

HARVESTING ETD METADATA FROM 'SHODHGANGA' TO NATIONAL DIGITAL LIBRARY OF INDIA: PRESENT SCENARIO AND IMPLEMENTATIONS



ETD 2022
Novi Sad
Sep 7 – 9, 2022

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 allow an initial assessment of compatibility. It provides an avenue for merging information or for exchanging information between systems. Integrating Indian ETDs based on metadata has received wide attention for better management and utilization. This study explores the current state of metadata harvesting of the Shodhganga or Indian ETD repository to the 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 significant objectives realized through this activity includes increased visibility and enhanced discoverability of resources.

Dr. Sukanta Kumar Patra

Librarian, Vidyasagar College For Women
(under University of Calcutta), West Bengal,
India

Mousumi Patra

Librarian, Govt. College of Nursing, ID & BG Hospital
West Bengal, India

INTRODUCTION

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.

OBJECTIVE

- Need for metadata harvesting;
- To explore the metadata challenges;
- Current state of metadata harvesting from Shodhganga to NDL.

METHODOLOGY

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.

RESULTS

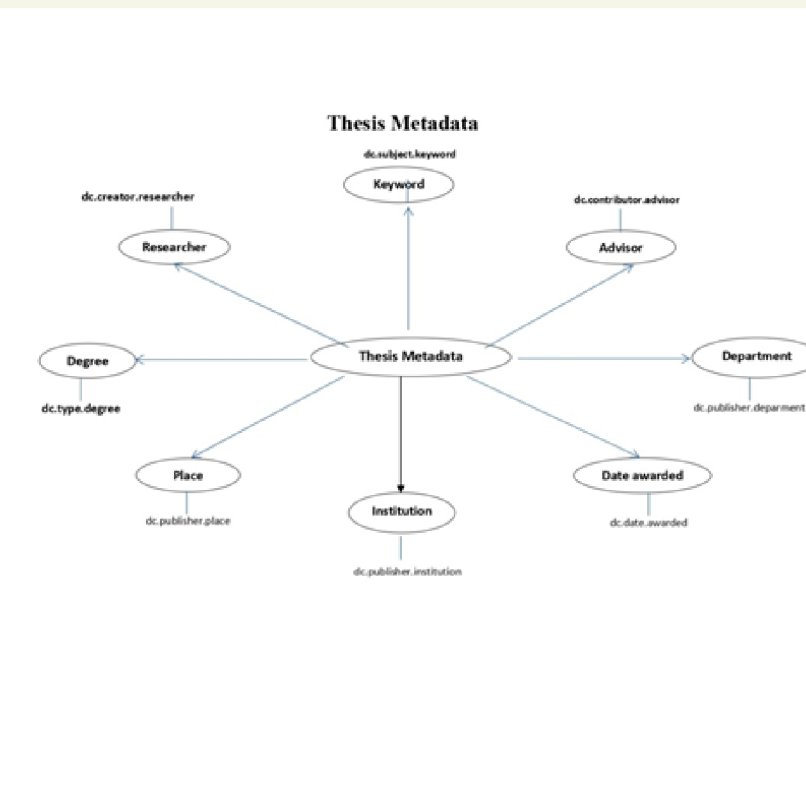
Online availability of electronic thesis through centrally maintained digital repositories, ensures easy access and archiving of Indian doctoral theses but also helps in the long run, raising the standard and quality of research, this would overcome the serious problem of duplication of research and 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.

ANALYSIS

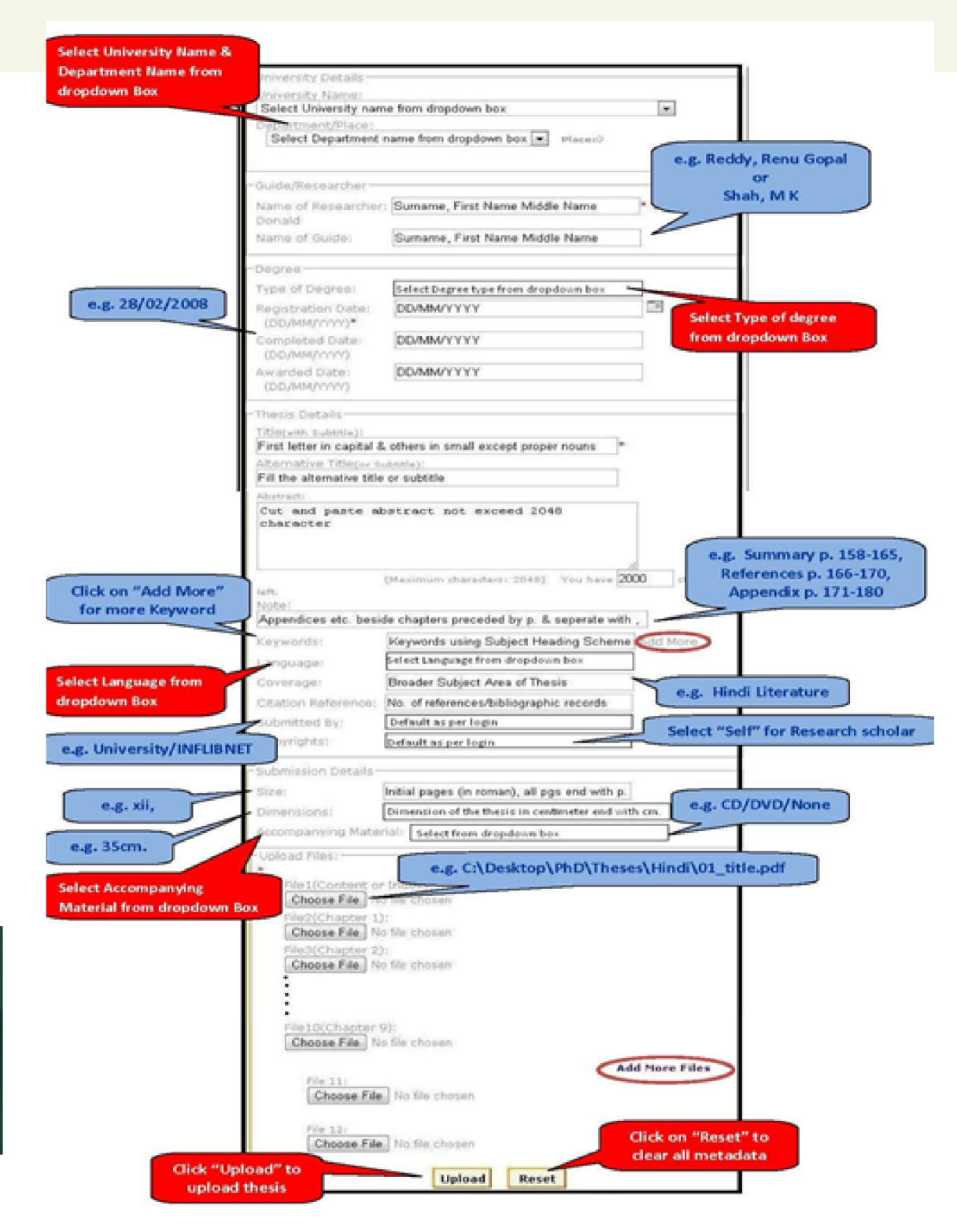
OAI-PMH provides an application-independent interoperability framework based on metadata harvesting. If a set of digital resources are described in the same metadata schema according to similar rules, then a DL application can use this metadata to provide services more easily. The OAI-PMH protocol requires 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.

Thesis Metadata:

NDL has been envisaged to be a huge repository of digital content and harvesting from varying domains and content categories. Variation of content category motivated the exploration of different metadata standards to define metadata schema for NDL. Based on the variation in contents, metadata schema in NDL has been categorized into three classes, such as Generic Metadata, Educational Metadata and Thesis Metadata (Qualified DC). Shodhganga thesis metadata standard has been used to represent NDLI-ETD metadata set.



Metadata	Multi-value?	Standard	ShortDescription
dc.date.awarded	single	yyyy-mm-dd	Date on which the degree was awarded
dc.type.degree	single	Controlled Vocabulary	Degree awarded to the recipient
dc.creator.researcher	single	lastName, firstName	Researcher
dc.publisher.institution	multi	country, state, city (Capitalize each word)	Name of institution
dc.publisher.department	single	country, state, city (Capitalize each word)	Name of the department
dc.publisher.place	multi	country, state, city (Capitalize each word)	Place, typically the institute location
dc.publisher.date	single	yyyy-mm-dd	Publishing data
dc.contributor.advisor	multi	lastName, firstName	The supervisors on the thesis



CONCLUSION

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 doctoral 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.

Related literature

1. Das, A., & Sutradhar, B. (2016). 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/10.1108/0974800910954271>
3. Foulonne, M., & Riley, J. (2008). Metadata for digital resources: implementation, systems design and interoperability. Oxford: Chandos Publishing, p.124-125.
4. Hakimjavad, 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.iiitkgp.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. (2016). Interoperability and retrieval. UNESCO, p.28-31. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000262195>
9. Park, E. G., & Richard, M. (2011). Metadata assessment in e-theses and dissertations of canadian institutional repositories. The Electronic Library, 24(3), 394-407. doi:<http://dx.doi.org/10.1108/02640471111141124>
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. Yiotsis, K. (2008). Electronic theses and dissertation (ETD) repositories. OCLC Systems & Services, 24(2), 101-115. doi:<http://dx.doi.org/10.1108/10650750810875458>