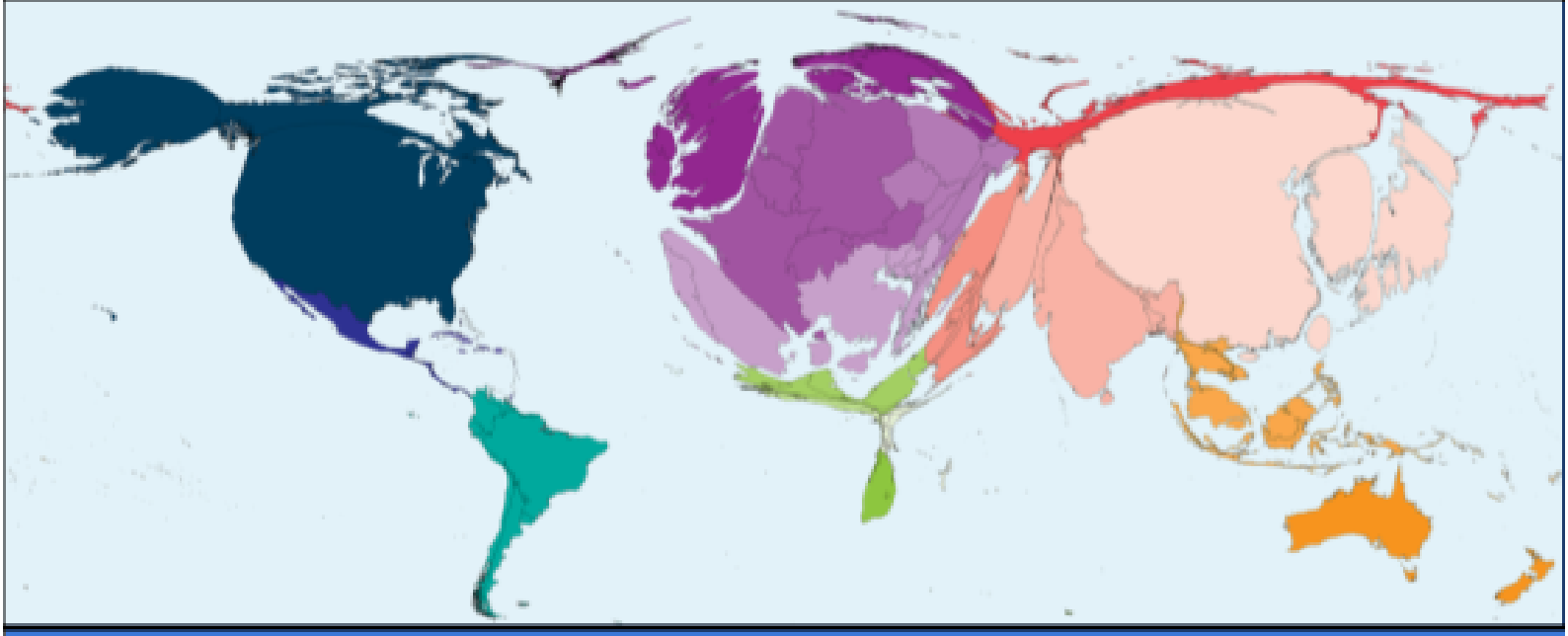


ETD 2024

27th International Symposium on Electronic Theses and Dissertations

"Electronic Thesis and Dissertation Visibility at a Global Scale"

November 4-6, 2024

Livingstone · Zambia

Contents

About

27th International Symposium on Eletronic Theses and Dissertations

The 27th International Symposium on Electronic Theses and Dissertations (ETD 2024) aims to bring together global leaders and researchers working in the broad areas of digital libraries, institutional repositories, scholarly research and electronic theses and dissertations. The theme of ETD 2024 is "ETD Visibility at a Global Scale", and will explore innovative approaches that make use of Electronic Theses and Dissertations (ETDs), including the use of modern-day Artificial Intelligence techniques such as Large Language Models and exploration of advances that will result in increased visibility of ETDs at a Global Scale.

Thank you so very much for coming to Livingstone, Zambia to participate in the ETD 2024 conference events. We sincerely hope you enjoy the conference and your short stay in Livingstone, and, more importantly have productive and worthwhile discussions and meet new people!

Zikomo Kwambili!

Best Wishes
Lighton Phiri (Chair)

Local Conference Organizing committee

Lighton Phiri, UNZA	Denny Nsokolo, HEA	Francina Makondo, UNZA
Phyela Mbewe, UNZA	Thabiso Mwiinga, UNZA	Kaoma Daka, UNZA
Matildah Muchinga, LAMU	Stein Mkandawire, ZAMREN	Buumba Dubeka, ZCAS
Adrian Chisale, UNZA	Dokowe Tembo, ZCAS	Cecilia Kasonde, KNU
Habeenzu Mulunda, UNZA	Francis Kawesha, HEA	Fabian Kakana, UNZA
Brian Munkondya, MU	Elijah Chileshe, UNZA	Mpande Ntumbo, ZAMREN

Volunteers

Mubanga Chibesa, UNZA	Gift Muwele, UNZA	Lwiimwe Shansonga, UNZA
Albertina Mooka, UNZA	Christabel Kunda, UNZA	

Programme committee

Ana Pavani	Pontifícia Universidade Católica do Rio de Janeiro
Charles Greenberg	NDLTD
Edward Fox	Virginia Tech
Gabriela Mejias	DataCite
Hussein Suleman	University of Cape Town
Iryna Kuchma	EIFL
Jian Wu	Old Dominion University
Joachim Schöpfel	University of Lille
Lazarus Matizirofa	University of the Witwatersrand
Lighton Phiri	University of Zambia
Maïté Roux	ABES
Mirjana Brkovic	University of Novi Sad
Nabi Hasan	Indian Institute of Technology Delhi
Ramesh Gaur	Indira Gandhi National Centre for the Arts
Shantashree Sengupta	P.E.S Modern College of Arts Science and Commerce
Tainá Batista De Assis	Brazilian Institute of Information in Science and Technology
William Ingram	Virginia Polytechnic Institute and State University

Partner Institutions and Sponsors

The ETD 2024 conference was organised by the Networked Digital Library of Theses and Dissertations (NDLTD) and, additionally hosted by The University of Zambia (UNZA) and co-hosted by The Higher Education Authority (HEA) of Zambia and Zambia Research and Education Network (ZAMREN).

Gold, Silver and Bronze Sponsors

Partner Institutions/Organisations

Full Papers Track

Improving the Mkulima Repository Content: Utilizing Theses, Dissertations, and LLMs for Agricultural Knowledge Dissemination in Kiswahili

Joseph P. Telemala[1]

[1] Sokoine University of Agriculture, Tanzania

The Sokoine National Agricultural Library (SNAL) at the Sokoine University of Agriculture (SUA) faces significant challenges in disseminating agricultural information to Swahili-speaking communities, as most research outputs are predominantly in English. This language barrier hinders the effective transmission of vital agricultural knowledge to key stakeholders in the agriculture-food value chain, who use Kiswahili in their daily activities. To address this gap, SNAL established the Mkulima Collection and Repository, dedicated to collecting agricultural content in Kiswahili. Despite these efforts, the Swahili content in the repository remains limited.

This study aims to enhance the Mkulima Repository by translating abstracts from English-language theses and dissertations using MarianMT, a machine translation (MT) model based on large language models (LLMs). The selected abstracts underwent pre-processing, machine translation, and subsequent quality assessment by multilingual experts.

Our findings reveal significant challenges in using LLMs like MarianMT for low-resource languages such as Kiswahili. While the MT system offers a rapid and scalable method for translating academic content, the accuracy and fluency of the translations were found to be suboptimal, as indicated by the evaluators. Common translation errors, particularly in agriculture-specific terminology and scientific names, highlight the limitations of current MT models in handling specialized agricultural content. These issues underscore the need for a more refined approach, including the development of a curated dataset of Swahili-English pairs that focus on agricultural jargon, as well as the integration of a knowledge base to improve the translation of scientific terms.

E-Theses and Dissertations in Zambia: A Case Study of Two Universities in Kabwe

Jive Lubbungu[1], **Chewe Mumba**[1]

[1] Kwame Nkrumah University, Zambia

This study investigates the successes and challenges in the implementation of Electronic Theses and Dissertations (ETDs) at Kwame Nkrumah University and Mulungushi University in Kabwe district, Zambia. Employing a qualitative research approach, data were collected from four purposively selected key informants using structured interview guides. These informants were personnel from the e-resources departments of the university libraries.

The findings reveal that both institutions share common attributes, such as the establishment of institutional repositories and the successful initial implementation of ETDs. However, the study identifies significant challenges, including staff resistance to depositing their dissertations into institutional repositories, a lack of expertise in configuring the repository platforms, and intermittent network connectivity. Thematic analysis was utilized to analyze the data.

The study concludes that while some progress has been made in the implementation of ETDs, the current state at the two institutions has not yet reached the desired level. To address these issues, the study recommends enhanced training programs for staff on the importance and use of ETDs, improved technical support and infrastructure for repository management, and strategies to foster a culture of compliance and participation among academic staff.

Posters Track

Posters

Automatic Electronic Thesis and Dissertation Guideline Verification For Consistently Formatted Manuscripts

Mubanga C. Chibesa[1], **Albertina Mooka**[1], **Gift Muwele**[1], **Lwiime Shansonga**[1], **Lighton Phiri**[1]

[1] The University Of Zambia

Abstract:

Higher Education Institutions worldwide enforce guidelines and academic approaches to ensure scholarly integrity and adherence to academic standards(Razı et al., 2019).The University of Zambia, is not an exception. Just like most HEIs it offers training to postgraduate students and one of the key aspects of postgraduate training is producing an Electronic Thesis and Dissertation manuscript The Directorate of Research and Graduate Studies (DRGS) provides guidelines which stipulate how ETD's should be formatted. Successful graduation is dependent on postgraduate students' manuscripts of which their conformance to the guidelines is a key aspect. Examiners and other relevant authorities are expected to verify and check if postgraduate students' manuscripts conform to the guidelines. However, the process of checking for conformance is a manual and tedious procedure, resulting in submission of inconsistently formatted manuscripts. Although this research seeks to ascertain the exact reasons why this is so, it is apparent that the primary task for the examiners is approving the content written in these manuscripts, and the secondary task being making sure that each and every manuscript is in conformance to the postgraduate guidelines and this is arguably an intense exercise. To address this challenge our project seeks to implement a tool that will automate the process of checking ETD's compliance against established postgraduate guidelines. The tool will leverage data mining techniques to perform these tasks. More specifically, document layout analysis (DLA)(Binmakhashen & Mahmoud, 2019). The tool will flag off portions of ETD manuscripts that do not conform to established guidelines. Hence, this will help resolve the inconsistencies in the format of submitted manuscripts.

Methodology:

We will employ a mixed method approach combining both qualitative and quantitative methods for data collection and analysis. Primary data is being collected through structured interviews and questionnaires, while secondary data involves document analysis of postgraduate guidelines and archived historical ETDs. Convenient sampling and purposive sampling will be used to select participants, the selection is based on judgement and their willingness to participate in the study. The study is targeted on postgraduate examiners, alumni and current postgraduate students. Document analysis will be employed to understand the postgraduate guidelines stipulated in the (The University of

Zambia, 2015) document and the identified guidelines are to be categorised into themes and a checklist will be prepared which will be used as a measurement instrument and each guideline, will be set as a question and each question will be given a measurement scale of 1-5; content analysis will be used on randomly sampled ETDs in order to experimentally determine the extent of the problem using Park's metrics of metadata quality (accuracy, consistency and completeness) and, finally, a DLA Natural Language Processing model will be implemented and will be evaluated using standard DLA metrics such as Structure Similarity Index and Intersection over Union. This DLA pipeline is hinged on Artificial intelligence and Natural language processing(NLP)(Mishra & Kumar, 2020). We also leveraged an open source package called Deepdoctetion for the implementation of the software tool. For the development process, Agile methodologies, specifically scrum, will be used for software development, with document layout techniques applied to enhance the Deepdoctetion package in the process. We also accessed the accuracy and usability of the tool through a benchmark dataset composed of historical ETDs archived on The University of Zambia institutional repository. Measurement instruments include structured interview guides and a checklist to evaluate the threshold of compliance of historical ETDs. Preliminary Findings and Discussion Empirical Analysis of Historical Archived ETDs The analysis of historical archived manuscripts against postgraduate guidelines revealed that most manuscripts did not adhere to the guidelines. The preliminary findings indicate that the majority of the students from all the schools at the institution performed poorly on sections such as the copyright declaration page which had the lowest compliance, with much lower scores in accuracy (2.227), consistency (2.194), and completeness (2.037). This suggests that this section of the manuscripts frequently falls short of meeting the required standards, highlighting a critical area for improvement. This is followed by the declaration page with the second lowest scores in accuracy (3.042), consistency (2.681) and completeness (2.835). The compliance to miscellaneous requirements also had some low scores in accuracy (3.774), consistency (3.760) and completeness (3.841). Overall middle-ranking sections include the title Page with the following compliance scores in accuracy (4.078), consistency (3.861) and completeness (4.023). Whereas compliance on the certificate of approval indicates the following scores in accuracy (4.293),consistency (4.304) and completeness (4.322). Similarly, the end matter section with the following compliance scores in accuracy (4.581), consistency (4.471) and completeness (4.551).Lastly, compliance to pagination had the following scores in accuracy (4.634), consistency (4.629) and completeness (4.623). However, while these sections generally show good compliance, the moderate degree of variation suggests that there are some inconsistencies in adherence to the guidelines, and targeted improvements could enhance their overall performance. Overall, the "Order Of Text" demonstrates exceptional compliance with nearly perfect scores in accuracy (4.969), consistency (5) and completeness (5). Similarly, compliance to the prescribed length of an ETD indicates remarkable scores in accuracy (5), consistency (5) and completeness (5). Survey responses for current postgraduate students According to the responses from current postgraduate students, the most common challenges in

ensuring correct document formatting in accordance with the universities guidelines included difficulties with citation management, generating the correct table of contents, and formatting tables and figures. Out of 22 respondents, seven (7) highlighted that they find difficulties in adjusting their manuscripts to the prescribed margins, another seven (7) highlighted difficulties when working on tables and figures. Ten respondents stated that the most challenging aspect is with regards to generating references and citations. Five respondents also highlighted that following the prescribed table of contents is challenging for them. Of the 22 respondents, 50% stated that they were not very familiar with the postgraduate guidelines of the institution while 6

Survey results indicate that postgraduate alumni faced challenges with formatting their manuscripts, particularly in adding tables, maps, figures (5 respondents), managing citations (4 respondents), and inserting footnotes (2 respondents). These difficulties align with the low scoring of miscellaneous guidelines (covering citations, figures, and tables) in historical ETD analyses.

Examiners also reported challenges in manually verifying manuscript compliance. With a high student-to-examiner ratio in some schools, the workload can be overwhelming, limiting their ability to be thorough. Due to time constraints, examiners often review up to 50 manuscripts per sitting to provide timely feedback and avoid delaying student research. Furthermore, examiners stated that they find themselves either resending the guidelines to the students or scheduling short seminars mid-research to re-educate the students about the guidelines when they notice a lot of inconsistencies in the submitted manuscripts. Conclusion The use of DLA methods to create a pipeline that will curtail the challenge of ensuring and checking compliance of ETDs to postgraduate guidelines at the university of Zambia has proven to be the most efficient solution. The development of an ETD automatic guideline verification tool presents an opportunity to enhance efficiency as well as promote consistency in the quality of ETDs while alleviating the challenges faced in the process of manually checking for compliance consequently reducing the workload for students, supervisors and examiners alike.

References:

Mishra, B. K., & Kumar, R. (2020). Natural Language Processing in Artificial Intelligence. Razı, S., Glendinning, I., & Foltýnek, T. (2019). Towards Consistency and Transparency in Academic Integrity. Peter Lang Gmbh, Internationaler Verlag Der Wissenschaften. Binmakhashen, G. M., & Mahmoud, S. A. (2019). Document Layout Analysis: A Comprehensive Survey. ACM Comput. Surv., 52(6), 1–36. The University of Zambia. (2015). Directorate of Research and Graduate Studies: Postgraduate regulations.

AUTOMATIC ELECTRONIC THESIS AND DISSERTATION GUIDELINES VERIFICATION

A Data Mining Tool For Improved Compliance Against Postgraduate Guidelines

Mubanga Chibesa | Albertinah Mooka | Gift Muwele | Lwiime Shansonga | Lighton Phiri

DEPARTMENT OF LIBRARY AND INFORMATION SCIENCE

UNIVERSITY OF ZAMBIA

INTRODUCTION

Higher Education Institutions worldwide enforce guidelines and academic approaches to ensure scholarly integrity and adherence to academic standards. The University of Zambia is not an exception. Just like most HEIs it offers training to postgraduate students and one of the key aspects of postgraduate training is producing an Electronic Thesis and Dissertation manuscript. The Directorate of Research Innovation and Development (DRID) at the University of Zambia provides guidelines [4] which stipulate how ETD's should be formatted. However, the process of checking for conformance is a manual and tedious procedure, resulting in submission of inconsistently formatted manuscripts in the Institutional Repository (IR).

To address this challenge our project seeks to implement a tool that will automate the process of checking ETD's compliance against established postgraduate guidelines. The tool will leverage data mining techniques to perform this task. More specifically, Document Layout Analysis (DLA)[1] will be the core approach used in the implementation. The tool will flag off portions of ETD manuscripts that do not conform to established guidelines. Hence, this will help resolve the inconsistencies in the format of submitted manuscripts.

RESEARCH OBJECTIVES

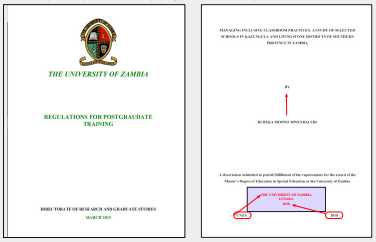
1. To Identify the University of Zambia postgraduate guidelines that ETDs manuscripts should conform to.
2. To analyse the University of Zambia archived historical ETDs compliance to postgraduate guidelines.
3. To Investigate the challenges faced when checking for ETDs compliance to postgraduate guidelines.
4. To Design and implement a software tool that will utilise document layout analysis (DLA) techniques that will automatically flag off portions of manuscripts that do not conform to the institutional guidelines.

METHODOLOGY

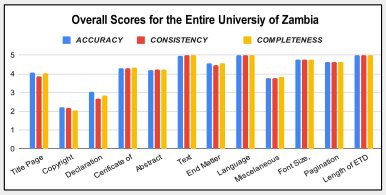
Using a mixed methods approach, document analysis will be employed to understand the postgraduate guidelines stipulated in the "Regulations and Guidelines for Postgraduate Studies" guidelines document; content analysis was used on randomly sampled ETDs in order to experimentally determine the extent of the problem and, finally, a DLA Natural Language Processing model[3] will be developed and evaluated using standard DLA metrics such as Structure Similarity Index and Intersection over Union[2]. This DLA pipeline is hinged on Artificial intelligence and Natural language processing(NLP). We also leveraged an open source package called Deepdoctetion for the implementation of the software tool.

RESULTS & DISCUSSION

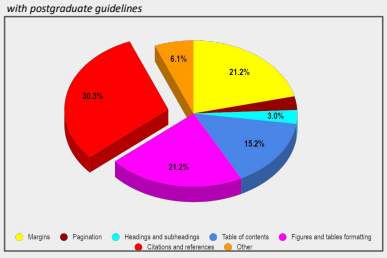
1. Below are images representing the Guidelines (Right) and Sample ETD script with parts not complying pointed out (Left).



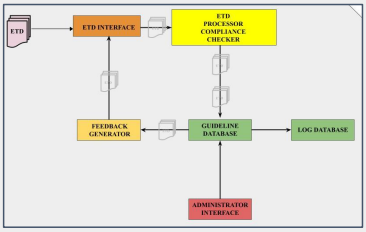
2. Below is a chart representing the average scores from the analysis of historical achieved ETD Manuscriptof the entire schools at the University of Zambia on each preliminary section of the manuscript



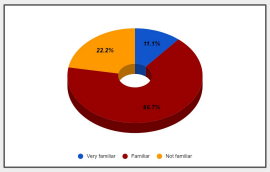
3. The chart below shows the responses gotten from Alumni Students on the challenges they faced when checking if their manuscript were in compliance with postgraduate guidelines



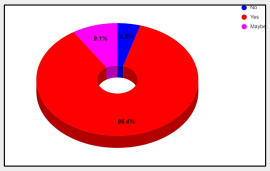
4. The diagram below illustrates how the software tool will operate once implemented.



3.1. The diagram on the left show responses gotten from Alumni Students when asked How familiar they were with the DRID poespstgraduate guidelines for ETD formatting



3.2. The diagram below show responses gotten from Current Students when asked if there is a need for an automated software tool to assist with checking if their manuscrit conform to institutional guidelines?



CONCLUSION

The development of an ETD automatic guideline verification tool presents an opportunity to enhance efficiency as well as promote consistency in the quality of ETDs while alleviating the challenges faced in the process of manually checking for compliance consequently reducing the workload for students and examiners.

BIBLIOGRAPHY

1. Binmakhashen, G. M., & Mahmoud, S. A. (2019). Document Layout Analysis: A Comprehensive Survey. *ACM Comput. Surv.*, *52*(6), 1–36.
2. Mishra, B. K., & Kumar, R. (2020). *Natural Language Processing in Artificial Intelligence*.
3. Razi, S., Glendinning, I., & Foltýnek, T. (2019). *Towards Consistency and Transparency in Academic Integrity*. Peter Lang Gmbh, Internationaler Verlag Der Wissenschaften.
4. https://graduate.unza.zm/images/files/pg_regulations.pdf

Project Team Members

- Chibesa Mubanga <2020031345@student.unza.zm>
- Mooka Albertinah <2020019299@student.unza.zm>
- Muwele Gift <2020014203@student.unza.zm>
- Shansonga Lwiime <2020002353@student.unza.zm>

Project Supervisor

- Dr.Lighton Phiri <lighton.phiri@unza.zm>

Authors Index

Delegate Name	Organisation	Country
booktabs Andy Neale	Clarivate	United Kingdom
Austin Mclean	ProQuest, a part of Clarivate	United States
Benaiah Akombwa	University of Zambia	Zambia
Bertrand Thomas	Abes	France
Birbal Boniface Musoba	Higher Education Authority	Zambia
Boniface Banda	The University of Zambia	Zambia
Brenda Van Wyk	University of Pretoria	South Africa
Cecilia Kasonde	Kwame Nkrumah University	Zambia
Charles Greenberg	Networked Digital Library of Theses and Dissertations	United States
Denny Nsokolo	Higher Education Authority	Zambia
Dokowe Tembo	ZCAS University	Zambia
Douglas Mwale	St Bonaventure University	Zambia
Dr Beatrice May Banda	Chipata College of Nursing and Midwifery	Zambia
Edward Fox	Networked Digital Library of Theses and Dissertations (NDLTD)	United States
Efraim Dumeni	Namibia University of Science and Technology	Namibia
Elijah Chileshe	The University of Zambia	Zambia
Elizabeth Namonje	Higher Education Authority	Zambia
Ernest Banda	Zynle Technologies Limited	Zambia
Fabian Kakana	University of Zambia	Zambia
Fitzwell Kabunda	Bridging Gap Solution Limited	Zambia
Francina Makondo	University of Zambia Library	Zambia
Francis Kawesha	Higher Education Authority	Zambia
Gabriela Mejias	DataCite	Germany
Gift Muwele	The University of Zambia	Zambia
Godfrey Marera	St. Bonaventure University	Zambia
Hope Chinganya	CIDRZ	Zambia
Hussein Suleman	University of Cape Town	South Africa
Johanssen Obanda	Crossref	Kenya
Joseph Muyunda Walusiku	Centre for Infectious Disease Research in Zambia	Zambia

Joseph Telemala	Sokoine University of Agriculture	Tanzania
Josiline Chigwada	University of South Africa	South Africa
Kabuku Mwandamena	Chipata College of Nursing and Midwifery	Zambia
Kabwe Kasoma	CIDRZ	Zambia
Kamani Perera	Chartered Institute of Personnel Management	Sri Lanka
Kaoma Lamba Daka	University of Zambia	Zambia
Kazhila C. Chinsembu	Higher Education Authority	Zambia
Kelvin Chibomba (Phd)	Information and Communications zuniversity	Zambia
Kudakwashe Siziva	Datacite	South Africa
Lamia Salsabil	Old Dominion University	United States
Lazarus Matizirofa Banda	University of the Witwatersrand, Johannesburg	South Africa
Lewis Cheelo	Zambia Air Force (ZAF)	Zambia
Lighton Phiri	University of Zambia	Zambia
Lombe Tembo	ORCID	Zambia
Manoj Kumar K	INFLIBNET Centre	India
Mark Phillips	University of North Texas	United States
Martin Mushumba	Higher Education Authority	Zambia
Martin Nkumbula	Zynle Technologies Limited	Zambia
Marvin Kabubi	Information and Communications University	Zambia
Matamela Michael Mulaudzi	University of the Witwatersrand, Johannesburg	South Africa
Matildah Mercy Muchinga	Lusaka Apex Medical University	Zambia
Mduduzi Aubrey Ntetha	University of South Africa	South Africa
Melusi Ncube	Bridging Gap Solutions	Zambia
Mercy Banda	St. Bonaventure University	Zambia
Mosebjadi (Mose) Petje	Times Higher Education	South Africa
Moses Makungu	Zambia Air Force Center of Advance Learning	Zambia
Moses Mupeta	Information and communications university	Zambia
Mpundu Chilonga	KWAME NKRUMAH UNIVERSITY	Zambia
Mubanga Chibesa	The University Of Zambia	Zambia
Olatunbosun Obileye	DataCite	Nigeria
Olivier Cian	ABES	France

Phyela Mbewe	University of Zambia	Zambia
Shahzeb Hasan	Akal University	India
Thabiso Mwiinga-Mwanachingwala	University of Zambia	Zambia
Vivek Ranjan	Silver Oak University	India
Wendel Fabian Chinsamy	DataCite	South Africa
Yinlin Chen	Virginia Tech	United States
Zachary Zulu	THE UNIVERSITY OF ZAMBIA	Zambia