

Towards Metadata Completeness in National ETD Portals for Improved Discoverability

By

Adrian Chisale and Lighton Phiri

(University of Zambia, Department of Library and
information Science)

26th International symposium, ETD, Gujarat, India 26—28 October
2023.

.

About us

Data lab research group from UNZA is composed of the faculty staff, student (undergraduate and postgraduate) working in main three areas

- Data mining
- Digital Libraries
- Technological enhance learning

<https://datalab.unza.zm/project/zambia-national-etd-portal>

Digital Libraries

The field of Digital Libraries (DLs) generally involves the study of digital collections of information and corresponding network-based services used to retrieve data from the collections. DLs are in effect information systems that are used to persistently store digital objects, manage the digital objects and, facilitate access to the digital objects.

Our focus in the field of DLs, as a research group, mostly involves experimenting with techniques that can potentially facilitate efficient and effective access to digital objects stored in DLs.

Project Members

Zambia National Electronic Theses and Dissertations Portal

The Zambia National Electronic Theses and Dissertations (ZANETD) portal is meant to be a national ETD portal that will aggregate ETDs from all the the Higher Education Institutions (HEIs) in Zambia.

Project Members



Lighton Phiri
Academic Staff



Adrian Chisale
Postgraduate Student
MA Library and Information Science



Cecelia Kasonde
Postgraduate Student
MA Library and Information Science



Angela Banda
Postgraduate Student
MA Library and Information Science



Frazer Nyambe
Undergraduate Student
B.ICT With Education



Richard Mufuzi
Undergraduate Student
B.ICT With Education



Geoffrey Ngoma
Undergraduate Student
BA Library and Information Science



Nkole Mulenga
Undergraduate Student
BA Library and Information Science

Project Publications

Data Mining

With the proliferation of data, the field of Data Mining has gained rapid popularity. Data Mining focuses on the discovery of patterns in large datasets by making use of statistical and machine learning techniques.

Our current focus involves leveraging machine learning techniques to facilitate efficient and effective delivery of services in the health and educational domains—two areas that are of significance in the so-called developing world.

Project Members



Lighton Phiri
Academic Staff



Robert M'sendo
Postgraduate Student
MSc Computer Science



Ivy Chisha
Undergraduate Student
B.ICT With Education



David Pungwa
Undergraduate Student
B.ICT With Education

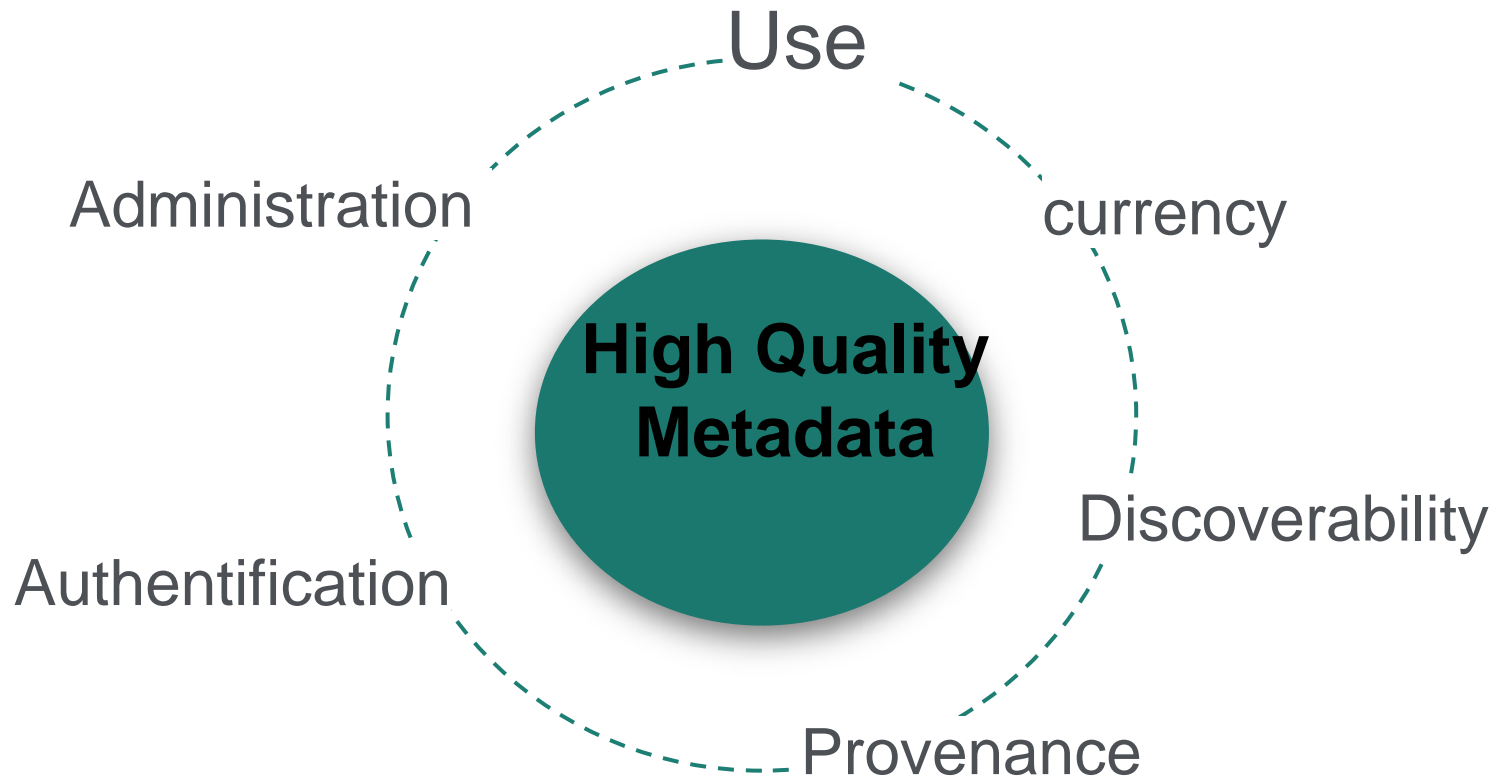


Outline

- ❖ Introduction
- ❖ Problem statement
- ❖ Research objectives
- ❖ Methodology
- ❖ Results and Discussion
- ❖ conclusion and future works

Introduction:

Core bibliographic function



(Park, 2009)

Introduction cont.

- ❖ The need for **adequate metadata** is a necessity for all digital objects.
- ❖ However, the increased rate at which Electronic Theses and Dissertations (ETDs) are being generated requires
- ❖ that much emphasis is placed on ETD metadata due to the heterogeneous sources .i.e various (HEIs).
- ❖ The need is further necessitated by the increase in the number of downstream aggregation services

Introduction Cont


Downstream aggregation services




Global portals (NLTD's UC and OATD)



National portal (e.g
ZANETD portal)



Higher Education Institutional
Repositories (HEIRs) in Zambia



<http://union.ndltd.org>

<https://oatd.org>

Problem statement

- ❖ In year 2018, Zambia Launched and is currently implementing the National ETD Portal project with intent to aggregate ETD metadata from all HEIs.
- ❖ Large quantities of metadata from heterogeneous IRs **necessitates** the need effective ways of Searching and browsing of ETD metadata harvested from external sources.
- ❖ To be effective, this calls for standardized approach in metadata creation from all HEIRs.

Problem statement cont.

- ❖ Poor ETDs metadata quality is among the challenges the global downstream services are facing (Suleman, 2012 & Phiri, 2018).
- ❖ However, the extent has not been looked into from Zambian Higher Education Institutional repositories as most of them are recent establishments.
- ❖ Hence, the need for research

Research Objectives

1. To identify Higher Education Institutions that has functional and interoperable online Institutional Repositories in Zambia.
1. To determine the relative quality of ETD metadata originating from HEIs in Zambia.
1. To investigate the sources of missing metadata elements for the automated generated using NLP techniques

Methodology

Identifying HEIs that has functional and interoperable online Institutional Repositories in Zambia

- Questionnaires and Semi structured interviews were used to identify HEIs: that offer postgraduate programs and HEIs with functional and interoperable IRs.

Determining the relative quality of ETD metadata originating from HEIs in Zambia

- Metadata from HEIs with functional IRs were harvested using the Open Archives Initiative Protocol for metadata Harvester (OAI-PMH) and subsequently analyzed in order to assess the metadata completeness during ingestion, relative to the **ETD-MS metadata standard. and Those who never activated the OAIPMH were done manually** (Lagoze et al., 2002; Hickey, Pavani & Suleman, 2010).

Methodology Cont.

Investigation of the sources of missing metadata elements for the automated generated using NLP techniques.

➤ **Content Analysis**

policy documents: analyzed against prescribed standard outline and format. while the ETDs manuscript: were analyzed to ascertain location of metadata

- **At the same time identifying potential sources of missing ETD-MS metadata elements.**

Methodology cont.

- ❖ The saturation principle was used to determine the sample size for document analysis in this study.
- ❖ “This stage is reached when no new data is gathered and researchers are no-longer gaining new insights” (Merriam & Tisdell, 2016).
- ❖ Microsoft Excel was used to analyze quantitative data collected which resulted in graphs and tables.

Results

- ❖ Zambia had 61 Higher Education Institutions registered with the Higher Educational Authority as at 25/6/2022 (Higher Education Authority, 2022).
- ❖ Out of 61 HEIs, we had 43 responses.
- ❖ Further investigations from the other 18 registered institutions indicated that they
- ❖ were not operational and some were closed.

RESULTS CONT

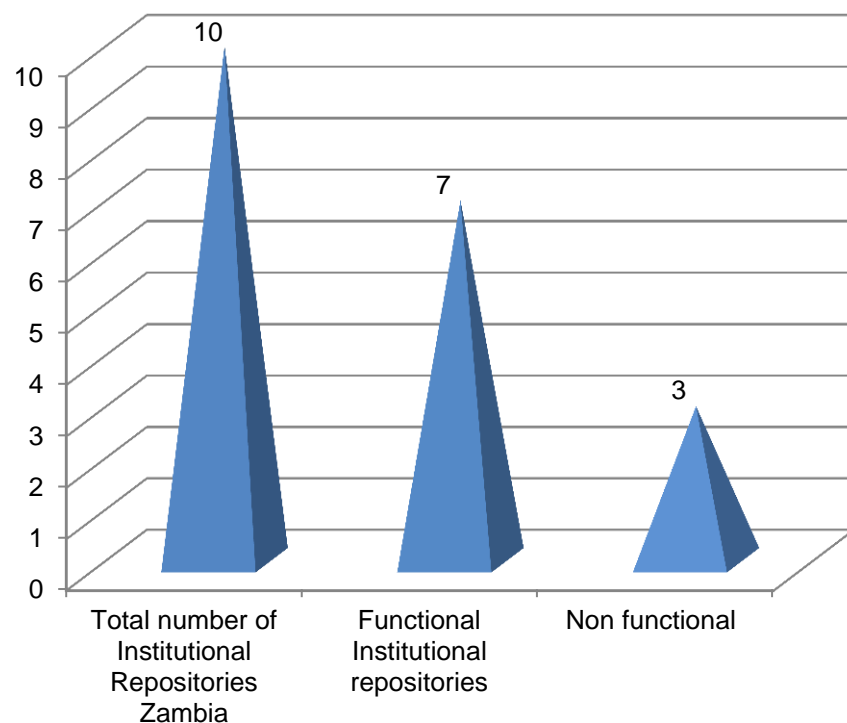
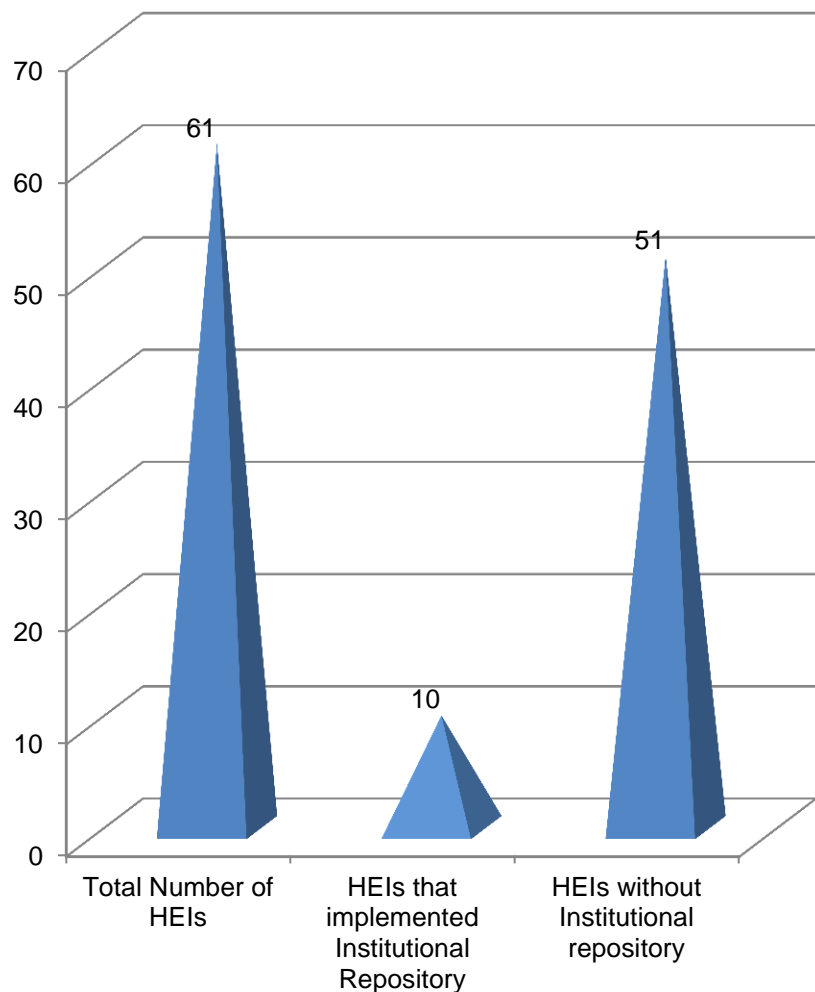
- ❖ Therefore this analysis was based on unique and completed responses only.

HEIs that have functional and interoperable online Institutional Repositories in Zambia

- ❖ Results indicated that out of the 61 HEIs in Zambia, 28 offered postgraduate programmes, with a mere 10 (16.4%) had implemented IRs
- ❖ and 51 (83.6%) of HEIs were still managing their ETDs in analogue format.
- ❖ Out of 10 that implemented and adopted IRs in Zambia, 3 IRs were found to be non-functional
- ❖ 1 (2%) did not have ETDs content as shown figure 1 below.

Status of Institutional repositories in Zambia

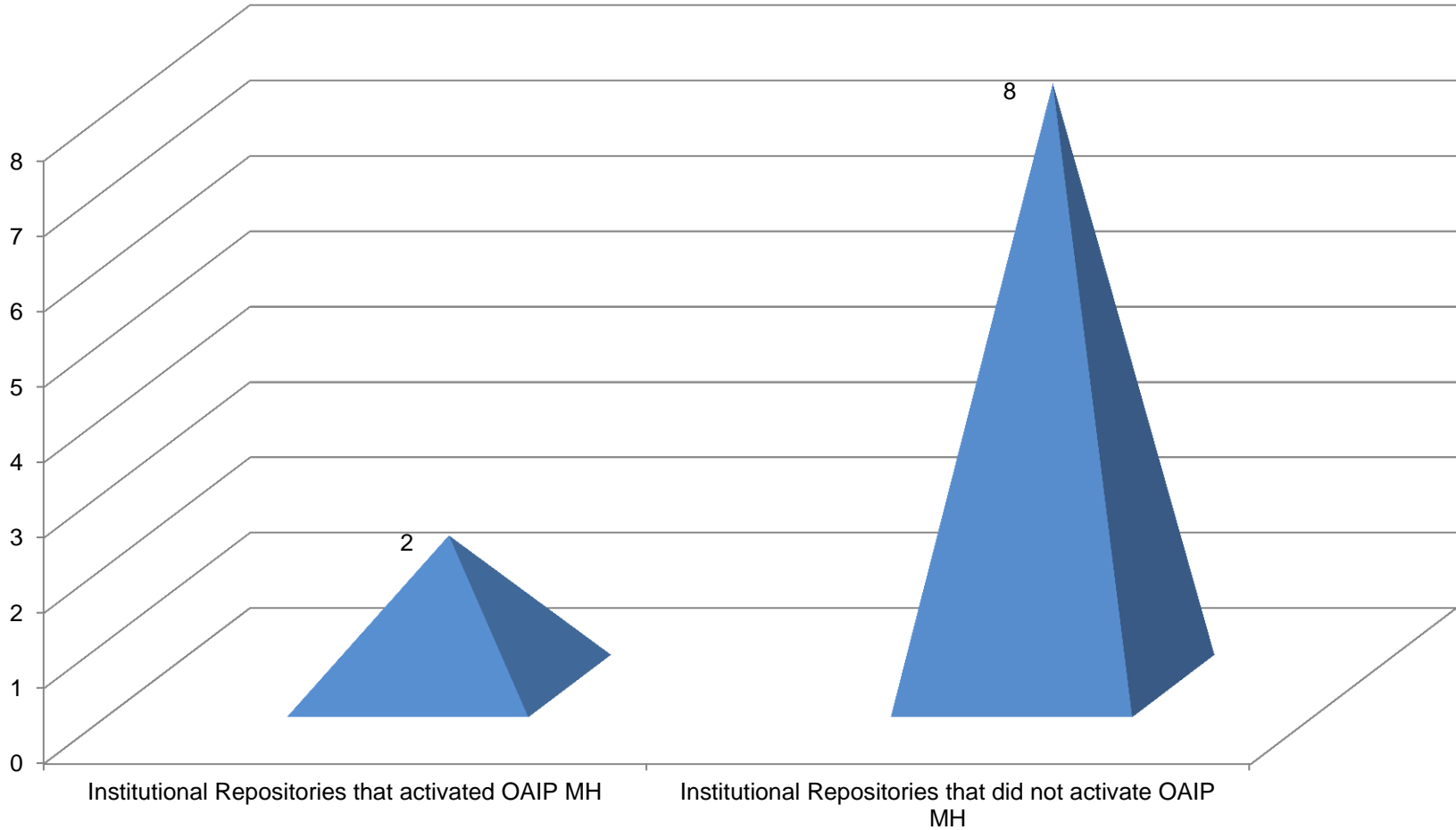
- **functionality status of implemented Institutional repositories**



Interoperable Repositories in Zambia

- ❖ It was found that eight (8) IRs representing 80% have not activated OAIP MH which define how metadata is exchanged amongst repositories.
- ❖ Only two (2) IRs representing 20 % activated the OAIP MH as shown in Figure 2 below

Interoperable Repositories in Zambia



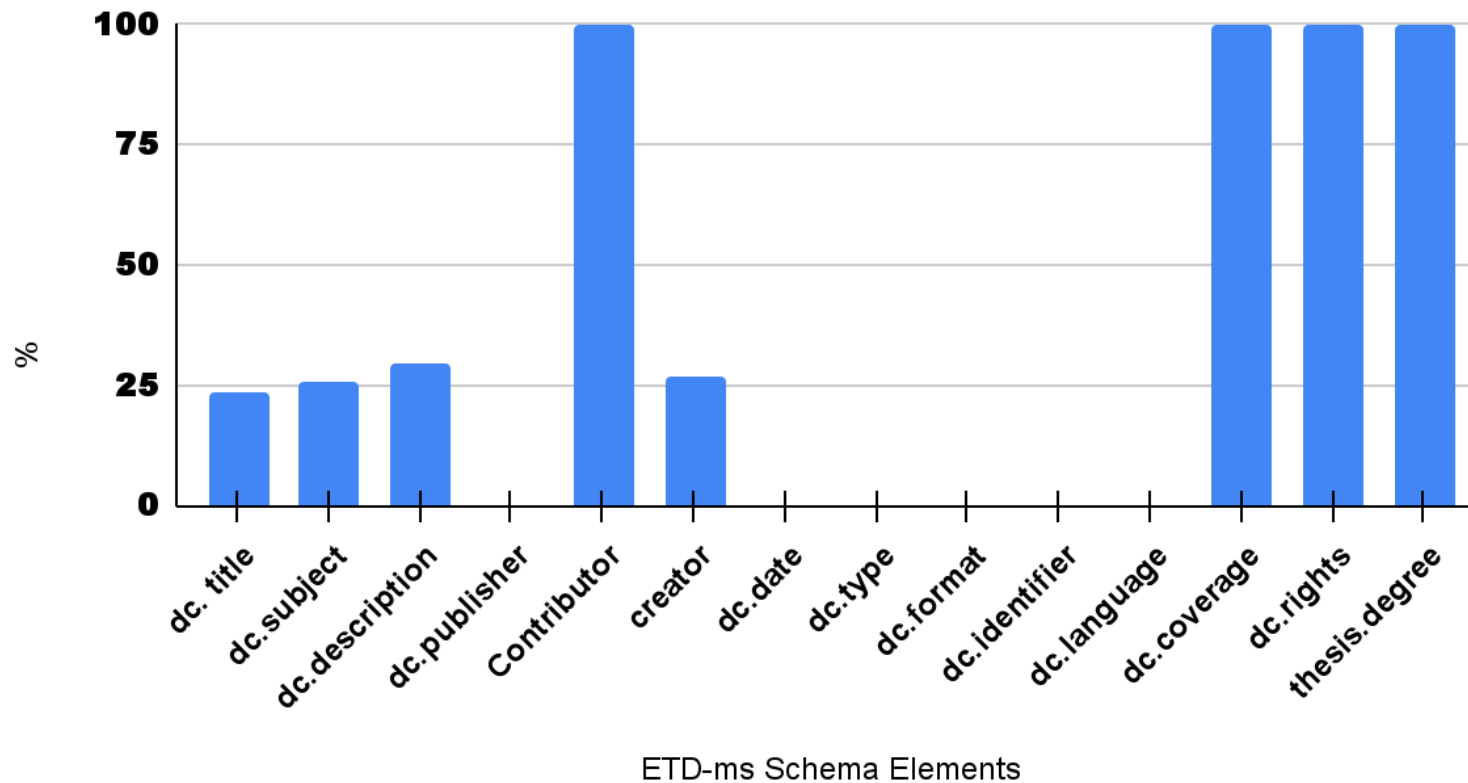
Situation analysis of metadata quality at The University of Zambia IR

- ❖ UNZA IR had 4003 records of ETDs as at(23rd February, 2022),
- ❖ Out of the 4003 records, 757 (23.3%) records had null title elements
- ❖ 845 (26.8%) null creator elements, 814 (25.5%) had null subject and 910 (29.4%) had null description elements.
- ❖ While coverage, right and thesis.degree had all their elements missing in all 4003 records. as shown below

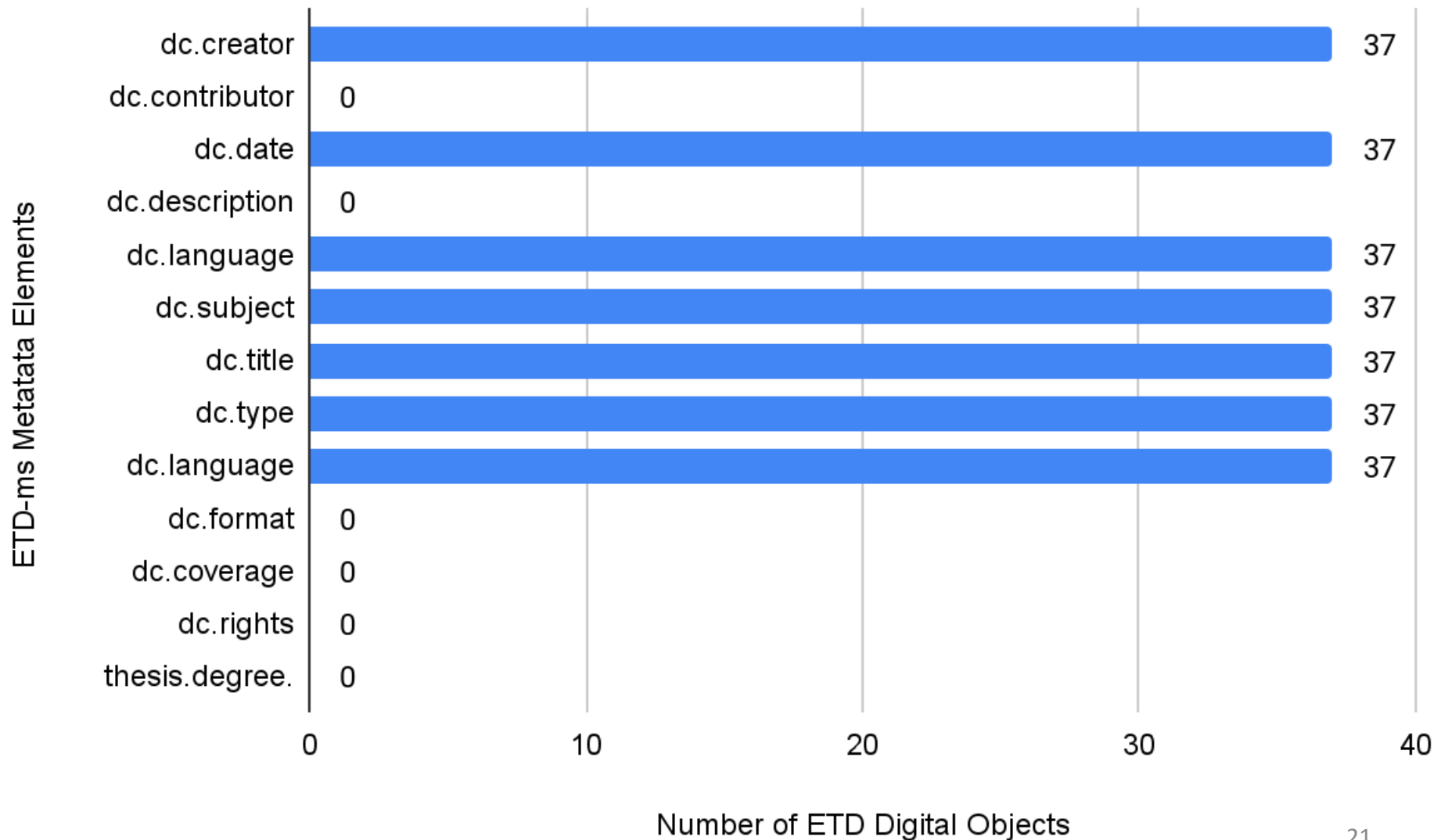
Situation analysis of metadata quality

University of Zambia IR

MISSING ETD-MS METADATA Elements



Metadata quality analysis at University of Lusaka IR



metadata elements used in HEIs in Zambia	Higher education institutions				
	Cavendish University	Mulungushi university	Zambia Institute of Accountancy studies University	Chalimbana University Institutional Repository	Lusaka Medical Apex University
dc.creator	✓	✓	✓	✓	✓
dc.contributor	x	x	x	x	x
dc.date	✓	✓	✓	✓	✓
dc.format	x	✓	x	x	x
dc.identifier	✓	✓	✓	✓	✓
dc.description.	✓	✓	✓	✓	✓
dc.coverage	x	✓	x	x	x
dc.language	✓	✓	✓	✓	x
dc.publisher	✓	✓	✓	x	✓
dc.subject	✓	✓	✓	✓	✓
dc.title	✓	✓	✓	✓	✓
dc.type	✓	✓	✓	✓	✓
dc.rights	x	x	✓	x	x
thesis.degree	x	x	x	x	x

Location of missing Metadata in manuscripts

- ❖ Policy documents and ETDs manuscripts analyzed indicated potential sources of missing metadata elements from various IRs.
- ❖ For instance, while supervisor/advisor details were found on the title page of most manuscripts for most HEIs,
- ❖ these details were not available in manuscripts for some HEIs as shown below

Policy documents: Standard outline on how ETD must be presented at UNZA



THE UNIVERSITY OF ZAMBIA

**REGULATIONS FOR POSTGRADUATE
TRAINING**

Preliminary page section

Section 26 Procedure for Preparation of Dissertations/Theses for Master's and Doctorate Degrees

26.1 A dissertation/thesis shall be satisfactory as regards **format** and literary presentation. It shall consist of preliminaries, text and end matter.

57

26.2 Preliminaries

The preliminaries shall consist of the following items, and in the order given:

(a) Title Page

The title page shall carry the following **information** suitably spaced and centred:

- (i) Full title of dissertation/thesis and sub-title, if any;
- (ii) If there is more than one volume, the total number of volumes and the number of the particular volume;
- (iii) By;
- (iv) Student's full name (as recorded on his/her birth certificate);
- (v) The statement "A thesis/dissertation submitted to the University of Zambia in partial fulfilment of the requirements of the degree of (Degree) in (Subject)" or "A thesis submitted to the University of Zambia in fulfilment of the requirements of the degree of (Title of the Degree) in (Subject)";

Policy document from Mulungushi



MULUNGUSHI UNIVERSITY

DIRECTORATE OF RESEARCH AND POSTGRADUATE STUDIES

**GUIDELINES AND REGULATIONS
FOR
POSTGRADUATE PROGRAMMES**

2017

standing instruction on what to include on the preliminary pages

11.0 STRUCTURE OF DISSERTATION AND THESIS

11.4 Front and Title pages

The front of title page must appear as follows:

- i. Title of dissertation/Thesis in CAPITALS (well indented and centered) in 14 font size and bold. A sub title must be in Capital and small letters.
- ii. The candidate should insert the name at the centre of the title page. The name should be in Capital and Small letters font size 12 and bold.
- iii. Below the name the candidate should insert the following “A Dissertation/Thesis (whichever is applicable) Submitted in Partial Fulfillment (for Dissertation) or in Fulfillment (for Thesis) of the Requirements for the Degree of (insert name of Degree) of Mulungushi University. Just below, give the date (Month and Year) of completion of dissertation/thesis (the year when the final corrections to the dissertation/thesis were made).

11.5 Certification

- i. The supervisor must certify that he/she read the dissertation/thesis, and found it to be in a form that is suitable and acceptable for examination and award of degree.

location of supervisor metadata in UNZA ETD manuscript

Title page section

UTILIZATION OF PLANT FUNCTIONAL TRAITS IN MITIGATING OVER-EXPLOITATION OF HIGH ETHNOBOTANICAL VALUE PLANTS OF CHONGWE DISTRICT, ZAMBIA.

BY

MWAMBO MATTHEWS

2018254138

A dissertation submitted to the University of Zambia in partial fulfilment for the award of Master of Science in Tropical Ecology and Biodiversity

The University of Zambia

Lusaka

2023

Activate Windows
Go to Settings to activate

Acknowledgement section

ACKNOWLEDGEMENTS

Great thanks go to my supervisor Dr. David Chuba who was always available for help and guidance during my research and Mrs. Florence Nyirenda for helping in the identification of plants.

Snipping Tool

location of metadata in university of lusaka manuscript

Title page section



SCHOOL OF POST GRADUATE STUDIES

EXAMINING THE IMPLEMENTATION OF SEXUALITY AND REPRODUCTIVE
HEALTH EDUCATION AMONG ADOLESCENTS IN SELECTED PRIMARY
SCHOOLS OF ZAMBIA

BY
MUKONKA REMMY
PHDED20120668

SUPERVISORS:
DR. CHRISTINE MUSHIBWE
PROF. DAVID S. MWANZA

A Thesis Submitted to the School of Post Graduate Studies in fulfillment of the requirements of the award of Doctor of Philosophy in Education

acknowledgement section

Acknowledgments


I am sincerely and profoundly grateful to all the people who helped and supported me to make this thesis a reality. My work would not have been complete without the guidance and support of many people to whom I am greatly indebted. Many thanks go to Dr. Christine Mushibwe and Professor David. S. Mwanza who supervised this work with all the enthusiasm and dedication. Their guidance throughout the writing process is greatly appreciated.

I also wish to extend many thanks to a number of Doctors and Professors of the University of Lusaka who include Dr. Nathan Musonda, Dr. Muchindu, Ms. Julie Mwewa and many more for the support rendered to me during the foundation phase of this work. Other appreciation go to lecturers and staff of the post graduate office of the University of Lusaka for their guidance and support. Further, my appreciation are also extended to all the respondents who took time to respond to my questions during fieldwork. Your time and insights are greatly appreciated.

I am also grateful to all my fellow students with whom I have made it this far in the study program of the PhD in Education and the Doctor of Business Administration (DBA) for their encouragement and support during the time of research. I am very grateful to all the individuals mentioned and those not mentioned but did contribute to the success of my research work, I am very humbled to you all for the strength, support, advice, and encouragement you inspired in me to be able to complete this research.

location of ETD metadata in Chalimbana University Manuscript

Title page section



CHALIMBANA UNIVERSITY
DEPARTMENT OF DISTANCE EDUCATION
EDUCATION LEADERSHIP AND MANAGEMENT

COMPUTER NO.:5201920099
COURSE CODE: MELM
YEAR OF STUDY: SECOND COHORT

**ADMINISTRATIVE EFFECTS OF STEREOTYPES AND LABELLING ON FEMALE
PUPIL'S ACADEMIC ACHIEVEMENT AT GRADE 12. A CASE STUDY OF
SECONDARY SCHOOLS IN NAKONDE DISTRICT, ZAMBIA.**

BY
SYLVIA NKOLE NSELUKA

**A DISSERTATION SUBMITTED TO CHALIMBANA UNIVERSITY IN
PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF EDUCATION IN EDUCATION
LEADERSHIP AND MANAGEMENT**

Activat
Go to Se

Acknowledgement section

Acknowledgment

First and foremost, I would like to express my sincere gratitude to my dissertation supervisor DR B. Chirwa for her unwavering guidance, direction and feedback. I would like to appreciate you for your selfless approach which helped me to rise up to the occasion. Thank you for sharing your knowledge and your continuous direction and support. Your time, effort, advice and feedback are most appreciated. Indeed, without your support, I would not have been able to produce this paper.

Finally, I would like to express my sincere gratitude and appreciation to my family for their patience, inspiration and support throughout the period of my study.

HEIs	Metadata element missing	Location of missing metadata in ETD (Sections of ETD)
University of Zambia IR	Supervisor / Advisor details	Acknowledgements/ certificate of approval
Mulungushi University IR		Title page/ Supervisor's Recommendation/ Acknowledgement
Chalimbana University IR		Approval/ Acknowledgement/ Declaration
Cavendish University IR	Supervisor / Advisor details	Title page/ Declaration/ Acknowledgement/
Lusaka Apex Medical University IR		Title page/ Supervisor's declaration

Location of the missing Elements in manuscripts

HEIs	Metadata element missing	Location of missing metadata in ETD (Sections of ETD)
University of Lusaka IR	Supervisor / Advisor details	Title page/ Certificate of Approval/ Acknowledgement
Zambia Centre for Accountancy Studies University I R		Title page/ Acknowledgement

conclusion

- ❖ The identification of sources of missing ETD metadata, presents opportunity for automatically generation of missing
- ❖ metadata for implementation of effective National ETD portal.
- ❖ As part of current and future work, Natural Language Processing models are being developed to automatically generate missing metadata elements.
- ❖ All HEIRs are not ETD-MS compliant

THE END

QUESTIONS AND DISCUSSION

THANKS VERY MUCH FOR YOUR TIME

Bibliography

Bowen, G.A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9 (2), 27-40.
<https://doi.org/10.3316/QRJ0902027>

Hickey, T., Pavani, A., & Suleman, H. (2010). ETD-MS v1.1: An interoperability metadata standard for electronic theses and dissertations. *Networked Digital Library of Theses and Dissertations*.
<https://ndltd.org/wp-content/uploads/2021/04/etd-ms-v1.1.html#>

Higher Education Authority. (2022). *Higher Education Authority*. Higher Education <https://doi.org/10.1177/02666669211015429>

Lagoze, C., et al. (2002). *Open archives initiative protocol for metadata harvesting*. Available on <http://www.openarchives.org/OA/openarchivesprotocol.html>

Merriam, S. B., & Tisdell, E. J. (2016). Qualitative research: a guide to design and implementation (4th ed.). San Francisco, CA: Jossey Bass.

Park, J. R. (2009). Metadata quality in digital repositories. a survey of the current state of the art. cataloging & classification quarterly. <https://doi.org/10.1080/01639370902737240>

Phiri, L. (2018). Research visibility in the global South: towards increased online visibility of scholarly research output in Zambia. In IEEE International Conference in Information and Communication Technologies.
<http://dspace.unza.zm/handle/123456789/5723>

Suleman, H. (2012). The NDLTD Union Catalog: Issues at a Global Scale: Standard for Electronic Theses and Dissertations. Available on <https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/622568/ETD2012>