Understanding metadata item records for electronic theses and dissertations in institutional repositories

Abstract. Institutional repositories of theses and dissertations require a high quality metadata to make its items relevant, discoverable and shareable. There are numerous metadata standards available with predefined metadata tags/fields for describing documents in a digital repository. In our study, the data is collected from the Open Directory of Open Access Repositories (OpenDOAR), to observe Institutional Repository software installed by universities, then examine the metadata item record of a thesis and document its distinct attributes. Preliminary study indicates that the countries having a large number of ETD repositories indexed in OpenDOAR have quality metadata. For each item their metadata record provides DOI, Copyright Statement, Degree Date awarded, use of controlled vocabulary and their repository interface maintains Researchers Profile, Altmetrics score alongside students' thesis. Some of the repositories are also having recommendations or findings of theses listed as separate entity. It was observed that in some repositories they mention about the supplemental materials and research data availability and integration, further details of which will be discussed in the paper.

1 Background

Institutional repositories of theses and dissertations require a high quality metadata to make its items relevant, discoverable and shareable. There are numerous metadata standards with predefined metadata tags/fields for describing documents in a digital repository. Popular metadata standards like DC, DDI, VERS, PREMIS, METS, MADS, EAD, TEI, ONIX, MARC, MARC-XML, MODS, NDLTD ETD-MS etc. are generally used to create metadata from conventional documents to interoperable digital documents. Over the years, metadata of theses and dissertations have improved with modification of many additional data elements. It has become more reliable, useful, relevant and a valuable resource for researchers and academic communities to retrieve scholarly materials from an institutional repository. With this, the repository interface has also improved with the inclusion of many additional features to enhance user experience in the scholarly landscape. Earlier, the research output in the form of students' thesis and dissertations published by the institutions used to have limited access and scope of further enhancements. With the scientific advancements, the revision and refinement of research has become possible. Furthermore, the demand for data and metadata residing in the thesis has risen among scientists, researchers and academic

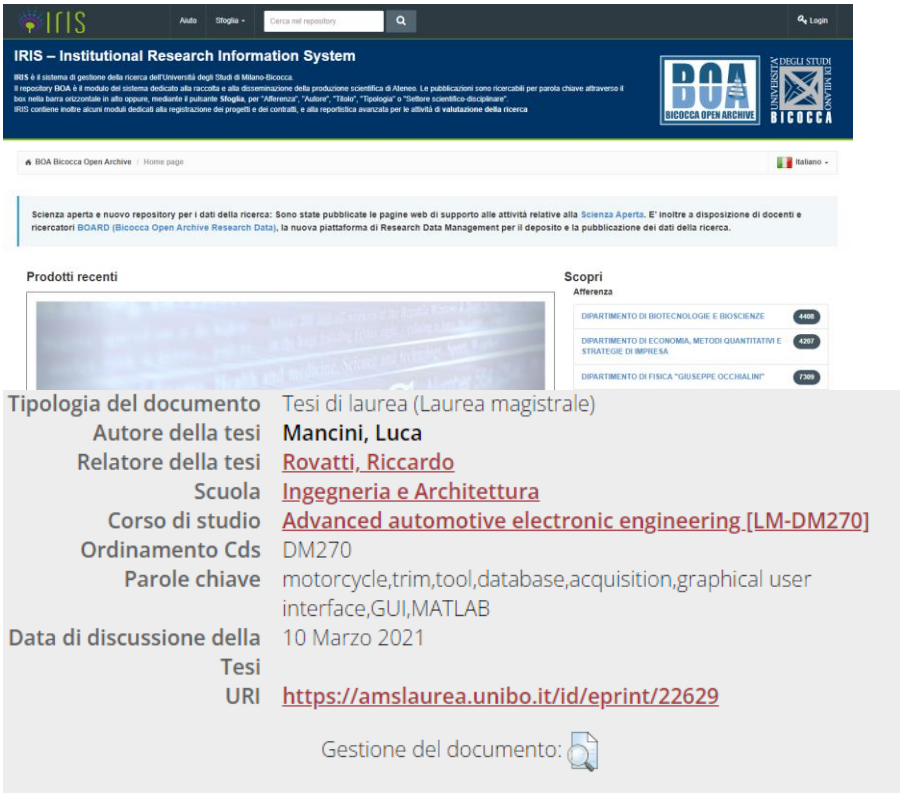
communities. They advocate for sharing and management of such data in universities. This paper examines the metadata item record of the thesis in a university repository across the countries for its distinct attribute and observes the latest features in their repository interface.

2 Objective

- To observe the metadata record of a thesis for its distinct field/data element in an institutional repository
- To examine the interface of the university thesis repository for the characteristic features
- To identify additional data elements described in their thesis record
- To trace availability of research data information in thesis metadata item record

3 Methodology

In our study, the data is collected from the Directory of Open Access Repositories, OpenDoar. It provides a list of digital repositories established by universities, research organisations worldwide. It is a collaborative project of University of Nottingham and Lund University funded by OSI, JISC, SPARC Europe and CURL, launched in the year 2005. As of April 2022 it shows 5858 Open Access repositories indexed in the OpenDoar platform. These institutional repositories of universities have a vast number of resources such as journal papers, conference proceedings, books chapters etc. In our study we have limited our search to examine university archives that have ETD content. From each country we have considered five university repositories starting from the first five alphabets. In case the examined repository does not have theses collection or is a research organisation we have excluded that repository and moved to the next repository in the list. In order to check a metadata item record and its attribute we tried to search the repository with the keyword say for example "Road" maintaining uniformity in search across university thesis repositories. We further examined their record structure and recorded the observations in a shared Google sheet. The data is analysed by citing the distinct features of metadata records provided by universities across countries. We used Google translator where there was a difficulty in understanding the record structure and the repository interface.



4 Result and Discussion

Metadata elements described for a particular thesis in an institutional repository vary across universities. Preliminary study shows that countries with a large number of repositories have high-quality metadata.Their metadata records are more comprehensive. In addition to this, the institutional repositories of some universities have readership maps, downloads, share, save, citation, copyright, licence, student profile such details are observed. It could be expressed that the Universities that have online thesis submission systems, at such places where students fill the metadata record of their thesis in the available template is verified by the faculty advisor and then the library further checks for its completeness and accuracy and preserve their record. And in the universities that do not have an online thesis submission system in place, in that situation, librarians are involved in developing the metadata record of the thesis into their LMS. They develop the metadata record from the provided CDs or print material or from digitised ETDs. In such a scenario, there is less autonomy available with the researcher or a faculty advisor to update or revise a thesis record or present the additional notes.

5 Conclusion

The preliminary results show that counties having a large number of repositories have well defined record structure and there are many features and functionalities incorporated in their repository interface. The metadata describing the thesis should have comprehensive information to facilitate retrieval from a repository. It should be checked that the metadata record is accurate, complete and reliable. Furthermore, for FAIR principles to be fully embraced it is essential that metadata administrators observe the quality and data elements of each record for seamless discovery of items. More in depth results of the study will be shared at the conference.

6 References

1. Woodley, M. S. (2008). Crosswalks, metadata harvesting, federated searching, metasearching: Using metadata to connect users and information. Getty Research Institute
2. .Park, J. R., & Tosaka, Y. (2010). Metadata creation practices in digital repositories and collections: Schemata, selection criteria, and interoperability. *Information Technology and Libraries*, *29*(3), 104-116.
3. Park, J. R., & Tosaka, Y. (2010). Metadata quality control in digital repositories and collections: criteria, semantics, and mechanisms. *Cataloging & classification quarterly*, *48*(8), 696-715.
4. Ulrich, H., Kock-Schoppenhauer, A. K., Deppenwiese, N., Gött, R., Kern, J., Lablans, M., ... & Ingenerf, J. (2022). Understanding the Nature of Metadata: Systematic Review. *Journal of medical Internet research*, *24*(1), e25440.