Harvesting Full text and Metadata of OpenDOAR through DSpace OAI-PMH : A Framework for Institutional Digital Repositories

Dr. Sukumar Mandal
Assistant Professor
Department of Library and Information Science
The University of Burdwan
Burdwan – 713 104
Email: sukumar.mandal5@gmail.com

Abstract *- Digital resource is increasing in this modern age due to exponential growth of technological impact and its peripherals. IDRs will additionally include digital materials that subsist outside the physical and administrative bounds of any one digital library. IDRs will include all the processes and accommodations that are the backbone and nervous system of libraries. The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a very high performance mechanism for client server architecture in repository interoperability. Broadly it can be classified in two ways such as data provider and service provider.* Data Provider can *access the structured standards global metadata by using OAI-PMH tool. On the otherhand* service providers *then make easily harvest the metadata through a set of six verbs of OAI-PMH based URL. Requests for data can be based on a datestamp range, and can be restricted to named sets defined by the provider. Data providers can only managed the XML based Dublin Core metadata format. It is possible to designing and developing an integrated institutional digital repository in different ways. In this research paper harvest the fulltext and metadata available in OpenDOAR by using the open source software DSpace. It is fully compatible with web-enabled OAI-PMH features. The whole tasks have performed in Ubuntu operating system for easy installation and configuration of DSpace OAI-PMH whereas all the resources import from the OpenDOAR.*

Keywords: DSpace, OAI-PMH, OpenDOAR, Open Access, and IDR

Introduction

OpenDOAR is a Directory of Open Access Repositories which consists of lists of website in the field of academic open access repositories. It is possible to easily search by locale, content and other measures. The service does not require complete repository details and does not search repositories metadata. It is maintained by the University of Nottingham under the SHERPA umbrella of services and was developed in collaboration with Lund University. The project is funded by the Open Science Institute, Jisc, the Consortium of Research Libraries (CURL) and SPARC Europe (http://www.opendoar.org/). Open archive initiative for protocol metadata harvesting is one of the important component in the field of open access for managing and retrieving the right information to the right users at right time for fulfilling the library objectives and users satisfactions. Open access resource is one of the important aspects in modern information services which increases the reading habits among the researchers in learned society for the development of an integrated knowledge society. The concept of repository is a part of digital library or in content management system. It is a one type of mechanism or techniques that can be managed and storing the digital objects in different subject fields. It can be classified in two ways such as subject and institutional based where putting the content and digital resources for maximum use of these resources of different purpose. Most of the repositories are support the common characteristics like learning,

research and administrative tasks and in this stage by using the open source standards or software for easily access the multiple resources. Import and export is easily possible by using the open source standards and it also store with metadata and full-text for the easy retrieval of information from a particular websites including lot of contents and sub-contents. Desinging and development of an integrated institutional digital repository is achieved in different ways but this research paper has explore the new innovative process for access fulltext and metadata by using the DSpace OAI-PMH tool. It can save the time of the library professionals and users for the better management and retrieval of digital information resources. All the fulltext and metadata is access from the OpenDOAR by using the open archive innitiative protocol base link.

Review of Related Works

Both national and international literature regarding this study have been consulted to find out the scope of the study and some of notable related studies have been recorded below--

Nikos and others in 2014 discussed the open archive initiative for protocol metadata harvesting for managing the repositories of different digital collections through multiple aggregators and metadata schemas. Deng & Reese in 2009 customizing the DSpace for uploading the Electronic Theses and Dissertations (ETD) by using the metadata standard Dublin Core and mapping the MARC records to Dublin Core for easy retrieval of information among the users. Theoretical is also an important component in institutional digital repository for successfull learning among the users by using the DSpace open source software. This framework is applicable in semantic level collections and searching (Solomou & Koutsomitropoulos, 2015). This is the series of 5th by comibination of two project such as UK Joint Information Systems Committee (JISC)-funded RoMEO Project (rights metadata for open-archiving). The survey explores two things including data provider and service provider (Gadd, Oppenheim & Probets, 2004). Academic library is one of the important aspect in modern society for the development of open access repositories and metadata analysis and also provides insight for academic libraries on how open access repositories development and metadata analysis can enhance new professional challenges for information professionals in the field of data management, data quality and intricacies (Leng, Ali & Hoo, 2016). The paper provides an overview of the functionalities of BASE and gives insight into the challenges that have to be faced when harvesting and integrating resources from multiple OAI servers (Pieper & Summann, 2006). The purpose of this paper is to demonstrate the relative effectiveness of a range of search tools in finding open access (OA) versions of peer reviewed academic articles on the world wide web (Norris, Oppenheim & Rowland, 2008).

In 2017, Sengar, Lohiya and Rai presented a paper on CSIR Institutional Digital Repository in a national conference held in Pune and they commented on CSIR IDR as a case study (Sengar, Lohiya & Rai, 2017). Both the authors Fortier & Laws made their focus on Marquette University's institutional repository and they told about an innovative concept on marketing of IDR in this digital era (Fortier, Rose & Laws, Emily, 2014). Passehl-Stoddart and Monge in their paper aims to highlight five IDRs collections emphasis on undergraduate students and showed how IDRs create an opportunity to build student-centric collections. Hamersly Library of Western Oregon University developed an institutional repository with collections based on the university's undergraduate education. In this repository, students are encouraged to publish their original articles, presentations and creative works (Passehl-Stoddart & Monge, 2014) Manjunatha and Thandavamoorthy in their paper present the trend of researchers of Karnataka University for deposition of their intellectual output to IDR. They

showed that the most of the science and technology scholars become aware of IDR and they were willing to deposit their papers to IDR. On the other hand, the Humanities and Social Science researchers are reluctant to deposit their intellectual output to the institutional repository (Manjunatha & Thandavamoorthy, 2011). Ashok Kumar stated that Institutional Repository (IR) is the prime source of digitized intellectual output deposited by research scholars. He also observed that the most of the research and development institutes and few academic institutes in India host their IR to provide service to users (Ashok Kumar, 2009). Royster in 2012 stated in his paper about the different aspects of OARs and the licensing policies attributed on OARs in this regard (Royster, 2012). Robin in his paper in VINE stated about Institutional repository and its features and development (Yates, 2003). Both Laxminarsaiah, Ashalatha & Rajgoli, Iqbalahmad U. and Zahid showed in their paper how to build and develop Institutional repository through open source software (Laxminarsaiah ; Rajgoli; Iqbalahmad ,2007 & Zahid , 2010). Chen & Jieh in their paper stated the innovative features of Institutional digital Repository and its development and creation though open source software (Chen, & Jieh, 2009). In the light of this study the following papers have been noted--Mandal in his paper 2015 states about thesaurus construction tool and he in 2016 his paper studied on search indexing tool (Mandal, 2015 & Mandal, 2016). In his paper in 2017, Mandal developed a domain specific cluster for college libraries (Mandal, 2017). Chakrabarti and Mandal in their paper on DOAB showed how single window search facility of books is helpful for researchers (Chakrabarti and Mandal, 2017). In a letter from the secretary of UGC dated the 16th August, 2016 { D.O.No.F.l -l12016 (Secy) } , it is stated that UGC requested IIT Kharagpur to build a National Digital Library which will harvest all metadata content from other Institutional Digital Repositories in India (UGC, 2017). National Knowledge Commission led by Sam Pitroda proposed to set up Indian Institutional Digital Repository and suggested that the higher education and R&D sectors should frame guidelines and open access policies to improve accessibility of research (NKC,2007). Moreover, Developing Library Network (DELNET, 2017) and Information Library Network (INFLIBNET, 2017) have initiated for the establishment the IDRs.

Objectives

There are many digital resources available in an Internet environment. To determine the recall and precission of open access resources for users. It is vary from person to person to analysed the relevant resources. The main objective of this study is to access both the fulltext and metadata available in OpenDOAR for easy access and better management of institutional digital repository by using the open archive innitiative protocol metadata harvesting in DSpace. Library professionals and advanced users in different academic institutions have benefitted by using this prototype integrated framework.

Methodology

The process and methodology is very simple and sophisticated for harvesting the open access resources from different OAI-PMH base URL. But here only select the OpenDOAR because it gives the full feature of metadata. Designing and developing of an integrated framework is one of the main task in the domain of metadata and fulltext harvesting which available in OpenDOAR. The whole tasks is to be performed in DSpace because it is a well known open source institutional digital repository software. It is developed by the MIT and HP Lab during 2002 written in Java and fully support the cross-platform. In this original research paper has select only one operating system i.e Ubuntu because its very reliable and secure for the users.

OAI Provider service interface

This is very interesting in DSpace for access the different documents available in an online environment. It is consists of major five components such as OAI provider, OAI set, metadata format, and harvesting options. Just copy the OAI-PMH base URL from OpenDOAR and paste into this interface for importing the fulltext and metadata in different collections respectively. The Figure – 1 is represents the open archive initiative service interface in DSpace which helps to import the open access resources in different item types. Finally click on the save option against in one collection which available in DSpace.

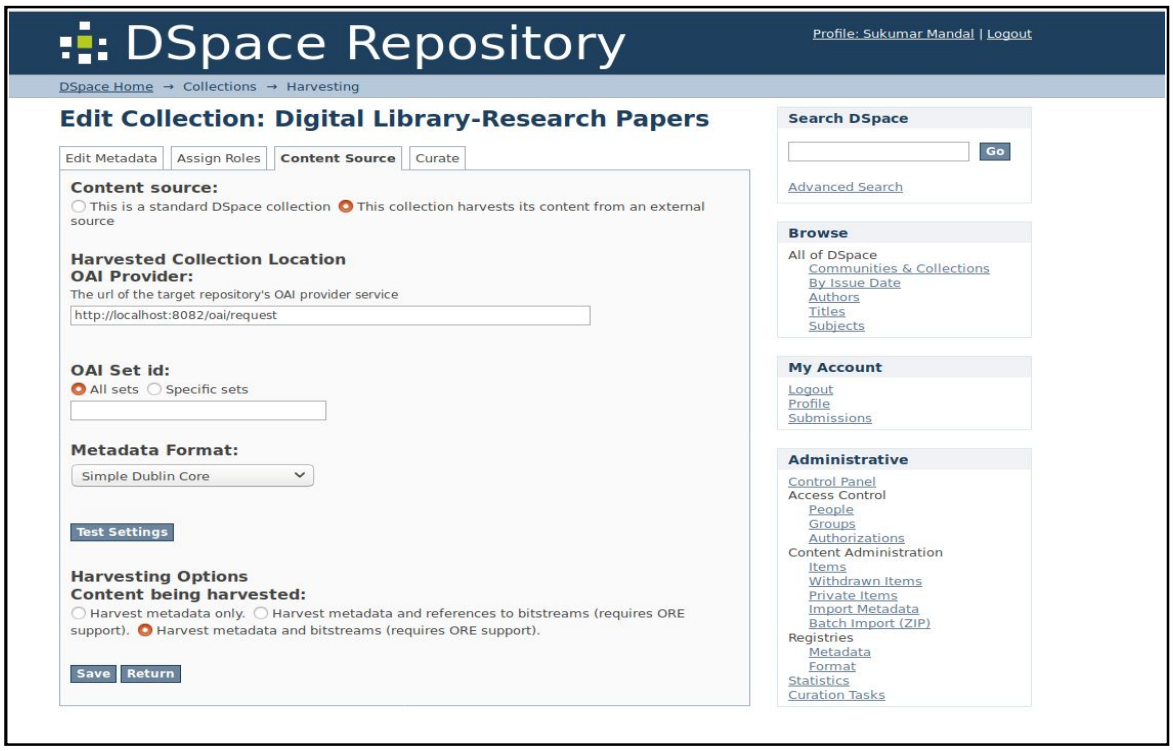

Figure – 1: DSpace OAI Provider Service Interface

Results of DSpace OAI

This interface is known as DSpace OAI-PMH data provider. Now, the results of open archive innitiative of different collections is display in the Figure-2. It shows the sets of metadata such as identify, sets, records, identifier, and metadata formats.

This is the metadata interface in DSpace and all the data relates with the Dublin Core global standard. The Figure-3 is display the Dublin Core metadata interface which helps to import different collections available in DSpace.

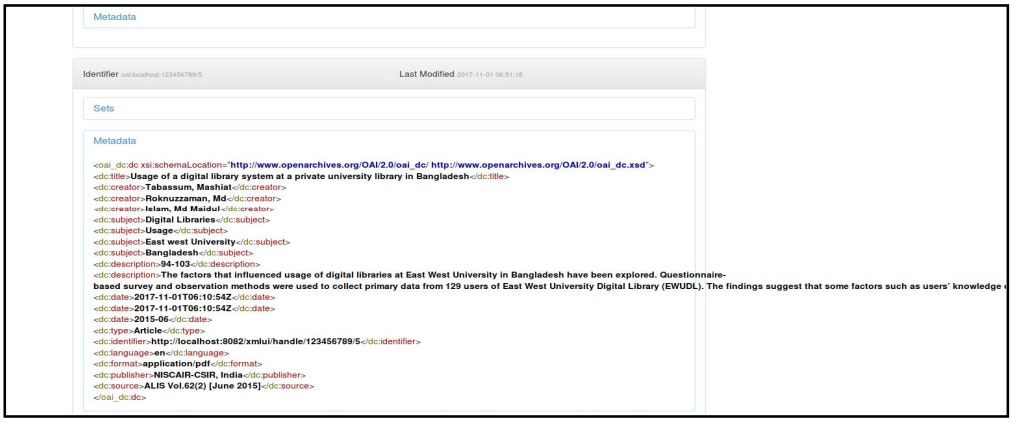

Figure – 3: Dublin Core Metadata Interface in DSpace

Generally users can enter the specific websites for access the relevant resources in different aspects. This is very time consuming process. But this research paper has successfully import all the necessary resources available in a particular website. In this paper only explore the OpenDOAR open access resources. The Figure – 4 is display the final results of different collections in DSpace which available in OpenDOAR. It is also display the number of resources available in a particular repository under the OpenDOAR.

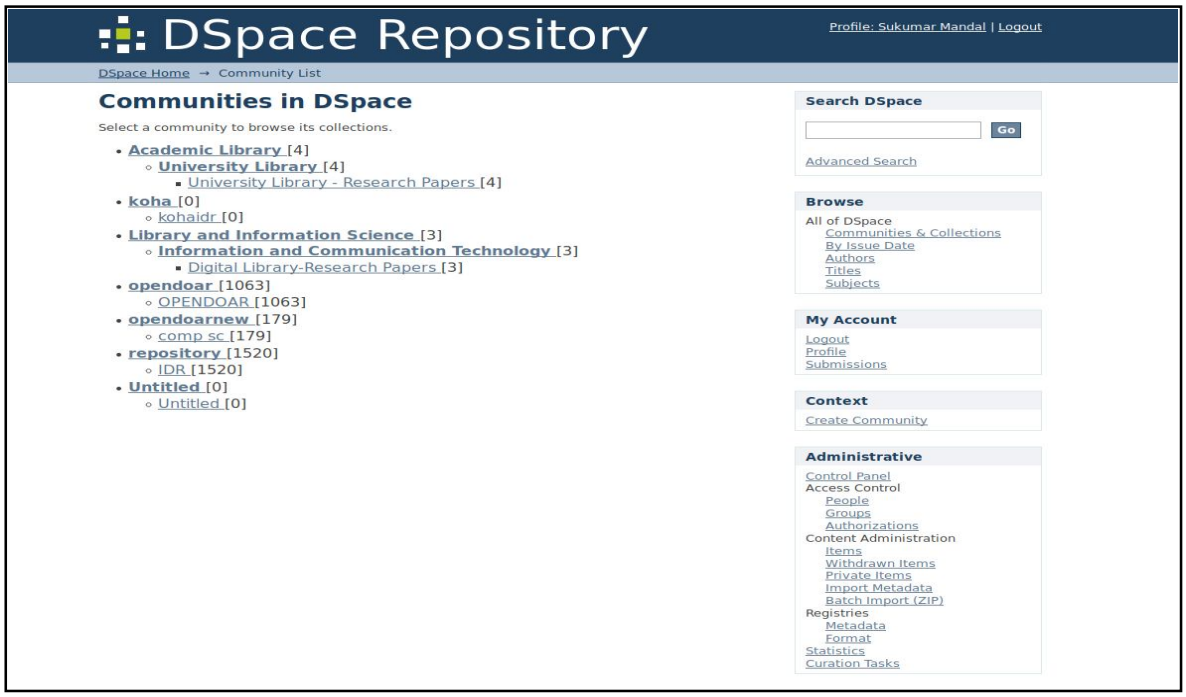

Figure – 4: Communities and collections in DSpace harvesting from OpenDOAR

Conclusion

In IDRs, a large number of resources are available to academic community for easy searching at any time anywhere. The awareness programs, training and workshops should be organized and funded by the leading organizations like UGC, CSIR etc. to educate and instruct the faculty members, researchers and scientists and all other working bodies to adopt the open access approach and make awareness regarding the creation of IDRs. It is noteworthy that the academicians, researchers are not at all fully aware of the IDRs and the maintenance of IDRs has not also properly been done. A mandatory order should enforce to harvest the metadata of

IDRs to Digital Library. A central repository should be built by using open source software to harvest metadata from various IDRs to provide a single window search facility to the academicians and researchers. From the above discussions it is clear that users have easily access the fulltext as well as metadata by using the DSpace interface because its fully support the OAI-PMH based URL which available in OpenDOAR. Overall this can be save time of the reader as well as library professionals for the better management and retrieval of open access resources which available in OpenDOAR.

References

1. Ashok Kumar, N. (2009, March). Institutional Repositories in India. Paper presented at the National Seminar on Strategic Information Management in the Digital Era, Tamil Nadu (India). Retrieved from http://eprints.rclis.org/17511/ (Accessed on 11.09.2017)
2. Chakrabarti, Abhijit & Mandal, Sukumar. (2017).Overview of Open Access Books in Libraryand Information Science in DOAB. International journal of Library and information Studies, 7(4),185-192.Retrieved from http://ijlis.org/img/2017_Vol_7_Issue_4/185-192.pdf (Accessed on 02.11.2017)
3. Chen, Kuang-hua & Hsiang, Jieh. (2009) .The unique approach to institutional repository: Practice of National Taiwan University. The Electronic Library,27(2), 204-221. Retrieved from https://doi.org/10.1108 /02640470910947 (Accessed on 04.12.2017)
4. DELNET. (2017). About DELNET. Retrieved from http:// www.delnet.org (Accessed on 09.10.2017)
5. Deng, S., & Reese, T. (2009). Customized mapping and metadata transfer from DSpace to OCLC to improve ETD work flow. New Library World, 110(5/6), 249–264. https://doi.org/10.11 08/03074800910954271
6. Fortier, Rose & Laws, Emily. (2014) . Marketing an Established Institutional Repository: Marquette Libraries' Research Stewardship Survey. Library Hi Tech News, 31(6), 12-15.
7. Gadd, E., Oppenheim, C., & Probets, S. (2004). RoMEO studies 5: IPR issues facing OAI data and service providers. The Electronic Library, 22(2), 121–138. https://doi.org/10.1108/0264047 0410699143
8. Houssos, N., Stamatis, K., Koutsourakis, P., Kapidakis, S., Garoufallou, E., & Koulouris, A. (2014). Enhanced OAI-PMH services for metadata sharing in heterogeneous environments. Library Review, 63(6/7), 465–489. https://doi.org/10.1108/LR-05-2014-0051
9. INFLIBNET. (2017). About INFLIBNET. Retrieved from http://www.inflibnet.ac.in (Accessed on 11.10.2017)
10. Laxminarsaiah, Ashalatha & Rajgoli, Iqbalahmad U. (2007) .Building institutional repository: an overview. OCLC Systems & Services. International digital library perspectives, 23(3),278-286. Retrieved from https://doi.org/10.1108/10650750710776413 (Accessed on 05.12.2017)
11. Leng, C. B., Ali, K. M., & Hoo, C. E. (2016). Open access repositories on open educational resources: Feasibility of adopting the Japanese model for academic libraries. Asian Association of Open Universities Journal, 11(1), 35–49. https://doi.org/10.1108/AAOUJ-06-2016-0005
12. Manjunatha, K. & Thandavamoorthy,K. (2011) . A study on researchers' attitude towards depositing in institutional repositories of universities in Karnataka (India).

International Journal of Library and Information Science,3(6),107-115.Retrieved from http://www.ac ademicjournals.org/ijlis.on 22.10.2017 (Accessed on 05.11.2017)

13. Mandal, Sukumar . (2016). Use of Search Indexing Tools in Library Automation: A Comparative Analysis. International Research Journal of Multidisciplinary Studies, 2 (6), 1-13.
14. Mandal, Sukumar (2017). Development of Domain Specific Cluster : An Integrated Framework for College Libraries under the University of Burdwan. Library Philosophy and Practice (Peer Review e-Journal)
15. Mandal, Sukumar (2015). MultiTes: A Knowledge Organization Thesaurus Construction Tool for College Libraries under the University of Burdwan. International Research Journal of Interdisciplinary & Multidisciplinary Studies (IRJIMS), 1 (8), 63-79.
16. NKC. (2017). About National Knowledge Commission. Retrieved from http:// www.NKc.org (Accessed on 05.12.2017)
17. Norris, M., Oppenheim, C., & Rowland, F. (2008). Finding open access articles using Google, Google Scholar, OAIster and OpenDOAR. Online Information Review, 32(6), 709–715. https://doi.org/10.1108/14684520810923881
18. OpenDOAR (2018). About OpenDOAR. Retrieved from http://www.opendoar.org/. (Accessed on 01.02.2018)
19. Passehl-Stoddart, E, & Monge, R. (2014). From Freshman to Graduate: Making the Case for Student-Centric Institutional Repositories. Journal of Librarianship and Scholarly Communication, 2(3),1130. Retrieved from http://dx.doi.org/10.7710/2162-3309.1130 (Accessed on 12.12.2017)
20. 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. https://doi. Org/10.1108/07378830610715473
21. Royster,P.(2012).Open-Access journals and institutional repositories. Library Conference Presentations and Speeches,82.Retrieved from http://digitalcommons.unl.edu/library_tal ks/82 (Accessed on 25.12.2017)
22. Sengar, K.P.S.;Lohiya, R.K.& Rai,V. (2017, February). CSIR Knowledgebase: An assessment of CSIR Institutional Digital Repositories of India. Paper presented at the National Conference on Library Innovations for Excellence,Pune(India).
23. Solomou, G., & Koutsomitropoulos, D. (2015). Towards an evaluation of semantic searching in digital repositories: a DSpace case-study. Program, 49(1), 63–90. https://doi.org/10.1108/ PROG-07-2013-0037
24. UGC. (2017). About UGC initiatives. Retrieved from http://ugc.ac.in/new_initiatives/etd_hb.pdf (Accessed on 23.10.2017)
25. Yeates, Robin (2003). Institutional repositories. VINE, 33(2), 96-101, Retrieved from https://doi.org /10.1108/03055720310509064 (Accessed on 15.12.2017)
26. Zahid Hossain Shoeb, (2010). Developing an institutional repository at a private university in Bangladesh, OCLC Systems & Services: International digital library perspectives, 26 (3), 198-213, Retrieved from https://doi.org/10.1108/10650751011073634 (Accessed on 05.11.2017)