

# Metadata Harvesting Services in India

Dr (Smt) Mangala Hirwade  
Librarian  
Shivaji Science College  
Congress Nagar, Nagpur  
[hirwade2004@indiatimes.com](mailto:hirwade2004@indiatimes.com)

Dr. Anil Hirwade  
Patent Information System  
Seminary Hills, Nagpur-6  
[hirwade2004@indiatimes.com](mailto:hirwade2004@indiatimes.com)  
(M) 9373217966

## Abstract

*Today more and more information is created in digital form. One of the key challenges faced by today's information managers is the need to inter-relate different sources and types of information with different formats, data structures and description standards. Using metadata to record data about information sources allows an initial assessment of compatibility and provides an avenue for merging information or for exchanging information between systems. Another service has gained momentum nowadays is the Metadata Harvester. Metadata Harvester provides indexes or harvests metadata, from different open access archives and open access journals. This paper describes Metadata, Open Access Initiative Protocol for Metadata Harvesting (OAI-PMH) and major Metadata Harvesting Services in India.*

Keywords: Metadata, Metadata Harvesting, *Open Access Initiative Protocol for Metadata Harvesting (OAI-PMH, Metadata Harvesting Services in India Search Digital Libraries (SDL), SJPI Cross Journal Search Service SEED (Search Engine for Engineering Digital-repositories) Open J-Gate and Knowledge Harvester@INSA*

## 0. Introduction

There has been a dramatic change in the world of learning, scholarship, business and governance brought about by Information and Communication Technologies. The ways and means of creating, accessing, distributing and managing information not only in text but in other forms such as audio-video and multimedia materials have undergone major changes. Today more and more information is created and managed in digital form.

One of the key challenges facing information managers today is the need to inter-relate different sources and types of information, whether it is in an internet search across a range of resources with different formats, data structures and description standards or an e-commerce system that needs to exchange data between proprietary applications in order to complete a transaction. Using metadata to record data about information sources allows an initial assessment of compatibility and provides an avenue for merging information or for exchanging information between systems.

Interoperability is the ability of two or more systems to exchange information and to use the information that has been exchanged. National Information Standards Organization (NISO) defines interoperability as "the ability of multiple systems, using different hardware and software platforms, data structures and interfaces, to exchange and share data" [1]

Providing access to information free of charge in electronic formats is the concept that is gaining momentum. Open access holds the promise to remove price and permission barriers to the scientific communication by using Internet. Open access literatures are available in open access journals, institutional repositories, subject repositories, digital archives and so on. Another service has gained momentum nowadays is the Metadata Harvester. Metadata Harvester

provides indexes or harvests metadata, from different open access archives and open access journals.

### 1. Metadata

Metadata can be defined as "data about data" describe the content, quality, condition, and other characteristics of data. Metadata is vital in helping potential users to find needed data and determine whether a data set will meet their needs before they spend the time and money to obtain and process it. [2]

### 2. Open Access Initiative Protocol for Metadata Harvesting (OAI-PMH)

Open access works are scattered across many disciplinary archives, institutional e-print archives, institutional repositories and open access journals. Therefore, it is difficult for scholars to locate all needed works on a particular subject.

One important international movement to solve this problem is the Open Archives Initiative (OAI), which aims to develop and promote the use of a standard protocol, known as the Open Archives Metadata Harvesting Protocol (OAMHP), designed for better sharing and retrieval of e-prints residing in distributed archives. [3]

### 3. Major Metadata Harvesting Services in India

A metadata harvesting service harvests or indexes metadata from OAI-compliant archives or repositories through harvesting software that supports a protocol known as OAI-PMH (Open Access Initiative Protocol for Metadata Harvesting). Some Indian institutions have been experimenting with metadata harvesting services and installed metadata harvesters. Major metadata harvesting services in India are [4]

- Search Digital Libraries (SDL)
- SJPI (Scientific Journal Publishing in India) Cross Journal Search Service
- SEED (Search Engine for Engineering Digital-repositories)
- Open J-Gate
- Knowledge Harvester@INSA

### 4. Details about Metadata Harvesting Services in India

Table No. 1 covers general information of Metadata Harvesting Services in India like URL, host and software used.

Table No. 1: Metadata Harvesting Services in India

Sr. No.	Name	URL	Host	Software used
1	Search Digital Libraries (SDL)	<a href="http://drtc.isibang.ac.in/sdl">http://drtc.isibang.ac.in/sdl</a>	DRTC, Bangalore	PKP System
2	SJPI Cross Journal Search Service	<a href="http://144.16.72.144/harvester/">http://144.16.72.144/harvester/</a>	NCSI, IISc	PKP System
3	SEED	<a href="http://eprint.iitd.ac.in/seed/">http://eprint.iitd.ac.in/seed/</a>	IIT, Delhi	PKP System
4	Open J-Gate	<a href="http://www.openj-gate.com/">www.openj-gate.com/</a>	Informatics India Ltd.	----
5	Knowledge Harvester@INSA	<a href="http://61.16.154.195/harvester/">http://61.16.154.195/harvester/</a>	INSA	PKP System

#### 4.1 Search Digital Libraries (SDL)

SDL harvests metadata from open access digital libraries and e-journals in the area of Library and Information Science. It uses PKP (Public Knowledge Project) software. The PKP OAI Harvester allows creating a searchable index of the metadata from Open Archives Initiative-compliant archives, such as sites using Open Journal Systems or Open Conference Systems. Harvester2 currently has **13572** papers from **21** archive(s) indexed. The details are provided in Table 2. [5]

Table No. 2: Open Access archives indexed by SDL

Sr. No.	Name of the archive indexed by SDL	No. of records
1	Librarians Digital Library (LDL)	251
2	Australian Library and Information Science Association (ALIA)	20
3	Bibliothèques Virtuelles Humanistes	98
4	CALTECHLIB	44
5	CCSD: Sciences de l'Information et de la Communication, France	693
6	CNR Bologna Research Library, Italy	58
7	Diálogo Científico utiliza, Brazil	656
8	DLIST, University of Arizona	724
9	E-LIS: E-Prints in Library and Information Science	2147
10	OCLC Research Publications	852
11	University of North Carolina, USA	296
12	WWW Conference Archive EPrint servers	504
13	Australian Library and Information Association e-prints	2792
14	OPUS Volltextserver des Berufsverbands Information Bibliothek (BIB)	225
15	Public Knowledge Project EPrint Archive	48
16	UFMG Database Group	4142
17	University of Technology, Sydney	22
18	Pokemon crater	0
19	Vespa scooter	0
20	Coin sorter	0
21	Kimbo fight video	0
Total		13572

#### 4.2 SJPI Cross Journal Search Service

The SJPI Cross Journals Search Service is part of the SJPI project. The goal of this project is to improve the accessibility of scientific literature published in Indian journals by introducing an indexing system. A sampling of journals are made OAI compliant using Open Journal System (OJS). They are harvested by the PKP harvester at this site. This demonstrates a search service

across multiple journals from a single point. This harvester indexes articles published in these journals and provides various ways of accessing them. Simple searching of keywords in Author(s), Title, Abstract or Index Terms is possible. Advanced Search and Browse facility is available.

The SJPI Harvester currently has **1047** papers from **13** journals indexed. It uses PKP (Public Knowledge Project) software. The list of journals indexed is given in Table No. 3. [6]

Table No. 3: List of journals indexed by SJPI

Sr. No.	Name of the journal indexed by SJPI	No. of papers
1	Bulletin of Materials Science	105
2	Current Science	229
3	Journal of Astrophysics and Astronomy	45
4	Journal of Biosciences	86
5	Journal of Chemical Sciences	51
6	Journal of Genetics	60
7	Journal of the Indian Institute of Science	46
8	Pramana - Journal of Physics	158
9	Proceedings Earth Planetary Sciences	46
10	Proceedings Mathematical Sciences	62
11	Resonance	62
12	Sadhana	54
13	SRELS Journal of Information Management	43
Total		1047

#### 4.3 SEED (Search Engine for Engineering Digital-repositories)

The IIT Delhi has developed a number of discipline-specific Research Support Tools (RST), which accompanies individual research studies indexed from e-journal and conference paper websites covering a wide range of disciplines. The RST utilizes the study's metadata to search relevant open-access databases for related studies, theory, news, policies, and other resources, as well as offering access to the study's metadata and citation, to a personal portfolio, and to email and comment options. The software used is again PKP system.

The Seed currently has **6176** papers from **4** archives indexed. Table No. 4 shows the list of archives indexed by SEED. Simple and advanced search facility along with browsing capabilities is available. [7]

Table No.4: List of archived indexed by SEED

Sr. No.	Name of the archive indexed by SEED	No. of papers
1	Earthquake Engineering	310
2	Dspace@NITR	262
3	Eprints@IISc	5005
4	Eprints@IIT Delhi	599
Total		6176

#### 4.4 Open J-Gate

Open J-Gate is an electronic gateway to global journal literature in open access domain. Launched in 2006, Open J-Gate is the contribution of Informatics (India) Ltd to promote OAI. Open J-Gate provides seamless access to millions of journal articles available online. Open J-Gate is also a database of journal literature, indexed from 3000+ open access journals, with links to full text at Publisher sites.

##### Salient features of Open J-Gate are

- It is a Portal with the largest number of e-journals. It indexes articles from 3000+ academic, research and industry journals. More than 1500 of them are peer-reviewed scholarly journals.
- Links to one million+ open access articles and this number is growing with 300000+ new articles added every year. Full-text links are regularly validated.
- The Open J-Gate site is updated every day.
- All journals are classified in a three-level hierarchical system to provide for better relevancy in search results.
- Users can browse the TOC (Table of Content) of latest issue and the back issues.
- Database allows various search options for the user's convenience. The subscriber can search by Title, Author, Abstract, Author's Address/Institution, Keywords, etc.

At present Open J-Gate indexes **3721** open access journals from different subject categories viz. Agricultural and Biological Sciences, Arts & Humanities, Basic Sciences, Biomedical Sciences, Engineering & Technology, Library & Information Sciences and Social & Management Sciences. [8]

#### 4.5 Knowledge Harvester@INSA

Knowledge Harvester@INSA is an experimental initiative from INSA indexes three archives viz. African Journal Online (currently has 248 journals), Archive of European Integration (currently has 5046 documents) and INSA (Indian National Science Academy) Digital Library. [9]

### 5. Conclusions

Metadata is a key part of the information infrastructure necessary to help create order in the chaos of the Web, infusing description, classification, and organization to help create more useful stores of information. Sources of metadata, like the sources of the resources themselves, will be of different quality and organized around different purposes to reflect the different objectives and business models of information providers.

The Open Archives Metadata Harvesting Protocol opens many new possibilities which are yet to be explored. OAI metadata harvesting offers a new bridge to bring new innovation in networked information services and applications, out of the research community more rapidly.

Researchers who want to explore new ways of organizing, presenting or using the large data resources now have a standardized way of extracting content without much disruption or cost to existing operational systems. Metadata harvesting services are powerful mechanism for enabling development of new applications and services that have never been possible.

## References

1. National Information Standards Organization (NISO). "Understanding Metadata"  
<http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>  
(Last accessed on 18.08.2006)
2. What is Metadata?  
[www.web2.gov.mb.ca/mli/glossary/m\\_z.html](http://www.web2.gov.mb.ca/mli/glossary/m_z.html) (Last accessed on 22.08.2006)
3. Rajshekhar, T.B. "Open Access initiatives in India"  
<http://www.istl.org/05-spring/article3.html#8> (Last accessed on 24.08.06)
4. Ghosh, S.B. and Das, Anup Kumar "Open access and institutional repositories – a developing country perspective: a case study of India"  
[http://www.ifla.org/IV/ifla72/papers/157-Ghosh\\_Das-en.pdf](http://www.ifla.org/IV/ifla72/papers/157-Ghosh_Das-en.pdf) (Last accessed on 22.08.2006)
5. Search Digital Libraries (SDL)  
<http://drtc.isibang.ac.in/sdl> (Last accessed on 26.08.2006)
6. SJPI Cross Journal Search Service  
<http://144.16.72.144/harvester/> (Last accessed on 26.08.2006)
7. SEED (Search Engine for Engineering Digital-repositories)  
<http://eprint.iitd.ac.in/seed/> (Last accessed on 26.08.2006)
8. Open J-Gate  
[www.openj-gate.com/](http://www.openj-gate.com/) (Last accessed on 26.08.2006)
9. Knowledge Harvester@INSA  
<http://61.16.154.195/harvester/> (Last accessed on 26.08.2006)