

Improving the visibility of Indian Research: An Institutional, Open Access Publishing Model

T.B. Rajashekar
National Centre for Science Information (NCSI)
Indian Institute of Science
Bangalore 560 012 (India)
(E-Mail: raja@ncsi.iisc.ernet.in)

(Position paper prepared for discussion in the Indo-US Workshop on Open Digital Libraries and Interoperability, June 23-25, 2003)

Abstract:

There is growing concern in India that Indian research, measured in terms of international publications, is losing its competitive edge over the past decade. Studies also indicate that research published in Indian sources is poorly cited compared to research published in international sources. Key challenge is how India can reciprocate the information flow and significantly improve access to, and hence impact, of its research? We propose an institutional level open access publishing model as a possible approach to improve the situation. Specifically, we propose that India evolves a network of distributed, inter-operable institutional digital research repositories covering, to begin with, research output of academic institutions and public-supported research laboratories. Each institutional repository will capture, preserve and provide open (free) access to its research output. More importantly, by adopting the inter-operability framework, research publications in these institutional repositories will be accessible online nationally and internationally in a seamless manner. We define open access publishing to include both “public domain” and “open access” scholarly material. Open access movement is driven by several enabling technologies, including open source digital library software; metadata standards and protocols for interoperability, particularly the OAI-PMH protocol. There is significant potential for open access publishing in India, considering the large number of public supported universities, institutions of higher learning and research laboratories in the country. Although these institutions produce substantial quantity of research output, only a very small portion gets published in formal medium like journals and conferences. It would appear that there is a significant case for improving the ‘research capacity’ of India by exploiting open access publishing technologies and bring much of the research material to the global online environment. There are several examples of freely available online scholarly Indian resources today, but these need to be brought under the inter-operability framework. eprints@iisc, the eprint archive of Indian Institute of Science, is probably the first OAI compliant institutional repository set up in India. We briefly discuss the proposed OAP system for India, including strategy for its development and implementation. We identify key challenges and issues that need to be met and resolved in the process. We also point to possible areas of collaboration in this area.

1. The problem:

There is growing concern that India is losing its competitive edge in research over the past decade or so compared to other Asian countries like China and Korea, measured in terms of research publications in international journals in science and technology. Studies also indicate that Indian research publications, published in Indian sources (e.g. journals) are poorly cited compared to those published in international sources. Two information related causes among others (for e.g. relatively poor infrastructure for research and inadequate research funding) often mentioned for this situation are: a) lack of quick and convenient access to international research publications, and b) lack of quality publishing channels for local research, at institutional and national level, and poor global visibility of such publications.

Thus, in the context of this workshop, two prime information related challenges faced by Indian research are:

1. How do we improve local access to global research? and,
2. How do we improve global access to local research?

While both the challenges are equally important, this position paper mainly addresses the second challenge and proposes an institutional level open access publishing model as a possible approach to improve the situation. We first very briefly touch upon the first challenge.

2. Local access to global research:

Several library consortia have been set up in India over the past couple of years, to obtain site licenses and enable desktop Internet access to scholarly e-resources like e-journals and databases. Prime examples include the INDEST consortium under the MHRD (Ministry of Human Resources) initiative covering leading academic institutions like IISc (Indian Institute of Science), IITs (Indian Institutes of Technology), IIMs (Indian Institutes of Management) and NITs (National Institutes of Technology); and CSIR (Council for Scientific and Industrial Research) consortia covering about 40 national laboratories. INFLIBNET, with UGC (University Grants Commission) support, is another major initiative covering universities. An interesting locally developed initiative in providing single-point access to global journal literature is the e-journal portal and gateway J-Gate from Informatics India. JCCC, an intranet product from the same company, goes a step further in facilitating sharing of journal resources held by all consortium libraries.

While these initiatives are expected to significantly improve the quality and productivity of teaching, research and learning, they have also given rise to several issues that need to be addressed urgently: perpetual access to e-resources (archiving); usage monitoring and impact assessment; usage promotion; personalization; and access management. It is to be hoped that this workshop will facilitate exchange of ideas and experiences between US and Indian researchers and practitioners in these areas.

3. Open access publishing (OAP) and global access to Indian research:

Key challenge is how India can reciprocate the information flow and significantly improve access to, and hence impact, of its research?

We believe that India can bring substantial portion of its research, including both published-but-poorly accessed and unpublished research, into the international scene by joining the rapidly evolving open access publishing (OAP) movement, which is itself driven by developments in inter-operable digital libraries and related technologies.

We propose that India evolves a network of distributed, inter-operable institutional digital research repositories covering, to begin with, research output of academic institutions and public-supported research laboratories. Each institutional repository will capture, preserve and provide open (free) access to its research output. More importantly, under the inter-operability framework, research publications in these institutional repositories will be accessible nationally and internationally in a seamless manner. Further, we also propose that other national research resources, for example science journals, be brought under this open access framework.

We broadly define OAP to mean provision of free online access to quality ‘Public domain’ and ‘Open access’ scholarly material (data and information). Several experts have defined open access. We subscribe to the definitions recently put forth in the “International Symposium on Open Access and the Public Domain in Digital Data and Information for Science” during 10-11 March 2003, organized by ICSU, CODATA, UNESCO and ICSTI. “Public domain” is defined in legal terms as sources and types of data and information whose uses are not restricted by statutory intellectual property (IP) laws and are available to the public for use without authorization. “Open access” is defined as proprietary information that is made openly and freely available on the Internet or on other media by the rights holder, but that retains some or all of the exclusive property rights that are granted under statutory IP laws. Open access may be provided by all types of public and private sector sources.

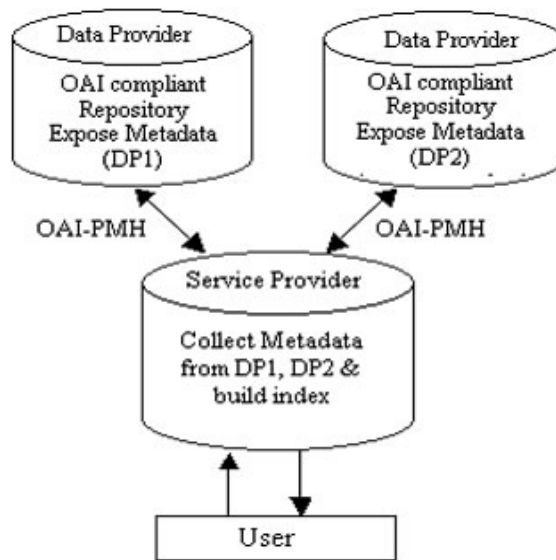
There is a rapidly emerging global movement in support of open access publishing. These include Open Archives Initiative (OAI), Budapest Open Access Initiative (BOAI), Scholarly Publishing and Academic Resources Coalition (SPARC), Free Online Scholarship (FOS), and Open Society Institute (OSI). Agencies like ICSU, UNESCO and ICSTI are increasingly in support of this movement. Several international workshops have also been held recently to discuss the technological, economic and legal aspects of open access.

4. Enabling technologies and framework for open access publishing:

Open access movement derives its strength from several enabling technologies and metadata-based inter-operability protocols that have become available recently. These include:

- Open source software for establishing and managing institutional digital repositories (e.g. Greenstone Digital Library Software from New Zealand Digital Library Project, DSpace from MIT, eprints archiving software from eprints.org, and CERN's CDSWare document server)
- Open source software for online journal and conference publishing (e.g. OJS system from the Public Knowledge Project, University of British Columbia, Canada)
- Metadata schemes and namespaces for dealing with digital objects of different user communities (e.g. Dublin Core for web-based publications; Encoded Archival Description (EAD) for Archives and special collections; Visual Resources Association (VRA) Core Categories for visual materials such as buildings, photographs, paintings, etc.)
- The OAI-PMH protocol for inter-operability (Open Archives Initiative Protocol for Metadata Harvesting)

With the free availability of such enabling technologies, it is to be expected that we are going to witness rapid growth in the number and variety of digital libraries. However, the much required 'glue' that is becoming increasingly acceptable for integration of these multitudes of digital libraries is the inter-operability framework of the Open Archives Initiative (<http://www.openarchives.org/>). The framework, through the OAI-PMH protocol for metadata harvesting, enables different digital libraries to expose their metadata for purposes of automatic metadata extraction from all participating digital libraries and build various services, including a central metadata index and search system as illustrated in the figure below. Such a metadata index enables users to quickly identify and select appropriate research material held in all participating institutions, and download specific publications online from the specific institutional repositories.



Interoperability using OAI-PMH requires that participating digital libraries are compliant to this standard. There are two key players under this model – Data Providers and Service Providers. Data providers maintain one or more repositories (web servers) that support the OAI-PMH as a means of exposing metadata. A data provider registers with OAI and publicize the fact that they have adopted the OAI-PMH. Service Providers issue OAI-PMH requests to data providers and use the metadata as a basis for building value-added services (e.g. central metadata index). They also register with OAI.

An example system where OAI-PMH is operational is the institutional e-print archive servers established using the eprint.org software (www.eprints.org). Eprints.org software is OAI-PMH compliant and defines a metadata standard, based on Dublin Core, for purposes of interoperability. Each institutional e-print server, registered with OAI, becomes a data provider. ARC is an example OAI service provider. ARC service (<http://arc.cs.odu.edu/>) provides a single point access, at metadata level, to content of all registered archives. The concept of inter-operable distributed institutional digital repositories is rapidly gaining ground today among the community of digital library developers.

We believe that India, given its proven IT competence, can take advantage of these developments to not only improve its research capacity but also to contribute to improving these technologies.

5. Open access publishing in India: Current status and potential:

There is significant potential for public domain and open access information initiatives in India, given the large number of public supported universities, institutions of higher learning and research laboratories in the country. There were an estimated 2,900 organizations involved in R&D in the country in 2001, in both commercial and non-commercial sector. These include a very large number of R&D labs under govt. science agencies in various domains (industrial, defense, agriculture, medical, biotechnology, environment, S&T, IT, space, energy, ocean development, etc.). In the academic sector there are close to 300 universities and institutions of higher learning, with graduate and research programmes. In 2001 there were an estimated 400,000 teachers in the country.

All these institutions produce substantial quantity of research output, only a very small portion of which gets published in formal medium like journals and conferences. In a study conducted by us at NCSI in IISc, using major bibliographic and citation databases, we identified that there were about 34,000 Indian research papers indexed by these databases during 2002. Web of Science database of ISI, which is often used for conducting scientometric studies, reported about 17,000 papers for the same year (WOS coverage is limited to about 5,900 so called ‘top ranking’ journals). Of these 17,000 papers, nearly 1/3rd (5,600) are from 49 institutions (IISc, 40 CSIR laboratories, 7 IITs, and TIFR).

It would appear that there is a significant case for improving the ‘research capacity’ of India by exploiting digital information technologies and bring much of the research material to the global online environment.

We are beginning to witness a few institutional and national level open access publishing and information dissemination efforts. Examples of open access publishing initiatives include: online access to scholarly journals (e.g. journals published by the Indian Academy of Sciences, Bangalore), theses (e.g. Vidhyanidhi project at University of Mysore), institutional e-print archives (e.g. eprints@iisc at Indian Institute of Science), books (e.g. Universal Digital Library project at Indian Institute of Science), and data sets (e.g. industrial micro-organism and biodiversity informatics at National Chemical Laboratory). There are also several initiatives that provide Internet access to open access material at bibliographic database level (e.g. NIC’s INDMED covering Indian medical journals). There are also a few innovative portal and gateway initiatives that attempt to integrate access to local and remote open access e-resources.

Probably the first OAI compliant institutional repository in India is eprints@iisc (<http://eprints.iisc.ernet.in>) at the Indian Institute of Science, Bangalore. It uses eprints.org archiving software. eprints@iisc facilitates the IISc researchers to self-archive their preprints and postprints. The archive supports online submission of variety of research publications and in various file formats like PDF, PS, Ms-Word, HTML etc. It has been registered with the service provider ARC, <http://arc.cs.odu.edu/>. We recently integrated the Greenstone Digital Library Software (GSDL) into this service, to support full-text searching. We have also developed variety of digital collections using GSDL, including UNICODE-based Indian language content. We have also recently completed successful installation of DSpace repository software.

6. Proposed OAP system for India:

These examples point to encouraging developments. While more such efforts are required, it is equally important that such open access systems are interoperable and facilitate seamless access nationally and globally.

We propose that India evolves a national network of distributed, inter-operable, open access digital repositories of S&T research material covering institutional digital repositories, science journals, conferences and other such scholarly material, and facilitate national and global access to these material.

Specifically, we propose that:

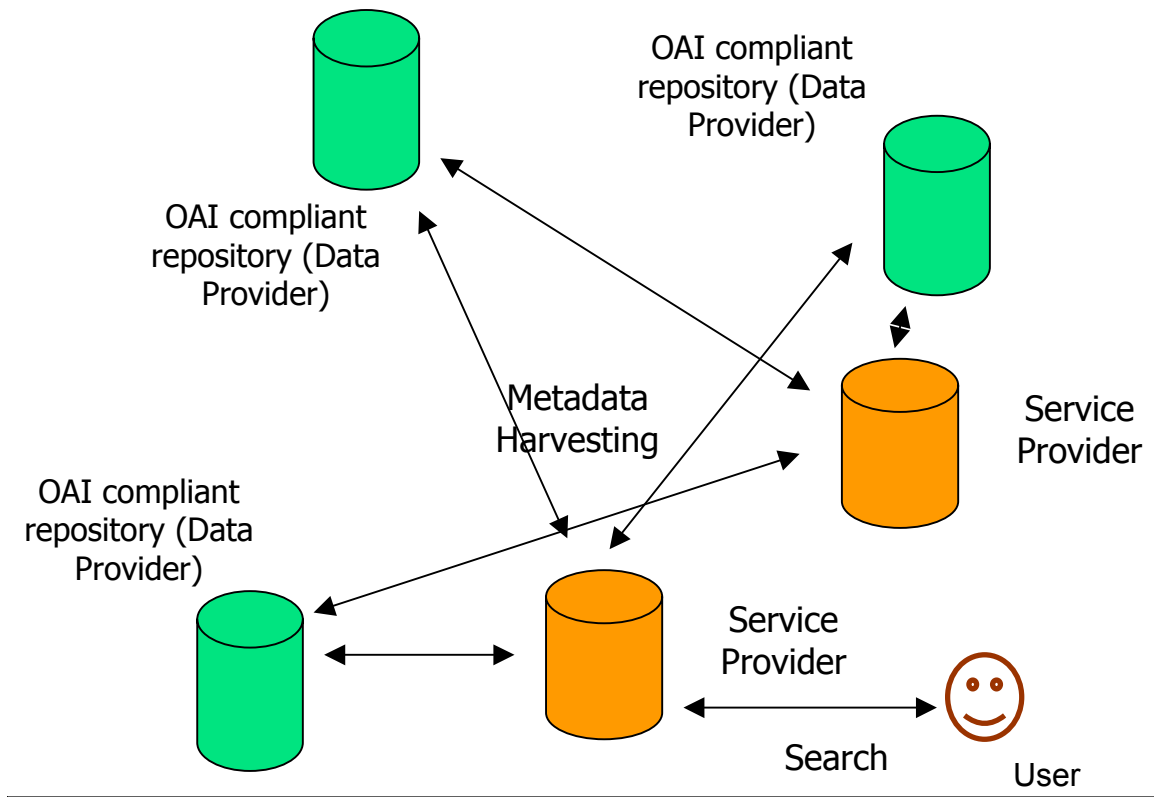
1. Academic and public-supported R&D institutions to set up online digital repositories of their research output, including institutionally organized conferences
2. Existing science journals published by academic and public-supported R&D institutions and professional societies to adopt open access publishing

3. New online-only open access journals be established, in areas of local strength (e.g. agriculture and medicine), and also graduate students journals
4. Wherever research material cannot be put online, either because it is not available electronically or because of copyright reasons, at least the metadata be put online

These organizations will act as ‘data providers’ under the OAI-PMH model. This requires that all these repositories are OAI compliant, and use appropriate metadata standards.

For the OAP system to be successful it is important that one or more ‘service providers’ emerge in the country to provide value-added services using metadata extracted from the data providers. The ‘service provider’ role can be taken up by one or more data providers themselves or could be new agencies, including private companies.

The OAP system is depicted in the figure below.



7. Development and implementation strategy:

The OAP system we have proposed above is of course very ambitious, considering the large number of potential institutions who will qualify as data providers, and also the financial reposes this would require. We propose a multi-pronged, phased approach for implementing the system.

1. Carry out ‘feasibility’ taking a few institutions within two administrative domains – for example IISc and couple of IITs within the MHRD domain and a few laboratories under the CSIR domain.
2. Demonstrate the operation of institutional repositories and their ability to act as ‘data providers’ to expose their metadata
3. Develop and demonstrate central search service through metadata harvesting (‘service provider’)
4. Firm-up the implementation mechanism based on this study.
5. Replicate the model across other institutions.
6. Expand the model to cover Indian science journals and other domain or resource specific national level data providers
7. Ensure that linkages exist with other global service providers

It is very important India adopts a very structured and planned approach to evolve this system, somewhat on the lines of UK. It appears that, at least initially, some agency should take on the role of a coordination agency to promote the concept, guide the implementation and support setting up of working models, and services. Such a centre could act as a national resource centre for open access publishing.

8. Challenges and issues:

Several technical, technological, economic, and legal challenges need to be met and issues resolved in the process of setting up the OAP system. We mention some of the important ones below, based on our experience in handling the IISc eprint archives and several open source digital library software.

1. What are the infrastructural requirements for establishing and operating institutional repositories (data providers) and service providers?
2. What content related standards and specifications (e.g. document genres and associated metadata, document formats, vocabulary) should be met by the repository software?
3. What authoring and publishing support is required at user (publisher) level?
4. What are the essential and desirable features of an institutional repository system?
5. What peer review and quality audit processes are required? How do we implement these?
6. What are acceptable workflows, procedures and responsibilities related to content management?
7. How do we provide OAI compliance to legacy repository systems (e.g. science journals)?
8. How do we simplify the submission and metadata assignment process? What submission processes can be automated?
9. Who will set up and manage the institutional repository?
10. How do we ensure perpetual availability of the repository and its contents (preservation)?
11. How do we encourage researchers to archive their research findings to the repository?

12. Who owns the copyright of content in the repository? Authors? Institution? Both?
13. How do we monitor the usage of content?
14. How do we measure the impact (ROI) of repositories?
15. What value-added services can be provided by harvesting content from repositories? Who will do this?
16. How does the OAP system exist along with the traditional publishing system?
17. What is the role of libraries in setting up and managing institutional repositories?
18. How institutions integrate access to variety of e-resources, including licensed external resources, free web resources and institutional resources? What is the role of institutional portals in achieving this integration?

We hope to learn from the rich experience US researchers and practitioners have had in these areas and contribute to future developments in a collaborative manner.

9. Benefits of OAP system to India:

Academic institutions and R&D labs generate significant number of internal research publications, including technical reports, manuals, guides, progress reports, presentations, etc., apart from preprints and postprints of papers in refereed journals and conferences. These publications contain very valuable and often detailed information, such as data, observations, conclusions, analysis, best practices, etc. In some sense these publications constitute the key research output and ‘intellectual capital’ of these organizations.

Only a portion of this research output is today catalogued and physically maintained in Indian libraries. Not only the organization does not have an integrated view of its ‘intellectual capital’, but a much more significant limitation is the inability of an Indian researcher to access the collective pool of research publications available in all the organizations.

Inability to efficiently and effectively capture, preserve and access the output of internal research has significant negative impact on R&D productivity of researchers in Indian academic and R&D labs.

We believe that establishment of inter-operable institutional digital repositories as proposed here will help in overcoming many of these shortcoming. Additional benefits of OAP system include:

1. Help researchers in establishing priority for their research findings and thereby protect their intellectual property
2. Remove access barriers
3. Enhance national research capacity
4. Provide global platform for local research and hence improved visibility
5. Facilitate improved research collaboration and information flow
6. Bring together intellectual output of an organization which otherwise gets scattered under the traditional publishing system

7. A well organized and content rich institutional repository will help an organization to demonstrate its value more easily; will bring enhanced status and reputation, which in turn could mean improved funding support.

10. Areas of collaboration:

We can identify several possible areas of collaboration at research, implementation and operational level, from the challenges and issues mentioned earlier. These include technological, economic and legal areas. We look forward to collaborative research and development in related areas, including the following:

1. Content related standards and specifications (e.g. document genres and associated metadata, document formats, vocabulary, citations) for institutional repositories
2. Evolving necessary and desirable features for repository software, including evaluation of currently available open source software
3. Peer review and quality audit norms and processes
4. Developing OAI-PMH support for legacy open access systems
5. Developing national level, domain-specific or omnibus harvesting services
6. Preservation of repository content for perpetual access
7. Promotion of repository with researchers for content publishing and access
8. Copyright
9. Usage monitoring and impact (ROI) measurement studies
10. Integration of institutional repositories with other institutional e-resources and e-services
11. Integration with traditional publishing systems
12. Access management
13. Personalization of e-information services

11. Conclusion:

Though this position paper is largely an Indian perspective, we believe that research, development, implementation and deployment of OAP systems will be of significant interest and benefit to both the countries.
