



Research Database Acquisition Model Using OAI-PMH Protocol

M. Miftakul Amin¹, Ali Firdaus², Yevi Dwitayanti³, Dembi Pratama⁴, Aditya Ferdiansyah⁵, Muhammad Fakhri Ikhbar Firasya⁶

^{1,2,4,5,6}Department of Computer Engineering, Politeknik Negeri Sriwijaya, Jl. Srijaya Negara Bukit Besar, Palembang, 30139, Indonesia

³Department of Computer Accounting, Politeknik Negeri Sriwijaya, Jl. Srijaya Negara Bukit Besar, Palembang, 30139, Indonesia
miftakul_a@polsri.ac.id

Abstract. The development of scientific repositories using the Open Journal System (OJS) opens up opportunities for their utilization as research databases. Currently in Indonesia, there is a Garuda database as an aggregator to collect the research database. This research aims to integrate the metadata of scientific articles that use the OJS platform in its management, so that it can be compiled in a university scientific repository for various purposes. The method used is the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to perform data interoperability, so that it can be collected in a single database. The results showed that by utilizing the OAI-PMH protocol, a web-based application has been obtained that has successfully collected data with a total of 5,444 records of scientific article metadata from 100 Open Journal Systems.

Keywords: Research Database, Open Journal System, OAI-PMH Protocol.

1 Introduction

Please In the late 1990s, there was a concept to come up with a technology that could openly collect metadata from digital archives. According to [1], at that time, the scientific community faced problems that made it difficult to collaborate and exchange knowledge between researchers due to limited and unavailable scientific data. The development of the Open Journal System (OJS) used by journal managers has brought opportunities for increased accessibility to scientific article information [2]. This development in accessibility is a positive signal as scientific repositories play an important role in solving problems faced in the broader field of scientific communication [3].

In Indonesia, there is a Garuda database as an Open Journal System (OJS) aggregator managed by research institutions, publishers, and universities to disseminate scientific articles that have been produced [4]. The Garuda database is also used as an initial screening in the accreditation process of scientific journals in Indonesia.

One of the technologies in data exchange and article metadata information is the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) as a very high performance mechanism for client/server architecture in scientific repository interoperability [5]. This interoperability is packaged in the form of standardized syntax as a mutual agreement between various parties and systems that access it [6]. The data format used in this OAI-PMH protocol is xml format. With this XML data format, it can be easily extracted by application developers for various application purposes to be built [7]. Since the goal of the OAI-PMH protocol is to increase the openness of repositories, it provides the ability to query from one document to a group of other documents, known as metadata. This metadata includes, but is not limited to, title, author, description (possibly abstract), and physical location of the document [8].

Some related research that has been done before includes Kumar [9], which has developed multilingual institutional repositories in library information systems, by accessing research databases. The OAI-PMH protocol is used as the metadata exchange standard. This research has successfully collected thousands of scientific article metadata with various languages such as Chinese, Japanese, German, and French. Similar research was also conducted by [10] using the OAI-PMH protocol to collect specialized research database metadata in the fields of biology, chemistry, genetics, molecular, protein, and similar fields. JSON and XML data formats are used for data interoperability.

Research conducted by [11] has developed an Open Educational Resources (OERs) platform. On this platform, lecturers and students can use the same scientific reference sources. The OAI-PMH protocol is used as one of the components in data integration.

In this research, a research database metadata acquisition model derived from several scientific journal repositories that use the Open Journal System was developed. The collected research database can then be used for various research purposes, such as duplication detection topics, journal finder, expert directory, plagiarism detection, and similar research related to research databases.

2 Research Method

2.1 Dataset

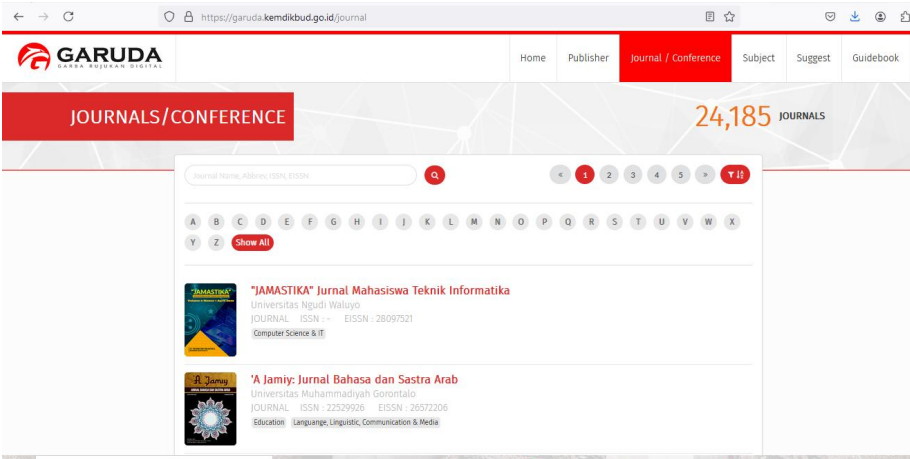


Fig. 1. Garuda Web Page Display

2.2 Database Design

To meet the needs of data storage, this research uses 3 main tables as persistent data storage. The tables that have been created include table institution, table oaipublisher, and table article. The institution table serves to store data on institutions or publishers that publish scientific articles. Furthermore, the oaipublisher table is used to store digital repository data. While the article table is used to store scientific article metadata information.

Figure 2 shows the relationship between tables formed from the three tables. In the institution table there is a Primary Key (PK) named `id_institution` and is related to the oaipublisher table with Foreign Key (FK) `id_institution`. The relationship formed is 1: n (one to many) where one institution or publisher can publish 1 or more digital repository or open journal system. While the oaipublisher table has a 1:n (one to many) relationship with the article table with a field called `request_url`. So from the relationship between oaipublisher and article, it can be understood that 1 digital repository or open journal system will have at least 1 or more articles.

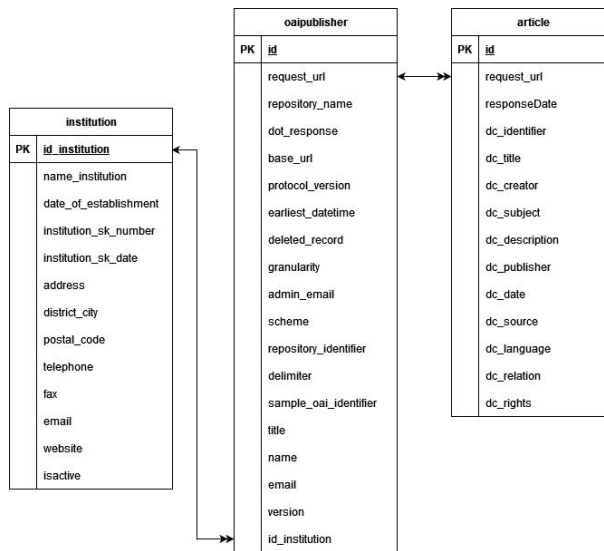


Fig. 2. Table Relationship Design

2.3 System Architecture

OAI-PMH stands for Open Archives Initiative - Protocol for Metadata Harvesting (<http://openarchives.org>) has several important methods that can be used to query metadata in the repository as shown in Figure 3. According to [12] There are several methods owned by the OAI-PMH Protocol, namely:

1. Verb=Identify, is used to obtain general information about the OAI-PMH repository. It includes the repository name, URL, supported protocol versions, administrator, date the repository was created, and other relevant metadata. This information is useful to know the basic configuration of the repository.
2. Verb=ListIdentifiers, is used to get a list of identifiers of records available in the repository. Each identifier usually refers to a specific metadata or object in the repository. ListIdentifiers is useful if you just want to get a list of identifiers without retrieving the full metadata details.
3. Verb=ListMetadataFormats, is used to find out what metadata formats are supported by the repository. OAI-PMH allows the provision of metadata in various formats, such as Dublin Core, MARCXML, MODS, and others. This verb provides a list of metadata formats that can be retrieved.
4. Verb=ListRecords, is used to retrieve complete metadata for multiple records at once. Compared to ListIdentifiers which only provide identifiers, ListRecords retrieve the full metadata in the specified format. This retrieval can be done in a specific time range or based on a specific set/collection.

5. Verb=ListSets, is used to get a list of sets or categories in the repository. These sets can be used to group records by subject, collection, or other criteria. ListSets is useful if you want to retrieve metadata only from a specific collection.

6. Verb=GetRecord, is used to retrieve the full metadata of a record based on a specific identifier. This is useful if you want to know the specific identifier of a record and want to get specific metadata in the desired format.

Of the 6 queries provided by the OAI-PMH protocol, only 2 queries are used in this research, namely Identify to get the identity of the scientific repository, and ListIdentifiers to get detailed information from each metadata of each scientific article. Referring to Figure 3, this research uses the Model-View-Controller (MVC) concept by applying Codeigniter 3 in developing web applications as a user interface between the system and the user.

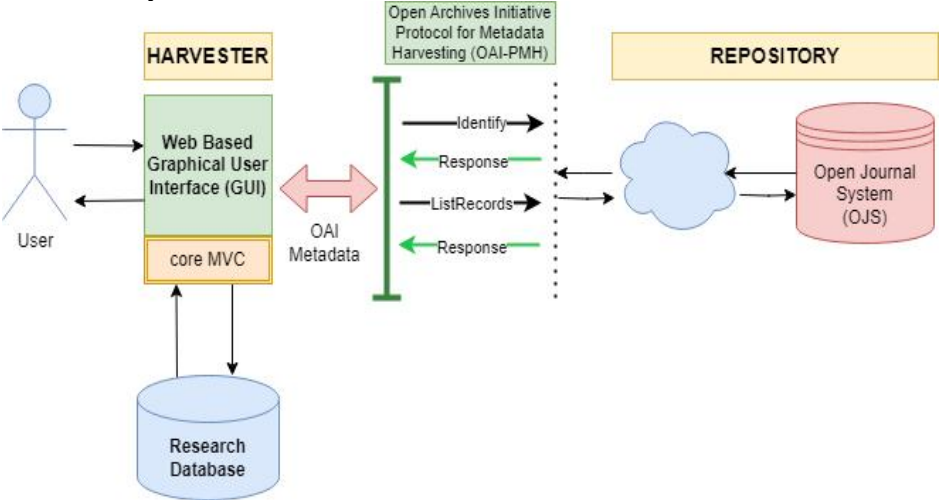


Fig. 3. System Architecture

3. Implementation and Results

3.1 Repository Identification system Architecture

Figure 4 presents a view of the web-based application that shows the identification process of the repository. The query used in the OAI-PMH protocol is verb=Identify. In the web application, the user simply enters the OAI URL data and is given the additional query verb Identify. Then the user simply presses the Identify button, so that important information related to the digital repository will be displayed. Furthermore, the system will record various important information from the digital into the database persistently. Figure 5 shows some of the data that has been stored in the database. Through this display, users can find out how much scientific article metadata has been collected in the database.

3.2 Data Acquisition Results

As shown in Figure 6, it can be seen that some metadata of scientific articles have been collected in the database. It can be seen from the graph presented, that the largest metadata obtained is 133 records and the least is 7 records.

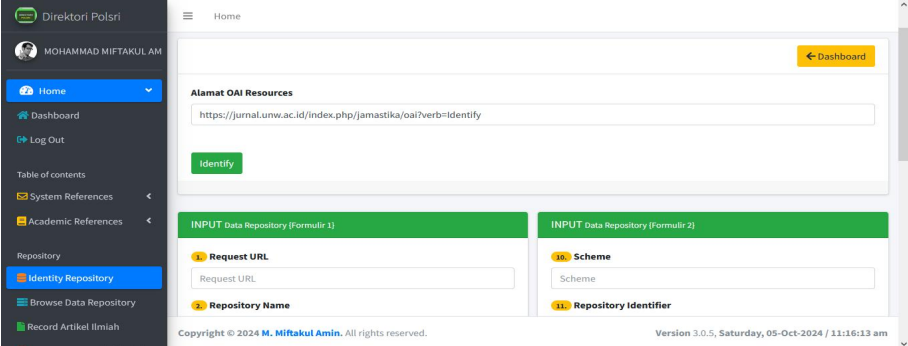


Fig. 4. Repository Identification View

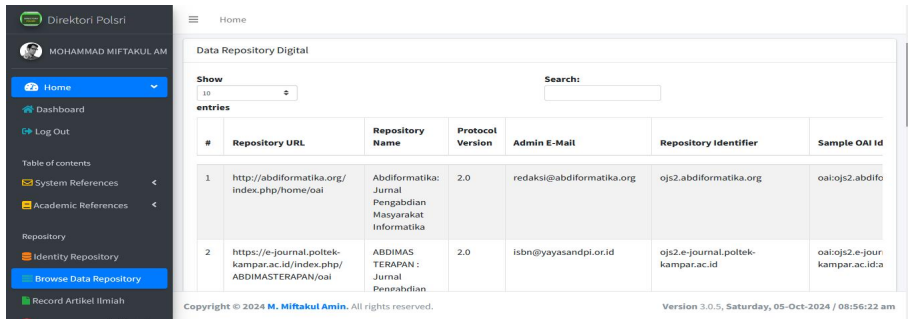


Fig. 5. Collected Data Repository

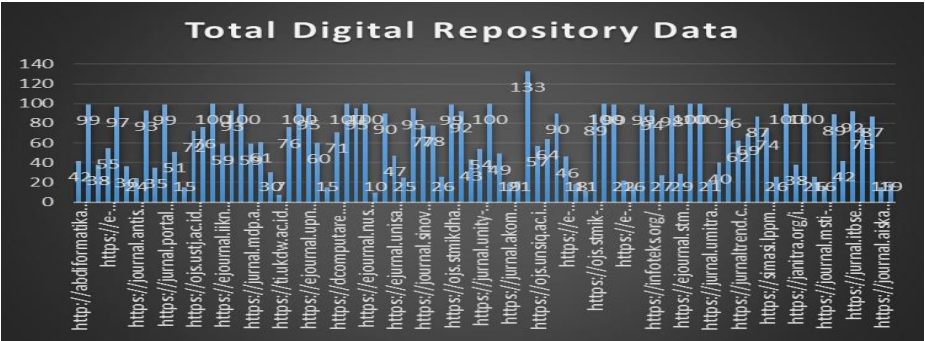


Fig. 6. Total Digital Repository Data

4. Conclusion

This research has produced a web-based application to acquire scientific article metadata from several sources that use the Open Journal System platform. By utilizing the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) protocol, 5,444 records of scientific article metadata from 100 repositories have been successfully collected. Further research that can be done from this research database that has been successfully collected is on the topic of duplicate detection, journal finder, expert directory, plagiarism detection, and other research that utilizes research database metadata.

Acknowledgments

We would like to thank the Ministry of Education, Culture, Research and Technology through the Directorate General of Vocational Education as the funder of this research in the 2024 vocational product research scheme. We would also like to thank Sriwijaya State Polytechnic through the Technical Implementation Unit of Information and Communication Technology as a partner in this research. Hopefully, this research can contribute and benefit the development and advancement of science and technology in Indonesia.

References

- [1] Y. Tkachov, "Open Archives initiative: A fast way of integration into global open science," *Challenges Issues Mod. Sci.*, vol. 2, no. June, pp. 432–445, 2024.
- [2] D. Gonzalez-Argote and J. Gonzalez-Argote, "Generación de grafos a partir de metadatos de revistas científicas con el sistema OAI-PMH," *Semin. Med. Writ. Educ.*, vol. 2, no. December 2023, p. 43, 2023, doi: 10.56294/mw202343.
- [3] J. M. Masinde and O. Sanya, "Analysis of interoperability, security and usability of digital repositories in Kenyan Institutions of Higher Learning," *Data Inf. Manag.*, vol. 6, no. 4, p. 100011, 2022, doi: 10.1016/j.dim.2022.100011.
- [4] M. Miftakul Amin, D. Stiawan, Ermatita, and R. Budiarto, "Proposed threshold-based and rule-based approaches to detecting duplicates in bibliographic database," *Bull. Electr. Eng. Informatics*, vol. 13, no. 3, pp. 2036–2047, 2024, doi: 10.11591/eei.v13i3.7665.
- [5] S. Mandal Assistant Professor, "International Journal of Library and Information Studies Harvesting Full text and Metadata of OpenDOAR through DSpace OAI-PMH: A Framework for Institutional Digital Repositories," *Int. J. Libr. Inf. Stud.*, vol. 8, no. 1, pp. 418–424, 2018, [Online]. Available: <http://www.openoar.org/>
- [6] Y. Hidalgo Delgado, E. Ortiz Munoz, and J. P. Febles Rodriguez, "A Method for Integrating Bibliographic Data from OAI-PMH Data Providers," *IEEE Lat. Am. Trans.*, vol. 15, no. 9, pp. 1695–1699, 2017, doi: 10.1109/TLA.2017.8015075.
- [7] H. A. Khajgiwale, V. M. Kharche, and A. U. Desai, "A Study on Harvester for

- OAI-PMH compliant Institutional Repositories for Academic Institutions,” in *International Conference on Intelligent Computing and Control Systems*, 2017, pp. 63–66.
- [8] R. Yuliant and N. Karna, “Knowledge sharing filtering on OAI-PMH,” in *2016 International Conference on Information Technology Systems and Innovation, ICITSI 2016 - Proceedings*, 2017. doi: 10.1109/ICITSI.2016.7858213.
- [9] D. R. Kumar, D. A. Kumar, and D. P. S.K., “Multilingual Repository on (DSpace): A Study of Open Archives Initiative and Protocol for Metadata Harvesting (OAI-PMH),” *Int. Res. J. Comput. Sci.*, vol. 10, no. 04, pp. 19–26, 2023, doi: 10.26562/irjcs.2023.v1003.02.
- [10] S. Neumann *et al.*, “Harmonising, Harvesting, and Searching Metadata Across a Repository Federation,” *Proc. Conf. Res. Data Infrastruct.*, vol. 1, pp. 1–4, 2023, doi: 10.52825/cordi.v1i.202.
- [11] C. G. Duran and C. M. Ramirez, “Integration of Open Educational Resources Using Semantic Platform,” *IEEE Access*, vol. 9, pp. 93079–93088, 2021, doi: 10.1109/ACCESS.2021.3092315.
- [12] N. Karna, I. Supriana, and N. Maulidevi, “Recommendation System on Knowledge Management System via OAI-PMH,” in *EECSI*, 2017, pp. 19–21.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

