

Assessing and Promoting Metadata Quality for Electronic Theses and Dissertations in Institutional Repositories Using a Policy-Driven Approach

Cecilia Kasonde <19001415@student.unza.zm>

Lighton Phiri <lighton.phiri@unza.zm>

Department of Library and Information Science
University of Zambia
Lusaka, ZAMBIA

26th International Symposium on Electronic Theses and Dissertations (ETD 2023)
Gandhinagar, India | October 26–28, 2023

About Us

- The Data Lab research group at The University of Zambia is composed of faculty staff and students—undergraduate and postgraduate—working in three main areas :
 - Data Mining
 - Digital Libraries
 - Technology-Enhanced Learning

<http://datalab.unza.zm>

Outline

- Introduction
- Problem Statement
- Research Objectives
- Methodology
- Results and Discussion
- Conclusion
- Recommendation

Introduction: Metadata in ETDs (1/3)

- Electronic theses and dissertations (ETDs) have revolutionized the way academic research is disseminated and accessed in higher education institutions (HEIs). With the emergence of institutional repositories (IRs), ETDs have found a secure and accessible home.
- These repositories play a vital role in preserving scholarly output, facilitating ease of sharing, and increasing the visibility of research contributions. However, the success of IRs in achieving these goals hinges on the quality of metadata associated with the ETDs they host.

Introduction: Metadata in ETDs (2/3)

- Metadata is the crucial information about a digital resource that enables effective discovery, retrieval, and management.
- In the context of ETDs, metadata includes essential details such as title, author, abstract, keywords, subject classifications, and more.
- High-quality metadata enhances discoverability, aids browsing, and ensures long-term preservation, contributing significantly to the overall effectiveness and usefulness of IRs.

Introduction: Metadata in ETDs (3/3)

- The study highlights the results and implications of a study that delves into the challenges of metadata quality for ETDs in institutional repositories.
- The research aimed to assess the existing metadata quality in all HEIs in Zambia, explore factors affecting metadata quality, and identify policy-centric approaches to promote metadata excellence.

Problem Statement (1/2)

- The discovery of digital library information objects highly depends on the quality of metadata, if the metadata quality is poor, so is the discovery of digital library information objects.
- Tani, (2013) observes that Metadata is a key element in the digital library domain. Actually, such a kind of data has characterized this domain since the beginning and for time it has been, in some cases this is still the case, the sole data digital library and repository systems have been requested to manage since they act as placeholders for real resources.

Problem Statement (2/2)

- Because of this core role, metadata quality is a characteristic that is directly associated with the digital library value and effectiveness.
- This study therefore intended to investigate the best practices that can lead to effective electronic theses and dissertations metadata and answer the question: What are some of the problems with metadata quality of ETDs ingested into the HEI IRs and what mechanisms can facilitate ingestion of quality metadata?

Research Objectives

- The study explored the assessment of ETD metadata quality and the analysis of the ingestion process in HEI IRs, aiming to establish the role of institutional repository policies in promoting metadata quality.
- Specifically, the study sought to comprehensively analyse the quality of ETD metadata in all HEIs in Zambia and, additionally, identify policy-centric factors that negatively affect the quality of metadata.

Methodology (1/7)

- To achieve these objectives, a mixed-method approach was employed.

Quality of Metadata Ingested into HEI IRs.

- Involved surveying 62 registered HEIs under the Higher Education Authority (HEA) in Zambia. 53 Private HEIs and 9 Public HEIs(as at 28/01/23).
- The study participants were persons of contact from an official list of universities that was collected from the Higher Education Authority (HEA) website.
- All 62 Universities (Private and Public) were sampled.

Methodology : Quality of Metadata(2/7)

- Data was collected using an online questionnaire.
- An online questionnaire was used in order to determine HEIs that have IRs.
- Once the HEIs had been identified and their IR URLs verified, the researchers harvested data from the IRs for analysis.
- The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) was used to harvest metadata.
- A quantitative analysis of ETD metadata from all HEIs with functional IRs was conducted, by focusing on metadata completeness—the completeness metric was arrived at by ascribing scores to individual metadata elements, relative to the ETD-ms metadata standard.

Methodology : Quality of Metadata(3/7)

- The completeness metric was introduced to assess the quality of ETD metadata.
- This metric was developed by the researchers based on the NDLTD ETD-ms (Electronic Theses and Dissertations Metadata Standard), a widely recognized standard for ETD metadata.
- In the completeness metric, each individual metadata element was assigned a weight.
- Mandatory metadata elements were given a weight of 1.
- Optional metadata elements were given a weight of 0.5.

Methodology : Quality of Metadata(4/7)

- A score was assigned to each ETD based on the presence and absence of metadata elements.
- For example, if an ETD included all mandatory elements (with a weight of 1 each) and some optional elements (with a weight of 0.5 each), it would receive a total score based on the sum of the weights of the elements it contained.
- The completeness score for each ETD was determined by dividing the calculated score by the recommended number of metadata elements according to the NDLTD ETD-ms (usually a fixed number) and then multiplying by 10 to scale it to a score out of 10.
- The formula for calculating completeness score would be: $(\text{Calculated Score} / \text{Recommended Number of Metadata Elements}) * 10$.

Methodology : Quality of Metadata(5/7)

- This completeness metric allowed the researchers to quantitatively evaluate the quality of ETD metadata in Institutional Repositories.
- It provided a standardized and structured way to measure the completeness and compliance of the metadata with established standards, with a completeness score reflecting the overall quality of each ETD's metadata.
- This approach enabled the researchers to identify areas for improvement and assess the effectiveness of metadata management within the IRs.

Methodology :Ingestion Process of Metadata (6/7)

- The sample was drawn from HEIs that have IRs. This comprised, individual contacts of IRs. This information was obtained from the Institutional URL provided in the first objective.
- Purposive sampling method was used to select only institutions with functional IRs.
- Primary data was collected using online interviews.
- Target population were library staff involved in the ingestion process.

Methodology :Institutional Repository Policy (7/7)

- In order to identify factors that affect metadata quality, interviews were conducted by 16 key stakeholders, from 8 HEIs, involved in drafting IR policies and, additionally, individuals involved in the ingestion of ETDs into IRs.
- Convenience sampling was used to draw a sample from HEIs that have functional IRs.
- Primary data was collected using online interviews.
- The interviews were recorded and transcribed using NVivo.

Results and Discussion: State of metadata quality in HEI IRs(1/6)

S/N	Institution	Mandatory Elements- 1	Optional Elements – 0.5	Score out of 10	Completeness Score
1	Apex	Title,Creator,Date,Subject,Identifier	Description,Publisher	6	4.29
2	Chalimbana	Creator, Date, Identifier, subject, title, type	Description,language	7	5
3	Mulungushi	Creator, Date,Identifier,subject,title,type	Publisher,language	7	5
4	Nkrumah	Creator, Date,Identifier,subject,title,type	Publisher,language	7	5
5	Texila	Creator, Date,Identifier,subject,title,type	Publisher,language	7	5
6	UNILUS	Creator, Date, Identifier, subject, title, type	Description,language	7	5
7	UNZA	Creator, Date,Identifier,subject,title,Type	Description,language, Publisher	7.5	5.34
8	ZCAS	Creator, Date,Identifier,subject,title	Publisher	5.5	3.93

Results and Discussion: Completeness Scores (2/6)

- University of Zambia achieved the highest score of 5.34 for metadata quality.
- Other institutions scored between 3.93 and 5, indicating moderate quality with room for improvement.
- Missing metadata elements includes: Contributor, Format, Coverage, Rights, and Thesis Degree.

Results and Discussion: Challenges of metadata (3/6)

- Lack of adherence to international standards like ETD-ms.
- ETD-ms standard includes 14 vital metadata elements.
- Missing elements vary across institutions, indicating inconsistencies in metadata practices and standards.
- Many optional elements missing, including contributor, format, coverage, rights, and thesis degree.
- Tendency to develop institution-specific standards. Hinders interoperability and aggregation.

Results and Discussion: Missing Elements(4/6)

Addressing Missing Elements:

- Including Contributor for proper attribution.
- Adding Format for technical characteristics.
- Incorporating Coverage for contextual information.
- Ensuring Rights for legal and ethical considerations.
- Including Thesis Degree for academic context.

Results and Discussion: IR Policy (5/6)

- To address the issue of metadata quality , the study explore the role of institutional repository policies.
- To provide structure, governance, and guidance for the effective management of digital content within educational institutions.
- They ensure the repository's sustainability, the quality of its content, and its alignment with international standards ETD-ms, institutional goals and values.

Study Revealed

- Six out of eight institutions lack IR policies
- Absence of formalized guidelines for incorporating digital content

Results and Discussion: IR Policy (6/6)

Interview Respondents stressed the need to :

- Incorporating standardized metadata practices
- Provide training and support for metadata creators
- Implement quality control mechanisms
- Adhering to established metadata standards, like ETD-ms
- Compliance with all essential metadata elements
- Once embedded in repository policies, would enhance the overall discoverability and accessibility of ETDs.

Conclusion

- Effective ETD management requires comprehensive attention to metadata quality.
- In the era of digital scholarship, institutional repositories have become pivotal platforms for archiving and disseminating electronic theses and dissertations.
- The quality of metadata associated with these digital resources plays a significant role in facilitating discoverability, accessibility, and long-term preservation.

Recommendation

Institutions should adopt policy-driven approaches to enhance metadata quality.

- Ensuring Consistency and Compliance
- Preserving and Archiving Content
- Access Control and Copyright Compliance
- Content Submission and Workflow
- Metadata and Descriptive Standards
- Quality Assurance
- Interoperability and Standards

Bibliography

- 1. Beall, J. (2006). Metadata and Data Quality Problems in the Digital Library. University of Colorado at Denver and Health Sciences.**
- 2. Lynch, C. A. (2003). Institutional Repositories: Essential Infrastructure for Scholarship in The Digital Age. portal: Libraries and the Academy,3(2),327336.https://doi.org/10.1353/pla.2003.0039**
- 3. Park, J., & Lu, C. (2009). Metadata professionals: Roles and competencies as reflected in job announcements, 2003–2006. Cataloging & Classification Quarterly, 47, 145–160.**
- 4. Phiri, L. (2018). Research visibility in the global South: towards increased online visibility of scholarly research output in Zambia. In Proceedings of the IEEE International Conference in Information and Communication Technologies. [Http://dspace.unza.zm/handle/123456789/572](http://dspace.unza.zm/handle/123456789/572)**

Q & A Session

- **Comments, concerns and complaints?**