International Journal of Information Science and Management
Vol. 15, No. 1, 2017, 73-88

Designing Metadata Harvesting Framework for OAI-based LIS Repositories: A Prototype

Bijan Kumar Roy
Assistant Professor, Dept. of LIS,
The University of Burdwan, WB,
Corresponding Author,
Email: bijankumarroy@yahoo.co.in

Subal Chandra Biswas
Professor, Dept. of LIS,
The University of Burdwan, WB,
Email: scbiswas_56@yahoo.co.in

Parthasarathi Mukhopadhyay
Associate Professor, Dept. of LIS, The University of Kalyani, WB,
Email: psmukhopadhyay@gmail.com

Abstract

The paper provides an overview over the functionalities of the proposed harvesting model (BURA – Burdwan University Research Archive) and describes the activities involved in harvesting resources from OAI compliant LIS (Library and Information Science) repositories. Repositories, registered in OpenDOAR database, containing 'Library and Information Science' as one of the key subject have been considered and finally 117 repositories have been short listed on the basis of framed criteria. The purpose of this paper is to report an integrated Web-enabled distributed harvesting model that can extract and update metadata index efficiently from any OAI-PMH driven repositories on global scale. The findings suggest that LIS repositories are not up to the global standards and lack contents compared to others disciplinary repositories. It should follow open standards and should respect essence of interoperability. The model may be an interoperability solution to the professional librarians in designing and developing federated search option for multiple repositories from a single-window search interface. Thus, it may be concluded that this proposed harvesting framework may work as a guiding tool to those who plan to set up new OAI-based service provider.

Keywords: Federated Search; Interoperability; Metadata Harvesting; Metadata Discovery; OAI-PMH; Open Access; Open Access Repository.

Introduction

Technological developments particularly in the field of Library and Information Science (LIS) have led to the emergence of new service paradigms. This development has become more complex with the advent of World Wide Web (WWW), Internet and digital resources. The new technologies, standards, and best practices have created a revolution in the resource generation, resource discovery and accessibility of information over the Web. Network technologies have helped professional librarian to lower many of the geographical barriers that impede access to public funded research outputs.

Many elite research institutions including universities across the world are facilitating open-access (OA) to their research outputs through respective open access repositories (OARs).

There is a growing need to document these OA knowledge resources acquired by different LIS organizations for future use. Information professionals expect to have a distributed system that enables access to separately managed collections in house as the number of on-line repositories available through the Web grows. The development of different sophisticated harvesting tools like OAI-PMH harvesters has fulfilled the goal of seamless access of resources from different heterogeneous, networked digital repositories through a single search interface.

In order to provide better Web-based services, it is the responsibility of the library professionals to establish OAI (Open Archive Initiative) services through open platform for dissemination of distributed research outputs under different systems. Federated Searching has long been a desirable goal for the scholarly community and initiatives like OAI has made it possible of timely dissemination of scholarly information. It (OAI) has developed Metadata Harvesting Protocol (PMH) that opens many new possibilities in the area of resource discovery and metadata harvesting from different OAI-PMH compliant repositories. With the current distributed technology, the users' expectations of harvesting metadata from domain-specific OA information system have become possible. In this context, federated search may be one solution for harvesting metadata across heterogeneous digital repositories in real-time with respect to user search request. This proposed mechanism (e.g. BURA) may be a powerful tool to the library professionals for enabling the development of new applications and services that have never before been possible in our field.

Literature Review

Crow (2002) suggested that systems (IR) must be able to support interoperability and able to harvest metadata through multiple search engines and other standard discovery tools in order to provide broader access to the academic and research community, users outside the institution. Several research studies (Hunter & Guy, 2004; Horwood et al., 2004; Ginsparg, Luce & Van de Sompel, 1999; Van de Sompel & Lagoze, 2000; Mazurek et al., 2006; Eaton, 2008; Alipour-Hafezi et al., 2010; Khazraee et al., 2011; Miller, 2004; Westell, 2006) explained the necessity of OAI Protocol for Metadata Harvesting (OAI-PMH) in digital repository environment.

Kellogg (2004) surveyed different open source OAI harvesting tools and PKP harvester proved an excellent metadata harvesting and presentation tool. Peset et al. (2007) highlighted the situation of the development of repositories that use the Open Archives Initiative (OAI) protocol for data collection. The other experts (Jerez et al., 2004) focused on the multi-faceted use of the OAI-PMH to access stored contents in repository architecture. Warner (2001) showed in practical through Perl code how OAI-PMH can be used to expose and harvest metadata from Archives. Pieper and Summann (2006) reported the activities of Bielefeld University Library in establishing OAI based repository servers and provides an overview of the functionalities of BASE (Bielefeld Academic Search Engine) and gives insight into the challenges that have to be faced when harvesting and integrating resources from multiple OAI servers. Leiner (1998) reported the role of NCSTRL (confederation of over 100 institutions) in the field of computer science material. The necessity of cross-archive search services is important and the development of UPS (a protype) is an initiative in this domain (Van de

Sompel et al., 2000). Multilingual information access and searching heterogeneous multilingual collections on the Web were covered by Powel and Fox (1998).

Many Indian authors reported the necessity of developing OAI-PMH compliant OARs (Sreekumar et al., 2007; Narang et al., 2005; Rajashekar, 2003; Singh & Pandita, 2005; Chand et al., 2004; Awasthi & Jaiswal, 2008; Roy, 2007, 2010, 2014b; Roy & Mukhopadhyay, 2011) and use of open archives (OAs) is one approach to achieve some degree of interoperability (Alexander & Gautam, 2004). During 2011 to 2016, couple of studies (Roy, Biswas & Mukhopadhyay, 2011, 2012a, 2012b, 2013, 2015, 2016a, 2016b, 2016c, 2016d) put emphasis on the development of OAI-based OARs to support OA to knowledge movement. Jayakanth, Sharada & Minj (2007) stressed the importance of having a union catalogue (a centralized database) of the participating libraries to make the inter library loan service more efficient. Another study (Deshmukh, Bhavsar & Bhavsar, 2012) covered OSS (Open Source Software) available for federated search and gave some federated search applications of public domain. Prasad and Guha (2005) discussed the need of interoperability in digital library arena and explored the potential of the OAI-PMH protocol and also illustrated its structural and functional models. Amin (2003) proposed for OAI compliant repositories and described OAI-PMH framework and its technical architecture. Couple of studies (Hirwade & Hirwade, 2006; Hirwade & Bherwani, 2011) described metadata, OAI-PMH and major metadata harvesting services available in India. Another study (Singh, Pandita & Dash, 2008) shared practical experiences and showed how metadata is harvested (extracted) from data providers (repositories) by service providers (search engines) using this model. Sarkar and Mukhopadhyay (2010) proposed a method of metadata harvesting from different OAI-PMH compliant repositories containing electronic theses and dissertations (ETDs). This study described in details harvesting framework using PKP harvester. In another paper they reported the designing of a prototype union catalogue of ETDs on health and medicine (Sarkar & Mukhopadhyay, 2012). Jayakanth and Minj (2012) shared their practical experiences in setting up a prototype metadata harvesting service using the PKP harvesting software for the OAI-compliant repositories in India.

Indian Scenario

Several prototypes cross search services based on metadata harvesting have already been established in the country. Some are subject based service providers and some are general in nature. For example, the Search Digital Libraries (SDL), http://drtc.isibang.ac.in/sdl/, is a federated search service related to the Library and Information Science discipline. SEED (Search Engine for Engineering Digital-Repositories - http://eprint.iitd.ac.in/seed/) is a federated search service for the engineering discipline. SJPI (Scientific Journal Publishing in India) indexes articles with the objectives of increasing accessibility of scientific literature published in Indian journals. Open J-Gate is another such initiative to global journal literature in open access domain. A comprehensive federated search service for all the OARs in the country has not been attempted as yet. CASSIR (Cross Archive Search Service for Indian Repositories - http://smart.ncsi.iisc.ernet.in/oai) is the first such attempt. It is an attempt by the Indian Institute of Science (IISc) to build and maintain metadata harvesting service for all the OARs from India. It has been developed using the Open Harvester System (OHS), a

FOSS (Free Open Source Software) metadata harvesting system developed by the Public Knowledge Project (PKP) to harvest the metadata.

Objectives of the Study

The specific objective of this study is to develop a distributed union catalogue of selected OAI-PMH compliant repositories in the field of LIS. The second objective is to design a set of parameters for selecting LIS repositories for metadata harvesting taking into consideration global recommendations and best practices. And finally, addresses the features supported by the service providers (Harvesters) for browsing, searching and presentation of search results along with proposing an interoperability solution for LIS repositories.

Scope and Coverage

This study is limited to only LIS repositories registered in OpenDOAR database. Repositories containing ' *Library and Information Science*' as one of the key subject have been considered. Selection of OARs is based on carefully crafted criteria mentioned in Table - 1. An exhaustive review and harvesting of existing OARs on LIS is out of the scope of this paper. BURA is basically a prototype.

Limitations of the Study

Some of the repositories (having fulfilled the criteria mentioned in Table 1) could not be incorporated in this harvesting model due to the following facts – i) *corresponding OAI base URLs were not working in few cases*; ii) *OAI base URLs were found invalid and were not responding against OAI-PMH Verbs*; iii) *OAI base URLs were responding but due to some technical problems metadata could not be harvested and metadata index could not be updated*; and iv) *stated URLs of the organization were found wrong in few cases*.

Methodology

Data for this paper was downloaded from the official websites of these metadata harvesting service providers during November - December, 2015. With respect to the strategy and methodology adopted to formulate the metadata harvesting framework, specific actions were taken into consideration – i) *study of OARs listed in OpenDOAR and ROAR databases*; ii) *study of global recommendations and best practice guidelines*; iii) *framing of criteria for selection of LIS repositories for metadata harvesting*; iv) *selection of LIS repositories from OpenDOAR database based on framed criteria (Table 1)*; and v) *demonstrating several key features and functionalities of the metadata harvesting framework (e.g. BURA).*

Selection Parameters of LIS Repositories

Finally, out of 117 LIS repositories (as on December, 2015), nineteen (19) repositories from OpenDOAR database have been selected for metadata harvesting against carefully crafted criteria (Roy, Biswas & Mukhopadhyay, 2015).

Table 1
Parameters for selection of LIS OARs

Criteria / Parameters	Conditions (order of preferences)
Support for OAI-PMH (version 2)	Available for metadata harvesting
Number of objects uploaded	10,000 and above
Language covered	English (at least)
User Interface	Available in English
Type of software used	Distributed architecture
Metadata Schemas	Open Standard
Data type	Textual (at least)
Data format	Variety of format supported
Data availability	Abstract or Full text
Licensing model	Standardized license
Access pattern	Open access/embargo
Browsing and Sorting	Metadata elements
Searching	Simple and Advanced (at least)

BURA: A Prototype

BURA is basically a prototype union catalogue of OAI based repositories that support harvesting updated metadata efficiently from a single-window search interface to facilitate retrieval of OA resources across the globe. It is a centralized repository of metadata with distributed data sources, the objectives of which is to harvest metadata from various data providers and making them available through a single access point. The designing part of this harvesting framework has been described in the research work (Roy, 2014a) and has also been reported in another work (Roy, 2015). It is basically a localized resource discovery service model based on open standard and open source software (OSS). A unified search interface has been developed using PKP2 harvesting software that can harvest and retrieve OAI metadata in a variety of schemas though single access point. This model can also be used in any other disciplines, subjects or objects type.

Administrative Operations

The system allows performing various administrative operations such as addition, deletion, sorting, updating metadata index etc. and can be done any time using the following interface (Fig. 1).

Figure 1: Administration Interface

Like other system, any time new repository can be added (Fig. 2). The system also allows editing repository related information.

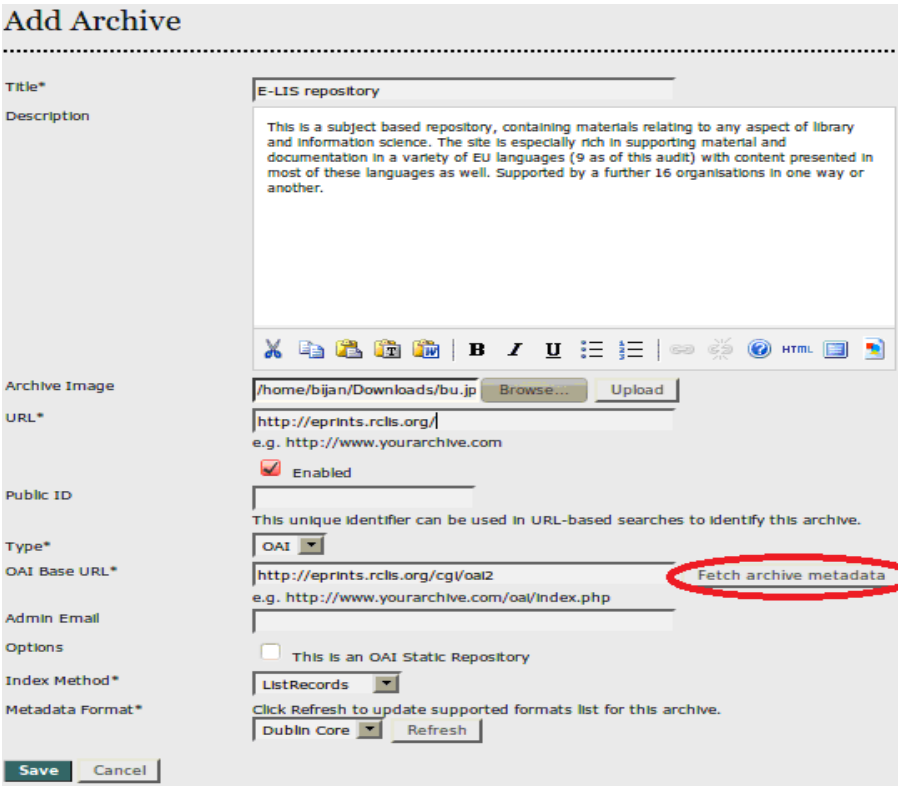

Figure 2: Addition of Repository

Another advantage of this system is that it supports updating metadata index by selecting any 'Sets' (such as harvesting by '*Date*' though it is optional) (Fig. 3).

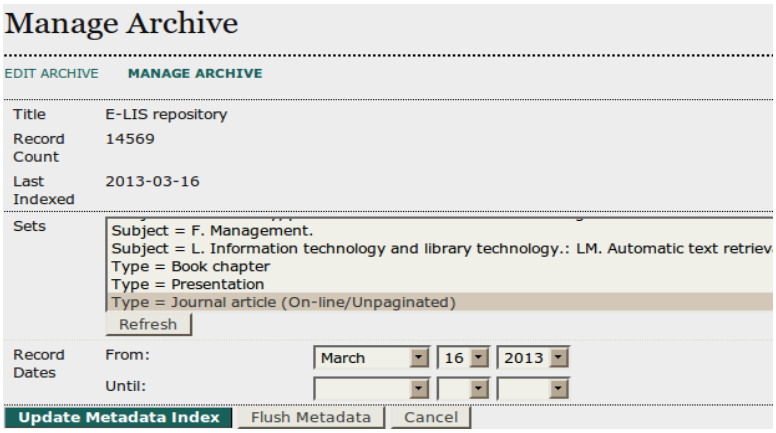

Figure 3: Update Metadata Records

Users have the flexibility of sorting and grouping repository(s) by *date*, *subject*, *repository type* or *archive name*. Even new sorting order can be re-established (Fig. 4).

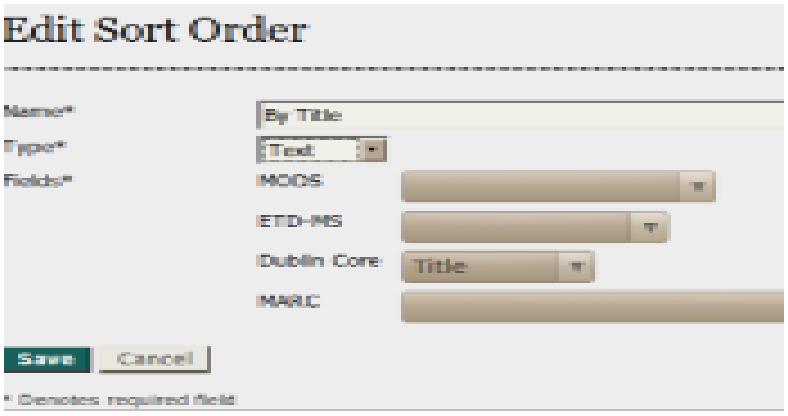

Figure 4: New Sorting Order

Browsing and Searching

BURA user interface supports browsing repositories on a record-by-record basis. Here, end-user can browse any repository (Fig. 5) and can limit his/her search within a single repository or group of repositories.

The system supports localized searching of metadata in two ways – *simple* and *advanced*. Advanced search provides a way for a user interactively to pick up the controlled values defined by specific archives via the search interface. Even search can be filtered by DCMES like title, author, date etc. In addition, system supports keyword search allows users to search all metadata fields across archives. Keyword search provides a simple and familiar way to conduct search across all archives, and the input can include Boolean operators (AND, OR, NOT).

Figure 5: Browsing Repositories

As sated earlier, user interface of BURA supports two different search capabilities - i) *simple search*, and ii) *advanced search*. User can search a particular repository or all repositories. This interface allows users to perform fielded, phrase and other hierarchical searches across these metadata sources. For example, repository '*E-LIS' (E-Prints for Library and Information Science)* has been selected here (Fig. 6).

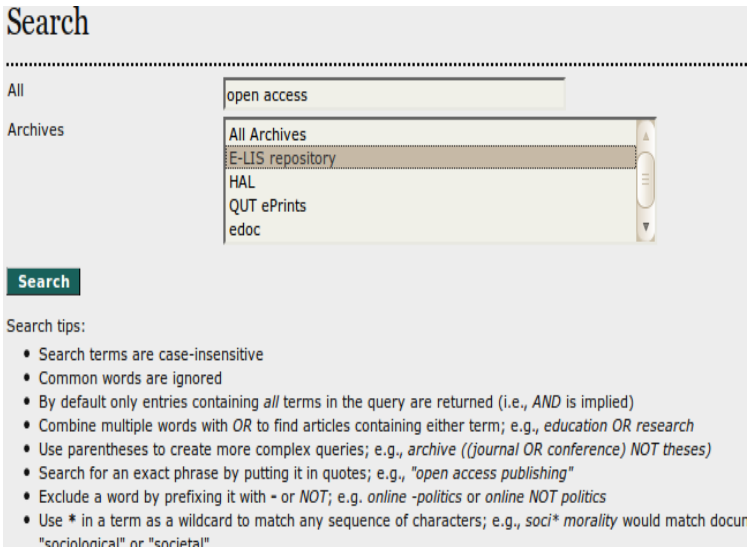

Figure 6: Searching specific Repository

Once the users enter their search term/criteria (here '*open access*' in Fig. 6) and perform

the search, the results summary page (Fig. 7) displays the total number of records found against the search term e.g. *open access*. This summary page also allows users to sort the results by *author*, *title* etc. It also shows a link to the full metadata record along with the full text. The search results against the search query '*open access*' (Fig. 6) are given below in Fig. 7.

Figure 7: Display of Search Results

Fig. 8 displays the records in details. When the user clicks the "*View Record*" link (Fig. 7) found on the summary page, the full record will be displayed. Even user can have full text clicking on '*View Original*' (Fig. 7).

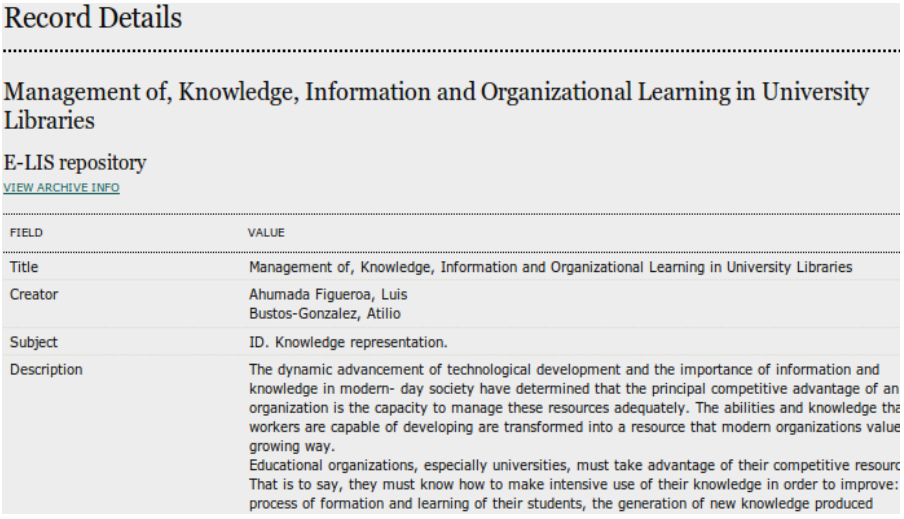

Figure 8: Short Display of Record details

Findings

After analyzing LIS repositories (both selected and rejected), the major findings have been grouped under the two broad headings viz. *General and Technical*. General section is basically based on Annex 1 and the key findings are - i) *number of uploaded documents is very low in compare to other disciplinary repositories*; ii) *all the repositories (except E-LIS & Ozone cover only LIS discipline) are multidisciplinary in nature and cover different subjects*; iii) *full text are not available in all cases and contents is accessible only to the community members (e.g. registered users)*; iv) *three repositories are from Asian countries*; and v) *most of the repositories don't have open access (OA) self archiving policies (see column 9 of Annex 1)*. The technical section covers some of the system related features of data providers such as - i) *majority of repositories don't respect interoperability and their corresponding OAI base URL are not functioning (OAI-PMH base URLs have been tested against OVAL validation tool - http://oval.base-search.net/)*; ii) *sometimes metadata could not be harvested from data provider e.g. repository*; iii) *majority of the data providers use the Dublin Core (DC) as the default metadata format for displaying metadata*; and iv) *majority of the repositories use metadata control device and use controlled metadata.*

Conclusions

There is possibly no such mechanisms of harvesting metadata from different OAI-PMH compliant repositories containing variety of digital objects in different forms and formats in the field of LIS. This framework has been developed using different open source software (OSS), open standard and open technologies and is the first of its kind in this domain. The system has made cross-library search for publications is now a reality. This system can be integrated with any national network and has the capability to include new OAI-PMH compliant OARs. Such a centralized system can only showcase the research outputs globally through a single access point and may help to exchange OA knowledge objects in a more faster, cost-efficient and cost-effective manner.

References

Alexander, M. L., & Gautam, J. N. (2004). Interoperability and Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). *2nd International CALIBER* (February 11-13, 2004, New Delhi) (pp. 338-345). Ahmedabad: INFLIBNET.

Alipour-Hafezi, M., et al. (2010). Interoperability Models in Digital Libraries: An Overview. *Electronic Library*, 28(3), 438 – 452.

Amin, S. (2003). The Open Archives Initiative Protocol for Metadata Harvesting: An Introduction. *DRTC Workshop on Digital Libraries: Theory and Practice* (March, 2003, DRTC, Bangalore). Paper H. Bangalore: DRTC. Retrieved January 5, 2014, from http://drtc.isibang.ac.in/xmlui/bitstream/handle/1849/40/H_OAIPMH_saiful.pdf?sequence=2

Awasthi, S. & Jaiswal, B. (2008). Open Archives Metadata Harvesting: An Overview. *6th Convention of PLANNER* (November 6-7, 2008, Nagaland University, Nagaland) (pp. 174-180). Ahmedabad: INFLIBNET.

Chand, P., et al. (2004). Institutional Repositories, Open Access Movement and OAI-PMH

Compliant Software. *2nd Convention of PLANNER* (November 4-5, 2004, Manipur University, Imphal) (pp. 52-64). Ahmedabad: INFLIBNET.

Crow, R. (2002). *The Case for Institutional Repositories: A SPARC Position Paper*. Washington, D. C: The Scholarly Publishing and Academic Resources Coalition.

Deshmukh, S., Bhavsar, S. & Bhavsar, S. (2012). Open Source Software for Federated Search. *DESIDOC Journal of Library and Information Technology*, 32(5), 427-430.

Eaton, J. (2008). Using the Open Archives Initiative Protocol for Metadata Harvesting. *Program: Electronic Library and Information Systems*, 42(4), 450 – 452.

Ginsparg, P., Luce, R. & Van de Sompel, H. (1999). *The Open Archives Initiative aimed at the further promotion of author self-archived solutions*. Retrieved August 8, 2015, from http://www.openarchives.org/meetings/SantaFe1999/ups-invitation-ori.htm

Hirwade, M., & Hirwade, A. (2006). Metadata Harvesting Services in India. *Library Herald*, 44(4), 275-282.

Hirwade, M. A. & Bherwani, M. T. (2011). Metadata Harvesting: Tools and Services in India. *SRELS Journal of Information Management*, 48(4), 389-398.

Horwood, L., et al. (2004). OAI Compliant Institutional Repositories and the Role of Library Staff. *Library Management*, 25(4/5), 170-176.

Hunter, P. & Guy, M. (2004). Metadata for Harvesting: The Open Archives Initiative, and How to Find Things on the Web. *The Electronic Library*, 22(2), 168–174.

Jayakanth, F., Sharada, B. & Minj, F. (2007). An OAI-based approach to build and maintain union catalog of OPACs. In A.R.D. Prasad & D.P. Madalli (Eds.), *Proceedings of the International Conference on Semantic Web and Digital Libraries* (February 21-23,, 2007, DRTC, Bangalore) (451-458). Bangalore: DRTC.

Jayakanth, F. & Minj, F. (2012). Federated Search Service for OAI-compliant Open-Access Repositories in India. *International Conference on Trends in Knowledge and Information Dynamics* (July 10-13, 2012, DRTC, Bangalore) (pp. 1-8). Bangalore: DRTC. Retrieved April 27, 2014, from http://eprints.iisc.ernet.in/43176/1/fj-minj_drtc_ictk_final.pdf

Jerez, H. N., et al. (2004). The multi-faceted use of the OAI-PMH in the LANL repository. *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries* (June 7-11, 2004, Tucson, Arizona, USA) (pp. 11- 20). USA: ACM.

Kellogg, D. (2004). *Open Source OAI Metadata Harvesting Tools.* Retrieved September 21, 2015, from http://old.diglib.org/aquifer/oct2504/harvesting.pdf

Khazraee, E., et al. (2011). EIAH data model: Semantic interoperability among distributed digital repositories. *Aslib Proceedings*, 63(1), 46–56.

Leiner, B. M. (1998). The NCSTRL Approach to Open Architecture for the Confederated Digital Library. *D-Lib Magazine,* 4(12). Retrieved November 11, 2015, from http://dlib.org/dlib/december98/leiner/12leiner.html

Mazurek, C., et al. (2006). Metadata harvesting in regional digital libraries in the PIONIER network. *Campus-Wide Information Systems*, 23(4), 241-253.

Miller, P. (2004). Interoperability. What is it and Why should I want it? *Ariadne,* 24. Retrieved March 4, 2015, from http://www.ariadne.ac.uk/issue24/interoperability/

Narang, S., et al. (2005). Creating health sciences theses institutional Repository model: An

approach at AIIMS, New Delhi, India. *Proceedings of 8th International Symposium on Electronic Thesis and Dissertations* (September 28-30, 2005, University of New South Wales, Sydney, Australia). USA: NDLTD. Retrieved July 21, 2015, from http://docs.ndltd.org:8081/dspace/bitstream/2340/264/1/044Narang.pdf

Peset, F., et al. (2007). Use of OAI protocol and its impact in digital libraries: a case study in Spain, Portugal and Latin America. In A.R.D Prasad and Devika P. Madalli (Eds.), *International Conference on Semantic Web and Digital Libraries* (February 21-23, 2007, Bangalore) (pp. 459-471). Bangalore: DRTC.

Pieper, D., & Summann, F. (2006). Bielefeld Academic Search Engine (BASE): An end-user oriented institutional repository search service. *Library Hi Tech*, 24(4), 614-619.

Powell, J., & Fox, E. A. (1998). Multilingual Federated Searching Across Heterogeneous Collections. *D-Lib Magazine*, 4(9). Retrieved July 21, 2015, from http://www.dlib.org/dlib/september98/powell/09powell.html

Prasad, A. R. D., & Guha, N. (2005). Interoperability and the OAI-PMH. *DRTC-HP International Workshop on Building Digital Libraries using DSpace* (March 7-11, 2005, DRTC, Bangalore). Paper J. Bangalore: DRTC. Retrieved December 11, 2015, from http://drtc.isibang.ac.in /xmlui/bitstream/handle/1849/245/J_OAI_paper.pdf?sequence=1

Rajashekar, T. B. (2003). Improving the visibility of Indian Research: An Institutional, Open Access Publishing Model. Paper presented at the *Indo-US Workshop on Open Digital Libraries and Interoperability* (June 23-25, 2003, Arlington, USA). Retrieved November 24, 2014, from http://fox.cs.vt.edu/IndoUSdl/raja.pdf

Roy, B. K. (2007). Indian Initiatives in the Development of Institutional Digital Repository. *Digital Media and library Information Services, Proceedings of 26th IASLIC Conference* (December 26-27, 2007, Jamia Millia Islamia University, New Delhi) (pp. 253-262). Kolkata: IASLIC.

Roy, B. K. (2010). Open Access Trends and Developments in India. *Librarian*,17, 83-87.

Roy, B. K. (2014a). *Designing Institutional Digital Repository for the University of Burdwan: A FLOSS Based Prototype*. A PhD Thesis, Library and Information Science Department, The University of Burdwan, Burdwan.

Roy, B. K. (2014b). Open Access Repository: An Alternative Model for Scholarly Communication. In G. Maity & Others, *Charaibeti: Golden Jubliee Commemorative Volume* (pp. 250 – 257). Kolkata: Jadavpur University.

Roy, B. K. (2015). *Institutional Digital Repository: From Policy to Practice*. Germany: LAP.

Roy, B. K., & Mukhopadhyay, P. (2011). Development of learning objects repositories in India. *Open Access: Gateway to Open Innovation, Proceedings of the 27th Annual Conference of the Society for Information Science* (November 24-27, 2010, Kolkata) (pp. 179-187). New Delhi: Society for Information Science.

Roy, B. K., Biswas, S. C., & Mukhopadhyay, P. (2011). An Analytical Study of Institutional Digital Repositories in India. *Library Philosophy and Practice.* Paper 692. Retrieved September 3, 2015, from http://digitalcommons.unl.edu/libphilprac/692

Roy, B. K., Biswas, S. C., & Mukhopadhyay, P. (2012a). Open Access Repositories in Asia: From SAARC to Asian Tigers. *Library Philosophy and Practice.* Paper 808. Retrieved

September 12, 2015, from http://digitalcommons.unl.edu/libphilprac/808

Roy, B. K., Biswas, S. C., & Mukhopadhyay, P. (2012b). Study of Open Access Repositories: A Global Perspective. *Information-Innovation-Technology: Creating Seamless Linkages, 29th Convention & Conference of Society of Information Science* (November 26 - 28, 2012, Silchar). Silchar: National Institute of Technology.

Roy, B. K., Biswas, S. C., & Mukhopadhyay, P. (2013). Global Visibility of Indian Open Access Institutional Digital Repositories. *International Research: Journal of Library and Information Science*, 3(1), 182-194.

Roy, B. K., Biswas, S. C., & Mukhopadhyay, P. (2015). Trends and Developments of Open Access Repository Movement in Europe. *International Research: Journal of Library and Information Science*, 5(3), 407-422.

Roy, B.K., Biswas, S.C., & Mukhopadhyay, P. (2016a). Status of Open Access Institutional Digital Repositories in Agricultural Sciences: A Case Study of Asia. *Library Philosophy and Practice*. [Accepted for publication].

Roy, B. K., Biswas, S. C., & Mukhopadhyay, P. (2016b). The COAPI Cats: The Current State of Open Access Repository Movement and Policy Documentations. *International Journal of Knowledge Content Development & Technology*. [Accepted for publication].

Roy, B. K., Biswas, S. C., & Mukhopadhyay, P. (2016c). Global Repository Movement in the Domain of Library and Information Science. *International Journal of Information Science and Management*. [Accepted for publication].

Roy, B. K., Biswas, S. C., & Mukhopadhyay, P. (2016d). *Open Access Repositories for Indian Universities: Towards a Multilingual Framework. IASLIC Bulletin*. [Mss. submitted].

Sarkar, P., & Mukhopadhyay, P. (2010). Designing single-window search service for electronic theses and dissertations through metadata harvesting. *Annals of Library and Information Studies*, 57(4), 356-364.

Sarkar, P., & Mukhopadhyay, P. (2012). Developing virtual union catalogue of ETDs on health and medicine: a practical approach. *Proceeding of the National Seminar on Challenges in Library Management System* (pp. 334-340). Indian association for the cultivation of science, Kolkata.

Singh, S., & Pandita, N. (2005). Building the open access self-archiving repository for the bio-medical sciences at National Informatics Centre. Paper presented at the *National Convention of Medical Library Association of India (*November 07- 09, 2005, Bangalore). Retrieved November 24, 2014, from http://openmed.nic.in/1108/02/mlai.pdf

Singh, S., Pandita, N., & Dash, S. S. (2008). Opportunities and challenges of establishing open access repositories: a case study of OpenMED@NIC. *Trends and Strategic Issues for Librarians in Global Information Society: ICCSR Sponsored Seminar* (March 18-19, 2008, Chandigarh) (pp. 98-104). Chandigarh: Panjab University.

Sreekumar, M. G., et al. (2007). Institutional Repositories for Knowledge Management in Academic and Research Institutions. In A. R. D. Prasad and Devika P. Madalli (Eds.), *International Conference on Semantic Web and Digital Libraries* (February 21-23, 2007, Bangalore) (pp. 260-273). Bangalore: DRTC.

Van de Sompel, H., & Lagoze, C. (2000). The Sante Fe Convention of the Open Archives Initiative. *D-Lib Magazine*, 6(2). Retrieved May 2, 2014, from www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html

Van de Sompel, H. et al. (2000). The UPS Prototype: An Experimental End-User Service across E-Print Archives. *D-Lib Magazine*, 6(2). Retrieved November 24, 2015, from http://dlib.org/dlib/february00/vandesompel-ups/02vandesompel-ups.html

Warner, S. M. (2001). Exposing and harvesting metadata using the OAI Metadata Harvesting Protocol: a tutorial. *High Energy Physics Libraries Webzine*, 4. Retrieved March 22, 2015, from http://library.web.cern.ch/library/Webzine/4/papers/3/

Westell, M. (2006). Institutional Repositories: Proposed Indicators of Success. *Library Hi Tech*, 24(2), 211–226.

Annex 1
List of LIS Repositories under Study

*Policies (self archiving): C stands for Content; D for Data; M for Metadata; P for Preservation and S for Submission

Name	Country	OAI-PMH	Software	Size	Subjects	Content	Languages	*Policies
HAL (Hyper Article en Ligne)	France	http://hal.archives-ouvertes.fr/oai/oai.php	HAL	1014552 items	Multidisciplinary	Articles; Theses; Conferences; Books; Unpublished	French	D
Queensland University of Technology ePrints Archive (QUT ePrints Archive)	Australia	http://eprints.qut.edu.au/cgi/oai2	EPrints	61446 items	Multidisciplinary	Articles; Theses; Conferences; Unpublished	English	C,D,M,S
edoc	Switzerland	http://edoc.unibas.ch/cgi/oai2	EPrints	36405 items	Multidisciplinary	Articles; References; Theses; Books	English	C,D,M,P,S
MADOC (MAnnheim DOCument Server)	Germany	http://nb-madoc.bib.uni-mannheim.de/cgi/oai2	EPrints	27865 items	Multidisciplinary	Articles; References; Conferences; Theses; Unpublished; Books	German & English	Undefined
Institutional Repository of Peking University (PKU Institutional Repository)	China	http://ir.pku.edu.cn/oai/request	DSpace	27756 items	Multidisciplinary	Articles: Books Conferences; Theses	English & Chinese	Undefined
ScholarBank@NUS	Singapore	http://scholarbank.nus.edu.sg/oai/request	DSpace	27179 items	Multidisciplinary	Articles; Conferences; Theses; Multimedia; Patents	English	Undefined
FAC (Flinders Academic Commons)	Australia	http://dspace.flinders.edu.au/dspace-oai/request	DSpace	27113 items	Multidisciplinary	Articles; Unpublished	English	Undefined
UniSA Research Archive	Australia	http://ura.unisa.edu.au/OAI-PUB	DigiTool	26855 items	Multidisciplinary	Articles; References; Conferences; Theses; Multimedia	English	Undefined
KOPS (Konstanzer Online-Publikations-System)	Germany	http://kops.ub.uni-konstanz.de/oai-dini/request	DSpace	24114 items	Multidisciplinary	Articles; Theses; Conferences; Books Unpublished	German & English	Undefined
Binus University Repository	Indonesia	http://eprints.binus.ac.id/cgi/oai2	EPrints	21565 items	Multidisciplinary	Articles; References; Theses	English & Malay	C,D,M,P,S
(IGSNRR OpenIR)	China	http://159.226.115.200/casirgrid-oai/request	DSpace	19997 items	Agriculture, Food and Veterinary; Library and Information Science	Articles; Conferences; Theses; Books	Chinese; English	Undefined

Name	Country	OAI-PMH	Software	Size	Subjects	Content	Languages	*Policies
E-LIS	Italy	http://eprints.rclis.org/cgi/oai2	EPrints	17615 items	Library & Information Science	Articles; References; Conferences; Theses; Unpublished; Books; Datasets; Learning Objects; Special	English; Italian; Spanish	C,D,M,P,S
Digital Library of the Czech Technical University in Prague	Czech Republic	https://dspace.cvut.cz/oai/request	DSpace	17442 items	Multidisciplinary	Articles; References; Theses; Multimedia	Czech; English	Undefined
CADAIR (Aberystwyth University Repository)	United Kingdom	http://cadair.aber.ac.uk/dspace-oai/request	DSpace	16141 items	Multidisciplinary	Articles; Theses	English; Welsh	S
OZone (OZone provided by Ontario Scholars Portal)	Canada	https://ospace.scholarsportal.info/oai/request	DSpace	15283 items	Library & Information Science	Articles; Unpublished; Datasets; Learning Objects	English	Undefined
D-Scholarship@Pitt	United States	http://d-scholarship.pitt.edu/cgi/oai2	EPrints	15263 items	Multidisciplinary	Articles; Multimedia Conferences; Theses; Books	English	C,D,S
Toulouse 1 Capitole Publications	France	http://publications.univ-tlse1.fr/cgi/oai2	EPrints	14718 items	Multidisciplinary	Articles;Conferences; Theses; Unpublished; Books	French & English	Undefined
eScholarship@UMMS	United States	http://escholarship.umassmed.edu/do/oai/	Digital Commons	14492 items	Health and Medicine; Library & Information Science	Articles; References; Theses; Books	English	D,M
InK (Institutional Knowledge at Singapore Management University)	Singapore	http://ink.library.smu.edu.sg/do/oai/	Digital Commons	14145 items	Social Sciences General	Articles; Multimedia Conferences; Theses; Unpublished; Books	English	Undefined