEO):

- Author can submit to personal or author's institutional server/ repository
- Published source must be acknowledged
- Must link to journal home page
- Publishers version/PDF cannot be used
- Articles in some journals can be made Open Access on payment of additional charge

4 Conclusion

It is the reviewers, who are domain experts and the reviewing process that makes a journal a highly ranked one and it will take a few years to consider a journal to be a good quality one. Considering the fact that commercial journals have been in exis-

tence more than 100 years and OA journals are hardly a decade old, it is too optimistic to expect OA will pervade soon. But it is difficult to deny the fact that the results of OA movement – OA journals, OA repositories is very impressive and significant.

The key to quick success of OA is highly dependent on the mandates from the authorized bodies. This will encourage the scholars to publish their publications in open access journals and to deposit it to institutional or self-archiving repositories. It is highly inadvisable to mandate direct deposit in a Central Repository (CR) -- whether discipline-based, funder-based, multidisciplinary or national. The right way to get OA content into CRs is to harvest it from the IRs (via the OAI protocol) (Harnad, 2006).

Unfortunately, many authors are apprehensive of the copyright issues. As many commercial publishers are slowly yielding to the pressure of OA movement, authors should be aware of the freedom of publishing their research work. There is nothing wrong with copyright as long as rightly practiced. It is the 'transfer of rights' that authors should concern themselves with. In any case, the author should only grant a non-exclusive right to publish and disseminate their articles. During the period of euphoria of the emergence of democracies, many founding fathers of democracy believed that free (universal) education will help in building nations. Similarly, universal access to information will help in building a better world bridging the gap between so called developed and developing nations.

References

- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities.
Conference on Open Access to Knowledge in the Sciences and Humanities held from 20 – 22 October, 2003, Berlin. <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>
- Budapest Open Access Initiative. Meeting organised by the Open Society Institute (OSI) held at Budapest on December 1-2, 2001. <http://www.soros.org/openaccess/>
- Coyle, K. 2002. Open source, open standards. *Information Technology and Libraries* 21(1): 33-36.
- eGovernment News – 28 June 2004 – Denmark/EU & Europe-wide – Interoperability. <http://europa.eu.int/idabc/en/document/3132/333>
- Harnad, Stevan (2006). Optimizing OA Self-Archiving Mandates: What? Where? When? Why? How? <http://eprints.ecs.soton.ac.uk/13098/01/arch.html>.
- LOCKSS: How it works. http://www.lockss.org/lockss/How_It_Works
- Open Access News: Who, what & why. BOAI. <http://www.biomedcentral.com/openaccess/www/?issue=4>

Open Access Now: What, Who and Why?. SPARC.

<http://www.biomedcentral.com/openaccess/www/?issue=3>

Ray English and Peter Suber. (June 2006). Public access to federally funded research: The Cornyn-Lieberman and CURES bills. *C&RL News*, June 2006. Vol. 67, No.

<http://www.ala.org/ala/acrl/acrlpubs/crlnews/backissues2006/june06/fedfundedresearch.htm>

Rhyno, Art. Using open source systems for digital libraries. Westport: Libraries Unlimited, 2004. pp26.

Robin Peek (January 2007). The Federal Research Public Access Act of 2006. *Information Today*, January 30, 2007. <http://www.infotoday.com/newsbreaks/nbo60508-2.shtml>

Vilbrandt, T., et al. 2004. Cultural heritage preservation using constructive shape modeling. *Computer Graphics Forum* 23(1): 25-41.

Weller, Anne C. (2001). *Editorial Peer Review: Its Strengths and Weaknesses*. Medford, NJ: Information Today.