ETD 2022
Novi Sad
Sep 7 – 9, 2022

25th International Symposium on Electronic Theses and Dissertations
– ETD 2022, Novi Sad, Serbia
September 7 - 9, 2022

Harvesting ETD Metadata from 'Shodhganga' to National Digital Library of India: present scenario and implementations

Abstract. *By using ICT, there has been a dramatic change in the world of e-learning and learning resource materials. Through the OAI-PMH, the records data about information sources allows an initial assessment of compatibility. It provides an avenue for merging information or for exchanging information between systems. Integration of Indian ETDs based on metadata has received wide attention for better management and utilization. This study explores the current state of metadata harvesting of Shodhganga or Indian ETD repository to NDL platform. One of the fundamental aims of NDL (National Digital Library of India) is to share the e-resources of multiple academic institutions and other resources across a common platform. One of the major objectives realized through this activity include increased visibility and enhanced discoverability of resources. Based on the variation in contents, metadata schema in NDL platform has been categorized into three basic classes, they are Generic Metadata, Educational Metadata and Thesis Metadata. Thesis Metadata fields like researcher, advisor, degree, etc. This Thesis Metadata standard has been used Qualified DC to harvesting the national ETDs to NDL repository. Some popular metadata schema of national and global ETDs repositories, such as ETD-MS (NDLTD) and UK-ETD (EThOS) also compared in the paper.*

1 Purpose of the paper

A digital library system can allow end-users to tag a resource with terminology meaningful to that user. Understanding user expectations is most important step in designing the discovery functionality for every effective digital library system, it enables retrieval of information and knowledge by making digital materials available to a large audience by removing barriers of language and culture. In order to avoid duplication of efforts and to increase accessibility, merging information or for exchanging information between systems metadata holds the key of success in digital library system. Integration of Indian ETDs based on metadata has received wide attention for better management and utilization. So, ETD metadata, as structured data about ETDs, enables user and system interactions with the resources themselves.

2 Objectives

The major objectives of the study are –

- To explore the need of metadata harvesting;

- To explore the metadata challenges;

- To show the current state of metadata harvesting from *Shodhganga* to NDL.

3 Design, methodology, approach

The data for the study were collected from the portals of 'Shodhganga' and 'National Digital Library of India' and also various primary literatures and human sources. A comparison overview, mainly on *Shodhganga* ETD metadata, ETD-MS (developed by NDLTD) and UK-ETD (EthOS) has been presented in the article. How to metadata harvest in the NDLI portal from *Shodhganga* has been presented in pictorial representations.

4 Findings

The National Digital Library of India (NDLI) is an integration of digital platform for all types of academic institutions, differently abled pupils and anybody who has a willingness to learn. Simply speaking, an information and knowledge repository gateway is a central location to access, collect, data management, networks, platforms and systems. Online availability of electronic thesis through centrally maintained digital repositories, ensures easy access and archiving of Indian doctorial theses but also helps in the long run, raising the standard and quality of research, this would overcome the serious problem of duplication of research and poor quality resulting from the "poor visibility" and the "unseen" factor in research output. Most ETDs are now born digital assets. As such, they often coexist with author- supplied metadata that has potential for repurposed. Harvesting metadata is also enhanced to facilitate discovery and access in an online environment. Proper indexing of Indian ETDs through NDLI makes wide usage of resources and also minimizes duplication of research. In a digital environment, information traceability has become extremely important. So a major motivation for universities in publishing their research output on the www is to improve discoverability of ETDs by existing users as well as by new users.

5 Research limitations and implications

The research limitations are mainly limited to Indian ETD repository i.e. *Shodhganga* and National Digital library of India. If a set of digital resources are described in the same metadata schema according to similar rules, then a DL application can use this metadata to provide services more easily. The OAI-PMH protocol requires all resources have metadata available in Dublin Core, and encourage professional communities to supplement this simple metadata format with additional formats as Qualified DC useful for ETDs. The study may play a key role for further research.

6 Practical implication

ETD metadata can be tightly coupled with resources and transported together with that ETD‟s contents in www. NDLI-ETD schema initially created a baseline for ETD metadata collection and dissemination. It would allow federated searching across ETDS within NDLI and facilitate a greater degree of consistency in the presentation of the records.

References

1. Das, A., & Sutradhar, B. (2018). Harvesting of additional metadata schema into DSpace through OAI- PMH: Issues and challenges. SRELS Journal of Information Management, 55(1), 1.

2. Deng, S., & Reese, T. (2009). Customized mapping and metadata transfer from DSpace to OCLC to improve ETD work flow. New Library World, 110(5), 249-264. doi: http://dx.doi.org/10.1108/03074800910954271

3. Foulonne, M. & Riley, J. (2008). Metadata for digital resources: implementation, systems design and interoperability. Oxford: Chandos Publishing; p.124-125.

4. Hakimjavadi, H., & Masrek, M. N. (2013). Evaluation of interoperability protocols in repositories of electronic theses and dissertations. Program, 47(1), 34-59. doi:http://dx.doi.org/10.1108/00330331211296303

5. https://ndlproject.iitkgp.ac.in/sites/default/files/files/Metadata_Schema_Executive_Summary.pdf

6. http://shodhganga.inflibnet.ac.in

7. McCutcheon, S., Kreyche, M., Margaret, B. M., & Nickerson, J. (2008). Morphing metadata: Maximizing access to electronic theses and dissertations. Library Hi Tech, 26(1), 41-57. doi:http://dx.doi.org/10.1108/07378830810857799

8. Mukhopadhyay, P. (2015). Interoperability and retrieval. UNESCO. p.28-31. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000232199.

9. Park, E. G., & Richard, M. (2011). Metadata assessment in e-theses and dissertations of canadian institutional repositories. The Electronic Library, 29(3), 394-407.

10. doi:http://dx.doi.org/10.1108/02640471111141124

11. Potvin, S., & Thompson, S. (2016). An analysis of evolving metadata influences, standards, and practices in electronic theses and dissertations. Library Resources & Technical Services, 60(2), 99-114. doi:http://dx.doi.org/10.5860/lrts.60n2.99

12. Suleman, H., & Fox, E. A. (2003). Leveraging OAI harvesting to disseminate theses. Library Hi Tech, 21(2), 219. doi:http://dx.doi.org/10.1108/07378830310479857

13. Thompson, S., Liu, X., Duran, A., & Washington, A. (2019). A case study of ETD metadata remediation at the university of houston libraries. Library Resources & Technical Services, 63(1), 62. doi:http://dx.doi.org/10.5860/lrts.63n1.62.

14. UNESCO & National Digital Library of India. (2017). UNESCO-NDL India International workshop on Knowledge Engineering for Digital Library Design [Brochure]. New Delhi: Author.

15. Veve, M. (2016). Harvesting ETD metadata from institutional repositories to OCLC: Approaches and barriers to implementation. Journal of Library Metadata, 16(2), 69-79. doi:http://dx.doi.org/10.1080/1051712X.2016.1215730

16. Yiotis, K. (2008). Electronic theses and dissertation (ETD) repositories. OCLC Systems & Services, 24(2), 101-115. doi:http://dx.doi.org/10.1108/10650750810875458.

……………………….