

# Open Harvester Systems with Special Reference to OAI-PMH Service Providers: A Study

Dr. Bairam Khan

Librarian, Bolpur College, Bolpur, Birbhum, PIN-731204, W.B.

---

## ABSTRACT

**Introduction:** This paper highlights the study of Open Harvester Systems with special reference to OAI-PMH Service Providers. The Open Harvester Systems is a free metadata indexing system developed by the Public Knowledge Project through its federally funded efforts to expand and improve access to research. Open Harvester Systems (OHS) is basically developed for the metadata harvesting. OHS allows you to create a searchable index of the metadata from Open Archives Initiative (OAI)-compliant archives.

**Objectives:** The goal of this paper is to discuss the importance of OAI-PMH supported metadata harvesting service providers for the next generation library interface and to trace OAI-PMH compliant various metadata harvesting service providers.

**Methodology:** For the purpose of this study, OAI-PMH service providers enlisted in Open Archives Initiative (<http://www.openarchives.org/service/listproviders.html>) have been taken and data have been analysed.

**Findings:** From this study, it is found that 19 OAI-PMH Service Providers enlisted in openarchives.org. It is observed that there is still no OAI-PMH service provider registered in Open Archives Initiatives from India, even from Asia, Oceania and Antarctica. It is also found that digital resources available through OAI-PMH service providers in different languages such as English, Croatian, German, Spanish, French, Italian, Norwegian, Greek, Polish, Deutsch etc.

**Keywords:** Open Harvester Systems, OAI-PMH Service Providers, Open Archives Initiative, Metadata Harvesting Protocol

---

Date of Submission: 22-05-2023

Date of Acceptance: 02-06-2023

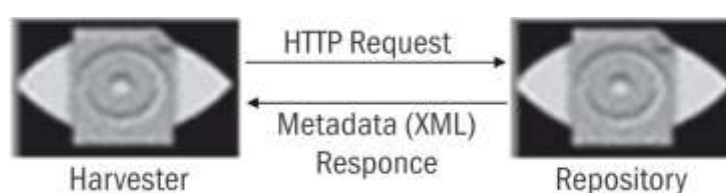
---

## I. Introduction

The Open Harvester Systems is a free metadata indexing system developed by the Public Knowledge Project through its federally funded efforts to expand and improve access to research. Open Harvester Systems (OHS) is basically developed for the metadata harvesting. OHS allows you to create a searchable index of the metadata from Open Archives Initiative (OAI)-compliant archives, such as sites using Open Journal Systems (OJS) or Open Conference Systems (OCS) (Public Knowledge Project). The main features of Open Harvester Systems include ability to harvest OAI metadata in a variety of schemas, extensible, advanced query tools, granular harvesting, filtering & normalization of metadata and customizable interface.

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) provides an application-independent interoperability framework based on metadata harvesting (Lagoze, Michael, & Warner, 2002). The OAI-PMH gives a simple technical option for data providers to make their metadata available to services, based on the open standards, HTTP (Hypertext Transport Protocol) and XML. Thus, metadata from many sources can be gathered together in one database, and service can be provided based on this centrally harvested or aggregated data.

OAI-based service providers harvest metadata from registered OAI-compliant systems and build a central index on the harvested metadata. This central index serves as a discovery tool for end-users, who need not be aware of the existence of distributed repositories.



## II. Background of the study

In the digital environment new methodologies of information management and access, coupled with advancements in digital information systems, have transformed to a great extent the ways and means of information management. Metadata, the systematic arrangement of data elements, aids the identification and location of information resources, thereby facilitating improved access to them.

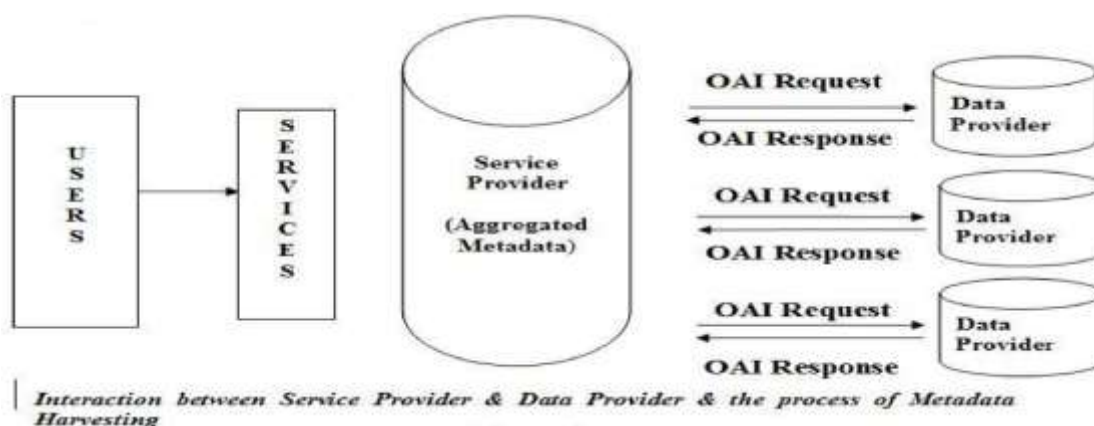
### Definition of the OHS and related terms

To define this subject some interrelated terms come together, these are Open Harvester Systems (OHS), OpenHarvesting Protocol for Metadata Harvesting (OAI-PMH). Each term can be defined as follows:

**Open Harvester Systems** - it is a free metadata indexing system which is basically developed for the metadata harvesting developed by the Public Knowledge Project.

**Metadata Harvesting Protocol:**As the term denotes, a metadata harvesting protocol sets rules or guidelines for harvesting metadata. In order to facilitate metadata harvesting, there ought to be some agreement on aspects such as: the transport protocol (HTTP or FTP etc.), the metadata format (Dublin Core, MARC, etc.); metadata quality assurance (mandatory element set, name and subject conventions, etc.) and intellectual property and usage rights.

There are two classes of players in the OAI-PMH framework: *data providers*, which administer systems that support the OAI-PMH as a means of exposing metadata, and *service providers*, which use metadata harvested via the OAI-PMH as a basis for building value-added services.



### Evolutionary history of OHS and OAI-PMH

The origin of OAI can be traced back to the efforts to increase interoperability among the e-print/ pre-print servers that hosted scientific and technical papers (Breeding, 2002).

Need was felt to build a framework to bring about a kind of integration of these e-print/pre-print archives to solve these problems. A meeting was convened in late 1999 at Santa Fe, New Mexico to address problems of the e-print world. The major work was to define an interface to permit eprint servers to expose their metadata for the papers it held, so that search services or other similar repositories could then harvest its metadata. These archives would then act as a federation of repositories by giving a single search platform for multiple collections.

After the meeting, the agreed principles were launched in January 2000 as the Open Archives Initiative specification by Herbert Van de Sompel, Rick Luce, and Paul Gisparg among others. The Digital Library Federation, the Coalition for Networked Information, and the National Science Foundation sponsored it.

The OAI Steering Committee was formed in August 2000 to give the strategic direction to the protocol. The protocol version 1.1 was launched in July 2001. The Open Archives Initiative Technical Committee (OAI-TC) was formed to develop and write version 2 of the Open Archives Initiative Protocol for Metadata Harvesting based on feedback from implementers. The OAI-PMH version 2.0 was eventually released in June 2002 (<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>).

Open Harvester Systems (OHS) (originally called "Open Archives Harvester," and later in 2005 "Harvester" was first developed in 2003 by Chia-ning Chiang, Patrick Inglis, Kevin Jamieson, Henry Kang, Daniel McLaren, Faith Shields, Robert Wickert, John Willinsky, and Eunice Yung, along with other support staff through the assistance of The Public Knowledge Project (PKP), a non-profit research initiative of the Faculty of Education at the University of British Columbia and other supporting entities. (Open Harvester Systems)

### **APUPA relations of OHS related terms**

From the history of the subject, it can be classified that OAI is boarder term that is developed for interoperability among the e-print/ pre-print servers that hosted scientific and technical papers whereas the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a low-barrier mechanism for repository interoperability and Open Harvester Systems was developed for metadata harvesting.

### **Scope and coverage of study**

Open Harvester Systems is open-source software for metadata harvesting. But due to lack of time and expertise, the study is limited to registered OAI-PMH service providers. Service provider issues OAI-PMH requests to data providers and uses the metadata as a basis for building value-added services. They are harvesters of the metadata for the purpose of providing a service such as a search interface, peer-review system, etc.

### **III. Literature Review**

A review of literature not only provides glimpses into the earlier studies carried out in this particular area but also reflects the direction in which it is moving. Amin (2003) discussed the importance of OAI-PMH and it is being regarded as important tools for information discovery in the digital library arena. This paper looks into the issues leading to its development as well as gives an inside view of the proposed model.

Peset, Ferrer, Subirats, & Muñoz, (2007) highlight the use of the Open Archives Initiative (OAI) protocol for data collection. These types of digital libraries are undergoing a worldwide boom. This work studies the current state of their implantation in Spain, Portugal and Latin America.

Hirwade and Bherwani (2009) make a study on metadata harvesting service providers. They studied sixty metadata harvesting service providers. The study reviewed metadata generation, preservation and harvesting, and various technical issues arising at these stages.

Roy, Biswas, & Mukhopadhyay, (2017) provide 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. This proposed harvesting framework may work as a guiding tool to those who plan to set up new OAI-based service provider.

Gaudinat, Beausire, Fuss, Banfi, Gobeill, & Ruch (2017) identify the need to collate the data (with a total of 42.3% OAI-PMH repositories specific to each meta-catalog) and improve current search tools, hence portraying the benefits of a comprehensible single unifying meta-catalog for end users.

### **IV. Objective of the study**

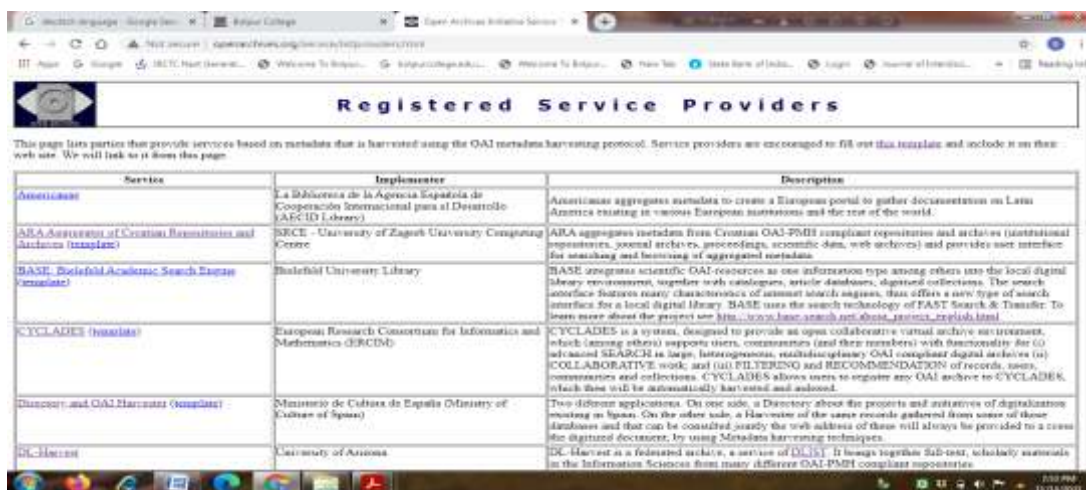
Open Access works are scattered across many disciplinary archives, institutional e-print archives, institutional repositories and open access journals. Therefore, it is difficult for users to locate all needed works on a particular subject. One important international movement to solve this problem is the Open Archives Initiatives (OAI), which aims to develop and promote the use of a standard protocol, known as the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), designed for better sharing and retrieval of e-prints residing in distributed archives.

The main objectives of the present study are:

- To discuss the importance of OAI-PMH supported metadata harvesting service providers for the next generation library interface
- To trace OAI-PMH compliant various metadata harvesting service providers.

### **V. Methodology**

The study focuses on the current status of nineteen (19) OAI-PMH service providers enlisted in Open Archives Initiative (<http://www.openarchives.org/service/listproviders.html>). The paper is largely based on a review of the literature, based on secondary data collected from the different websites using search engine (i.e. Google, Google Scholar) and academic web resources (i. e. BASE, Semantic Scholar, etc.). The data for this paper has been collected during December 2022 from Open Archives Initiative.



Snapshot of OAI-PMH Service Providers in openarchives.org

### VI. Data collection and analysis

The table-1 shows that OAI-PMH Service Providers to create a searchable index of the metadata from Open Archives Initiative (OAI)-compliant archives. These are as follows:-

**Table-1: OAI-PMH Service Providers**

Sl No	Name of Service	Name of Implementer	Description	URL	No of Collections
1	Americanae	AECID Library	Americanae aggregates metadata to create a European portal.	<a href="http://www.americanae.es/americae/es/inicio/inicio.do">http://www.americanae.es/americae/es/inicio/inicio.do</a>	-
2	ARA Aggregator of Croatian Repositories and Archives	SRCE - University of Zagreb University Computing Centre	ARA aggregates metadata from Croatian OAI-PMH compliant repositories and archives..	<a href="https://ara.srce.hr/index.php/index">https://ara.srce.hr/index.php/index</a>	Repository-155 Records-520397
3	BASE: Bielefeld Academic Search Engine	Bielefeld University Librar	BASE integrates scientific OAI-resources as one information type among others into the local digital library environment, together with catalogues, article databases, digitised collections.	<a href="https://www.base-search.net/">https://www.base-search.net/</a>	Documents-274380357, Content Provider-9204
4	CYCLADES	European Research Consortium for Informatics and Mathematics (ERCIM)	CYCLADES is a system, designed to provide an open collaborative virtual archive environment, multidisciplinary OAI compliant digital archives	<a href="https://www.ercim.eu/cyclades">https://www.ercim.eu/cyclades</a>	-
5	Directory and OAI Harvester	Ministry of Culture, Spain	Directory about the projects and initiatives of digitalization existing in Spain.	<a href="https://hispana.mcu.es/es/inicio/inicio.do">https://hispana.mcu.es/es/inicio/inicio.do</a>	-
6	DL-Harvest	University of Arizona	DL-Harvest is a federated archive, a service of DLIST.	<a href="https://repository.arizona.edu/handle/10150/105067">https://repository.arizona.edu/handle/10150/105067</a>	-
7	Galiciana	Xunta de Galicia	Galiciana is a project promoted by the Xunta de Galicia.	<a href="http://www.galiciana.gal/arpadhub/gl/inicio/inicio.do">http://www.galiciana.gal/arpadhub/gl/inicio/inicio.do</a>	463156 digital objects from 15 collections
8	Gavo data centre searchable VO registry	Zentrum für Astronomie, Universität Heidelberg and the German Astrophysical Virtual Observatory	The GAVO data center runs a searchable registry for the Virtual Observatory.	<a href="http://dc.zah.uni-heidelberg.de/wirr/q/ui/fixed">http://dc.zah.uni-heidelberg.de/wirr/q/ui/fixed</a>	30000 resouces

9	<a href="#">Harvester@UABT: Harvester of University of Tlemcen Algeria</a>	University of Tlemcen, Algeria	It is a system to collect information from anywhere repository which follow the OAI-PMH protocol.	<a href="http://dspace.univ-tlemcen.dz/">http://dspace.univ-tlemcen.dz/</a>	-
10	HISPANA	Ministerio de Cultura y Deporte, Gobierno de España	It aggregates content from Spanish libraries, archives and museums and is one of the largest data contributors to Europeana.	<a href="https://hispana.mcu.es/es/inicio/inicio.do">https://hispana.mcu.es/es/inicio/inicio.do</a>	9478358 digital objects from 227 repositories
11	MeIND	<a href="#">Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen</a>	The Service-Provider covers all subjects from different data-providers in Germany.	<a href="https://www.suchmaschinen-datenbank.de/meind-de-001212/">https://www.suchmaschinen-datenbank.de/meind-de-001212/</a>	408,587 data records
12	NORA	Norwegian Open Research Archive	It covers repositories from universities and university colleges in Norway and some research institutes.	<a href="https://nora.openaccess.no/">https://nora.openaccess.no/</a>	295257 records
13	OAIster	OCLC	It provides a search service on OAI archives.	<a href="https://www.oclc.org/en/oaister.html">https://www.oclc.org/en/oaister.html</a>	50 million records from more than 2000 contributors
14	Openarchive.s.eu	Horizons Unlimited srl (Bologna, Italy)	It is the European guide to OAI-PMH compliant digital repositories in the world.	<a href="http://www.openarchive.s.eu/">http://www.openarchive.s.eu/</a>	-
15	Perseus	Perseus	The Perseus system harvests registered OAI repositories and incorporates the information into its search interface.	<a href="https://www.perseus.tufts.edu/hopper/about">https://www.perseus.tufts.edu/hopper/about</a>	-
16	PIONEER Digital Libraries Federation	Poznan Supercomputing and Networking Center	It is another step in building a network of distributed digital libraries and repositories in Poland	<a href="https://fbc.pionier.net.pl/">https://fbc.pionier.net.pl/</a>	8.3 millions objects
17	<a href="#">PLEIADI: Portale per la Letteratura scientifica Elettronica Italiana su Archivia perti e Depositisti nazionali</a>	The CASPUR and CILEA consortia	PLEIADI project aims at building a national platform that offers centralized access to the scholarly literature archived in Italian repositories.	<a href="http://find.openarchives.it/info">http://find.openarchives.it/info</a>	3331871 documents
18	SuUB Bremen	State and University Library Bremen	The State and University Library Bremen - Germany offers a library catalogue called SuUB Bremen, where selected OAI-data providers are included.	<a href="https://www.suub.uni-bremen.de/ueber-uns">https://www.suub.uni-bremen.de/ueber-uns</a>	-
19	<a href="#">Université Numérique Ingénierie et Technologie</a>	UNIT is an association supported by the French Ministry for Education and Research	The UNIT Consortium associates all the public and private French actors of Higher Education.	<a href="https://unit.eu/">https://unit.eu/</a>	-

From the Table-1, registered OAI-PMH service providers may be analysed through the implementing agencies and their geographical coverage.

**Table-2 : Implementing agencies of registered OAI-PMH Service Providers**

Type of Implementing Agencies	No. of OAI-PMH	% (Percentage)
University	3	15.78
University Library	3	15.78
Ministry/Dept. of Govt.	3	15.78
Others	10	52.63

From the above table it is found that 3 (i.e. 15.78 %) OAI-PMH Service Providers are implemented by the university, university library and Ministry/Dept. of Govt. each. Whereas 10 (i.e. 52.63 %) OAI-PMH Service Providers are implemented by the research bodies, associations and centres. It may be graphically represented as:-

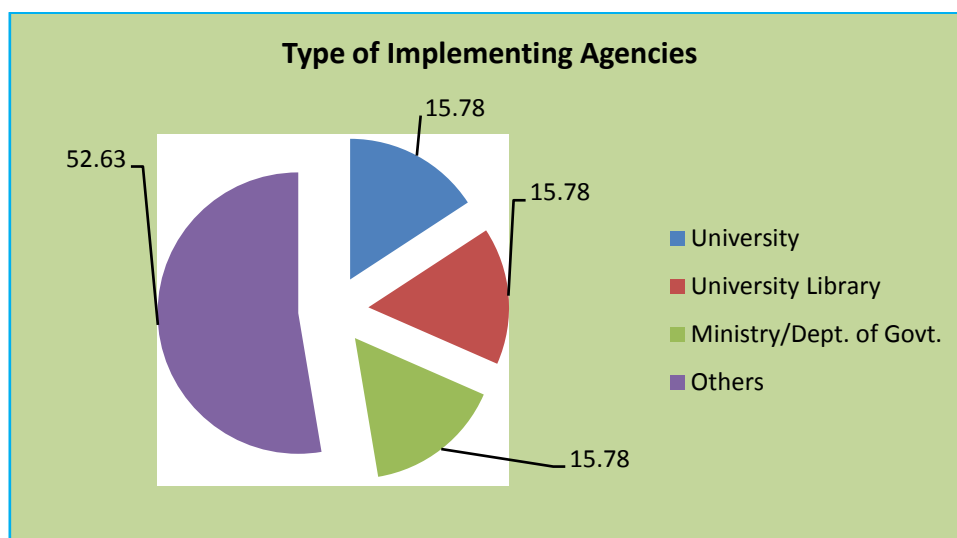


Figure-1 Type of Implementing Agencies of OAI-PMH Service Providers

Registered OAI-PMH service providers located at different countries as well as their geographical coverage have been derived.

Table-3 OAI-PMH Service Providers implemented countries and geographical coverage

Sl No.	OAI-PMH Service Providers	Name of Country	Geographical Coverage for OAI-PMH Compliant Repositories
1	Americanae	Spain	Latin America and rest of the world
2	ARA Aggregator	Croatia	Croatia
3	BASE	Germany	World
4	CYCLADES	Belgium	Europe
5	Directory & OAI Harvester	Spain	Spain
6	DL-Hervest	USA	World
7	Galiciana	Xunta de Galicia	Xunta de Galicia
8	GAVO data centre	Germany	Germany
9	Harvester@UABT,	Algeria	Algeria
10	HISPANA	Spain	Spain
11	MeIND	Germany	Germany
12	NORA	Norway	Norway
13	OAIster	USA	World
14	Openarchives.eu	Italy	World
15	Perseus	Greece	Greece
16	PIONEER	Poland	Poland
17	PLEIADL	Italy	Europe
18	SuUB Bremen	Germany	Germany
19	UNIT	France	France

From the Table-3, It is observed that 16 i.e. 84.21 % registered OAI-PMH Service Providers are located at European countries followed by America 2 i.e. 10.52 % and one implementing countries located at Africa i.e. 5.26 %.

## VII. Findings

From the above study, it is observed that there is still no OAI-PMH service provider registered in Open Archives Initiatives from India, even from Asia, Oceania and Antarctica.

University, Ministry/Dept. of Govt., Association and research centres are playing pivotal role for implementing OAI-PMH service provider to harvest metadata for facilitating searches to get digital resources across the world.

It is also found that digital resources available through OAI-PMH service providers in different languages such as English, Croatian, German, Spanish, French, Italian, Norwegian, Greek, Polish, Deutsch etc.

There is huge no of digital resources available in different languages through these OAI-PMH service providers for researchers and academic community in the world.

## VIII. Conclusion

The Open Harvesting Systems opens many new possibilities for metadata harvesting. The OAI-PMH provides an easy and effective solution for scholarly data circulation and academic communication. A number of OAI-PMH service providers are performing wonderful work in harvesting open access vehicles and open access literature scattered on the web. As a result, end users and librarian community would be beneficial, especially when considering the searching gaps of Google, Google Scholar and OAI-PMH service providers independently.

## References

- [1]. Amin, S. (2003, March). The Open Archives Initiative Protocol for Metadata Harvesting: An Introduction. DRTC Workshop on Digital Libraries: Theory and Practice . Bangalore, Karnataka, India: DRTC.
- [2]. Awasthi, S., & Jaiswal, B. (2008). Open Archives Metadata Harvesting: An Overview. PLANNER - 2008, (pp. 174-180).
- [3]. Baker, T. (. (2009). DCMI Usage Board Review of Application Profiles. Retrieved October 20, 2022, from <https://www.dublincore.org/specifications/dublin-core/application-profile-review/>
- [4]. Breeding, M. (2002). Understanding the Protocol for Metadata Harvesting of the Open Archives Initiative. *Computers in Libraries*, 22 (8).
- [5]. Gaudinat, A., Beausire, J., Fuss, M., Banfi, E., Gobeill, J., & Ruch, P. (2017, August 29). Global picture of OAI-PMH repositories through the analysis of 6 key open archive meta-catalogs. Retrieved October 25, 2022, from arXiv.org: <https://arxiv.org/ftp/arxiv/papers/1708/1708.08669.pdf>
- [6]. Hirwade, M. A., & Bherwani, M. T. (2009). Facilitating Searches in Multiple Bibliographical Databases: Metadata Harvesting Service Providers. *Liber Quarterly* , 19 (2), 140-165.
- [7]. Hodge, G. (2003). Metadata Made Simpler. Retrieved October 22, 2022, from NISO Press: [http://www.niso.org/news/Metadata\\_simpler.pdf](http://www.niso.org/news/Metadata_simpler.pdf).
- [8]. 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* .
- [9]. Lagoze, C. S., Michael, N., & Warner, S. (2002). The Open Archives Initiative Protocol for Metadata Harvesting – Version 2.0. Retrieved December 20, 2022, from [http://www.openarchives.org/OAI\\_protocol/openarchivesprotocol.html](http://www.openarchives.org/OAI_protocol/openarchivesprotocol.html).
- [10]. MacGregor, J., Stranack, K., & Willinsky, J. (2014). The Public Knowledge Project: Open Source Tools for Open Access to Scholarly Communication. In S. Bartling, & S. Friesike, *Opening Science*, (pp. 165-175).
- [11]. Naidu, G. H., & Rajput, P. S. (2008). Metadata Harvesting Tools and Services in Digital Era: A Guide for Professionals. *International CALIBER-2008* (pp. 430-436). Allahabad: University of Allahabad & INFLIBNET Centre.
- [12]. NISO (National Information Standards Organization). (2004). Understanding Metadata. Retrieved December 25, 2022, from <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>.
- [13]. Open Archives Initiative Organization. (n.d.). Open Archives Initiative-About OAI. Retrieved November 20, 2022, from <http://www.openarchives.org/organization/>
- [14]. Open Harvester Systems. (n.d.). Retrieved October 25, 2022, from LIMS wiki.org: [https://www.limswiki.org/index.php/Open\\_Harvester\\_Systems](https://www.limswiki.org/index.php/Open_Harvester_Systems)
- [15]. Peset, F., Ferrer, A., Subirats, I., & Muñoz, A. (2007). Use of OAI protocol and its impact in digital libraries: a case study in Spain, Portugal and Latin America. *ICSD-2007*, (pp. 459-471).
- [16]. Public Knowledge Project. (n.d.). Open Harvester Systems. Retrieved December 25, 2022, from <https://pkp.sfu.ca/ohs/>
- [17]. Roy, B. K., Biswas, S. C., & Mukhopadhyay, P. (2017). Designing Metadata Harvesting Framework for OAI-based LIS Repositories: A Prototype. *International Journal of Information Science and Management*, 15 (1), 73-88.
- [18]. Roy, S. G., Sutradhar, B., & Das, P. P. (2017). Large-scale Metadata Harvesting—Tools, Techniques and Challenges: A Case Study of National Digital Library (NDL). *World Digital Libraries*, 10 (1), 1-10.