Research Visibility in the Global South: Towards Increased Online Visibility of Scholarly Research Output in Zambia

Lighton Phiri
Department of Library and Information Science
University of Zambia
Lusaka, Zambia
lighton.phiri@unza.zm

Abstract—Scholarly research and publication forms an integral part of the core functions of Higher Education Institutions (HEIs). It is generally standard practice for HEIs to deposit scholarly output into publicly accessible Institutional Repositories (IRs). While Zambia has seen a rise in the number of HEIs, with a total of six Public HEIs and 60 Private HEIs, there is little online visibility of scholarly output generated by these HEIs. A bibliometric analysis, focused on electronic theses and dissertations (ETDs), was conducted by harvesting scholarly publications from HEIs IRs, in order to demonstrate the low online visibility of scholarly research output in Zambia. We also outline technological initiatives, by using case examples from The University of Zambia, that can be employed to potentially increase the online visibility of HEIs scholarly output. Specifically, we illustrate how subject repositories and downstream aggregate services can be utilised to increase the visibility of scholarly output. The study shows that only two HEIs have established IRs, with noticeably low scholarly publications by academic staff. In addition, there is a noticeably long delay between the publication date of the ETDs and the ingestion date into the IRs. In addition, while not comprehensive, the proposed initiatives demonstrate technological initiates that could be employed to increase the visibility of scholarly research output.

Index Terms—bibliometrics, digital libraries, OAI-PMH, repositories

I. INTRODUCTION

Higher Education Institutions (HEIs), in Zambia, play the crucial role of providing training towards the attainment of advanced degrees such as Masters and Doctoral degrees. In addition, HEIs conduct research that is aimed as solving many of society's pressing problems, with a key output being scholarly research publications.

The Higher Education Authority (HEA)[1] of Zambia—through the Higher Education Act of 2013 [1]—has been given the legal mandate to register Private HEIs and, more importantly ensure that HEI quality is not compromised. With the increasing demand of higher education, Zambia has seen a steady increase in the number of HEIs: there are a total of six Public HEIs [2] and 60 Private HEIs [3].

While there has been an increase in the number of registered HEIs and, corresponding enrolment rates of postgraduate stu-

[1]http://www.hea.org.zm

dents, the online visibility of scholarly research output is still noticeably low in Zambia. This paper outline the extent of the low online visibility of HEIs scholarly research output. In addition, the paper describes initiatives currently being undertaken to facilitate the increased online visibility of HEIs research output.

This paper contributes the following: (1) Empirical evidence showing the low online visibility scholarly research output generated by HEIs in Zambia. (2) Demonstration of initiatives that can potentially increase the visibility of scholarly research output generate by HEIs in Zambia.

The remainder of this paper is organised as follows: Section II discusses literature related to this work. Section III describes the methodology, while the results and discussion are presented in Section IV. Section V outlines initiatives that could potentially lead to increased online visibility of scholarly output and, finally, Section VI concludes the paper and outlines potential future work.

II. RELATED WORK

A. Bibliometric Analysis of HEIs in Zambia

Kalyambanino's dissertation examined faculty productivity at University of Zambia (UNZA) by analysing their research and publications. A mixed-methods approaches, involving questionnaires and interview guides with 251 participants, was used to gather data. The study suggests a low publication output, with 19.5% and 39% academic staff indicating having published books and articles in the previous two years, respectively [4].

Akakandelwa analysed publications authored by academic staff at UNZA by downloading publications papers authored between 2002 and 2007 from the Thomson Scientific database [5]. The publications were analysed in order to determine authorship patterns and collaboration. The average publication count was 36.7, with a highest publication count of 63, recorded in 2006.

Ahmed et al. conducted a mapping of postgraduate research in the School of Medicine, at UNZA, in order to explore research characteristics of the Master of Medicine programme [6]. A desk review of the Master of Medicine programme

dissertations was conducted by reviewing manuscripts that had been published between 1986 and 2009 and, deposited in the Special Collections of The UNZA Library. In contrast, this work is focused analysing the online visibility of ETDs that have been deposited on the UNZA institutional repository (IR).

B. Scholarly Research Visibility

In an attempt to explore alternative aspects for measuring the impact of The Medical Journal of Zambia[2], Kanyengo et al. reviewed online and hard-copy literature. Their online review was an online visibility assessment of the journal on platforms such as Google Scholar[3], ResearchGate[4] and academic databases such as Africa Journals Online[5]. Abrahams et al. state that the higher education sector in Southern Africa is, in part, dependant on universities' capacity to produce, communicate and use research output for educating future generations. However, they note that research output in the majority of Southern African universities is not visible [7].

Czerniewicz and Wiens conducted a study to assess the online visibility of poverty alleviation research in South Africa. Their analysis of indexed research on Google Scholar indicated relative online invisibility of research in the area [8]. In another study aimed at exploring the potential role of digital affordances in knowledge production and dissemination, Czerniewicz et al. observe that while Southern climate change researchers have a discoverable online presence, it is uneven and typically restricted to social media [9].

One of the the findings of SCAP was that Southern African research is marginal invincible in the global context [10]. Interestingly, another SCAP finding was that most universities typically have the technology required for effective scholarly communication.

C. Software for Increased Online Visibility

Scholarly publications are generally organised in collections referred to as Digital Libraries (DLs) [11]. There are a wide variety of open source DL software tools and services that are used for storing and, making available scholarly research output. HEIs generally use such DL tools for implementing IRs [12]–[14] and, increasingly, electronic journals [15].

While the DL tools have varying implementations [16], they offer generic services for facilitating core DL features like searching and browsing. More importantly, their implementations are standards based, integrating protocols for effecting ingestion and discovering of content. For instance, protocols such as the OAI-PMH are effective at facilitating the harvesting of metadata from external repositories.

III. METHODOLOGY

This section outlines the in-depth analysis conducted to explore the online visibility of scholarly publications for HEIs in Zambia. As a first step, the domains for the six Public HEIs

were crawled to determine if their scholarly output is visible only and, specifically, to determine if they have established IRs. Digital objects from HEIs with IRs were then harvested and analysed.

A. Harvesting Digital Objects

Open source Digital Library Systems (DLSes) that are used for setting up IRs are standards-based and implement interoperability protocols for effective storage and retrieval of digital objects. Digital Objects are generally composed of bitstreams—the digital resource consumed by end-users—and metadata—textual description of digital objects that provide for context about the digital resource.

The Open Archives Initiatives Protocol for Metadata Harvesting (OAI-PMH) [17] was used to harvest metadata from HEIs IRs, using the LibreCat Catmandu data processing toolkit [18]. The harvesting was done using the Dublin Core [19] metadata format—`metadataFormat=oai_dc`. In addition to the SetSpec field of the harvested metadata, the `Identifier`, `Date` and `Type` Dublin Core elements were used during the analysis stage, as outlined in Section III-B. Resources associated with each digital object were harvested using the Open Archives Initiative Object Reuse Exchange (OAI-ORE) standard [20]—`metadataFormat=ore`.

B. Processing Harvested Digital Objects

The harvested digital objects were analysed in order to classify the different types of digital objects and, additionally, determine when the digital objects were published and ingested into the IRs.

1) Metadata Processing: Metadata elements were processed in order to determine hierarchies the digital objects were associated with, the designated classification of the digital objects and, publication and ingestion dates associated with the digital objects. Specifically, the following metadata elements were analysed.

- `SetSpec`—Indicates the various hierarchical structures within which digital objects are nested.
- `Subject`—Indicates the research topics associated with the digital objects.
- `Creator`—Indicates the authors of the digital resource associated with the digital object.
- `Contributor`—Indicates the entities that contributed towards the creations of the resource associated to the digital object.
- `Description`—Indicates additional contextual overview of the digital resource.
- `Date`—Indicates the publication and ingestion dates.
- `Type`—Indicates whether the document is an ETD, pre-print or any other specified resource types.

2) Bitstream Processing : The digital object resources—PDF documents—were processed in order to determine if an ETD was Masters dissertation or Doctoral thesis. While analysing the `SetSpec` and `Type` metadata elements helped with the initial classification digital resources, parsing and processing the actual digital resource bitstream provided more

comprehensive details. In addition, the processing acts as a mechanism for validating the data contained in the metadata. Furthermore, processing the digital resource helped determine the format of the digital resource—whether it was born digital or digitised; the latter adversely affects full-text searching.

As earlier stated, OAI-ORE was used to harvest digital resources (`PDF` documents). For each `PDF` document, the `pdftk`[6] utility was used to extract the first page of the document—the cover page—and, thereafter, the `pdftotext`[7] utility was used convert the `PDF` page to plain text. Finally, the resulting text document was analysed for useful information such as the ETD classification: Masters or Doctoral; additionally, the result plain text document was used to determine if the `PDF` was born digital or digitised.

IV. RESULTS AND DISCUSSION

This section presents results from the analysis of digital objects harvested from HEIs IRs. Table I indicates that out of the six Public HEIs, only Copperbelt University (CBU) and UNZA have IRs. The results also suggest that the 168 digital objects in the CBU IR are exclusively ETDs. One obvious observation is the low count of digital objects, especially that both CBU and UNZA graduate relatively large number of Masters and Doctoral students. This is especially the case for CBU which only has 168 digital objects in its IR.

An important point worth noting is that the UNZA IR also consists of final year students' capstone project reports, scanned copies of past examinations and digital objects from external research institutes. For simplicity and consistency, subsequent analyses presented in this paper are restricted to pre-prints and ETDs, published between 2010 and 2017.

TABLE I
SCHOLARLY PUBLICATIONS AUDIT FOR PUBLIC HEI IRS

Institution	Repository	Output	Items
University of Zambia	DSpace@ UNZA†	ETD	3070
		Pre-print	253
		Capstone	1110
		Exams	356
		External	74
Copperbelt University	DSpace@ CBU‡	ETD	168
Chalimbana University	—	—	0
Kwame Nkhruma University	—	—	0
Mulungushi University	—	—	0
Mukuba University	—	—	0

†http://dspace.unza.zm:8080/xmlui
‡http://dspace.cbu.ac.zm:8080/jspui

A. Analysis 1. Digital Object Ingestion

In order to understand and better explain the low publication count, the dates the digital objects were published and their corresponding ingestion dates were analysed.

Figure 1 shows the HEI publications by year for the CBU and UNZA. The CBU IR only has digital objects published between 2011 and 2014, suggesting that nothing has been

ingested into the IR since 2014. While the UNZA IR seems to be regularly updated with publications, there are obvious inconsistencies in the rate of ingestion. The pattern suggests that digital objects are ingested in batches as opposed to when when are published.
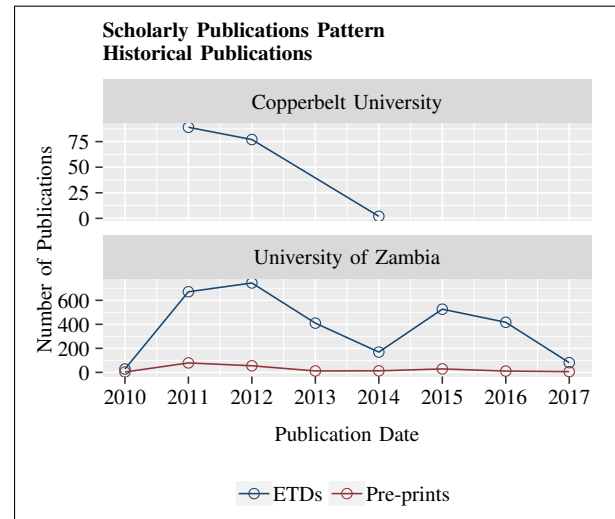

Fig. 1. Scholarly Publications by Year

The batch ingestion assumption for the UNZA IR is supported by Figure 2, which shows that the vast majority of publications were ingested between 2015 and 2016.
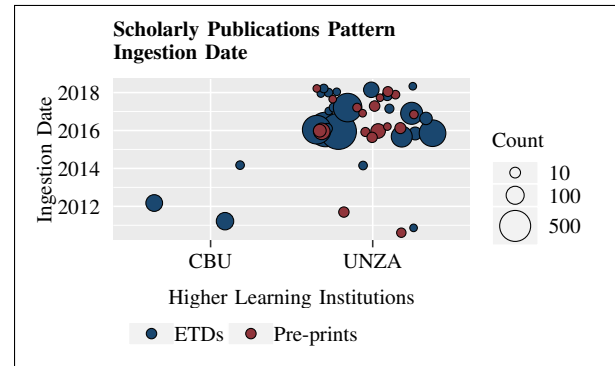

Fig. 2. Scholarly Publication Distribution by Ingestion Date

A further look at the publication distribution in Figure 2 is illustrated by Figure 3. The bubble plot indicates that most of the digital objects published are only ingested into the IR more a year after they are published, clearly affecting the online visibility of the resource. The long period has implications on not only the citation count of the digital object, but, more importantly, on other researchers potentially building up on related work—if content is not visible online, it becomes difficult for other researchers to realise this. In the case of ETDs, an argument could be made that this could ultimately result in the duplication of research conducted in various HEIs in Zambia.
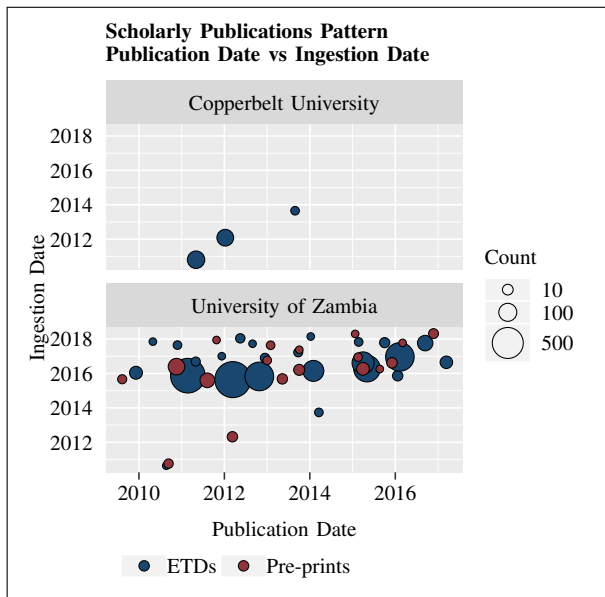
Fig. 3. Scholarly Publication Distribution by Ingestion Date

B. Analysis 2. Quality of Metadata

Listing 1. Descriptive Metadata for a Sample ETD From The UNZA IR

```
1  <oai_dc:dc>
2   <dc:title>
3     Evaluation of [...] networks(ZAMREN)
4   </dc:title>
5   <dc:creator>Mwiinga, Jervas</dc:creator>
6   <dc:subject>
7     High performance computing—Zambia
8   </dc:subject>
9   <dc:subject>
10    Research education networks—Zambia
11  </dc:subject>
12  <dc:description>
13    THESIS M.ENG
14  </dc:description>
15  <dc:description>
16    [...]
17  </dc:description>
18  <dc:date>2018−07−23T13:00:50Z</dc:date>
19  <dc:date>2018−07−23T13:00:50Z</dc:date>
20  <dc:date>2017</dc:date>
21  <dc:type>Thesis</dc:type>
22  <dc:identifier>
23    oai:dspace.unza.zm:123456789/5275
24  </dc:identifier>
25  <dc:language>en</dc:language>
26  <dc:format>application/pdf</dc:format>
27  <dc:publisher>
28    The University of Zambia
29  </dc:publisher>
30  </oai_dc:dc>
```

Digital object metadata provides descriptive information about the digital object resource. While metadata comes in different variations—administrative metadata, structural metadata and descriptive metadata—descriptive metadata plays the crucial role in facilitating the effective browsing and searching of digital objects. DLSes, in part, index metadata elements to facilitate the discovery of digital objects. The specific issues observed are explained below, by making reference to Listing 1.

1) Controlled Vocabularies: Metadata elements such as `Subject`, `Creator` and `Type` are vital for facilitating effective browsing and as such, require the use of controlled vocabularies. However, analysing the metadata harvested from the UNZA IR suggests otherwise. For instance Listing 1 clearly indicates that the two `Subject` elements do not make use of a controlled vocabulary. Incidentally, most popular DLSes like DSpace provide browsing features based on `Subject`, `Date` and `Author`, illustrating the importance of controlling the vocabulary used to populate these fields.

2) Missing Metadata Elements: One key observation made during the analysis of metadata was that crucial metadata elements was missing. Most ETDs did not have the supervisor/advisor field included—this is generally included using the `dc.contributor.supervisor` or `dc.contributor.advisor` qualifiers. This is especially important because downstream services such as OATD[8] harvest ETDs from IRs and, in certain instances, crosswalk them to different metadata schemes like ETD-ms[9].

C. Analysis 3. Bitstreams

Analysing the `PDF` documents yielded some interesting results. Only 38.66% of the digital objects were classified into their respective degree. Approximately 5.60% of the digital resources are suspected to be digitised since the resulting text file for the cover page had no content.

Interestingly enough, there were inconsistencies in the textual content on the cover pages of the ETDs. Further analyses would have to be conducted to determine if the inconsistencies are associated with publication dates for the ETDs.

V. TOOLING FOR ONLINE VISIBILITY

This section describes some initiatives that the author is involved with, which are aimed at increasing the online visibility of scholarly output at The UNZA.

A. Electronic Journals

The UNZA presently publishes seven official journals [21] and, additionally, three journals run by the Directorate of Research and Graduate Studies, aimed at publishing postgraduate research output [22]. In addition to these official journals, there are departmental journals that are subject specific. While there are a few journals such as the Journal of Preventive and Rehabilitative Medicine[10] and the Journal of Library and

---

[8]https://oatd.org
[9]http://www.ndltd.org/standards/metadata
[10]http://medicine.unza.zm/research/journal

Information Science[11] that have transitioned into electronic format, the vast majority of journals are still print-based.

The author is involved in institutional initiatives that aim to migrate print-based journals to electronic platforms, using Open Journal Systems journal management system. Migrating the print-based journals into electronic format is certain to increase the visibility of scholarly publications.

B. Subject Repositories

One of the reasons why there is a large time gap between the publication dates and ingestion dates of digital objects analysed in Section IV-A is possibly because the submission workflow is solely handled by the UNZA Library. A potential solution is to decentralise the process, using subject repositories, enabling authors to electronically submit their work. This could potentially ensure that the correct and appropriate metadata elements are submitted before the digital objects are actioned into the IR.

The OAI-PMH protocol can be used to integrate the subject repositories with the IR, as shown in Figure 4. The subject repositories could be school-specific or department-specific. Fundamentally, the subject repository acts as a data provider, enabling the IR to easily harvest metadata and bitstreams. Ongoing work is being conducted to assess the feasibility and effectiveness of this approach.
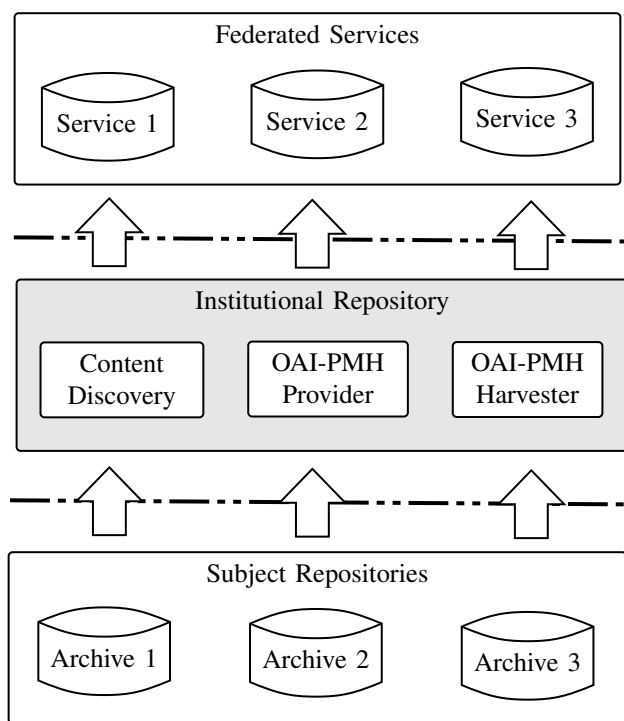

Fig. 4. Decentralised Architecture for Increased Visibility of Publications.

C. ETD Harvester

Harvester services typically take advantage of OAI-PMH protocol to collect and aggregate digital objects into a central portal, enabling end-users perform centralised searching and browsing of content. Popular portals include national initiatives such as the South African National Electronic and Dissertation portal[12] and the global Networked Digital Library of Theses and Dissertations Union Catalog[13]. Similarly, a Zambian National ETD portal[14] has been set up to aggregate ETDs from the various HEIs in Zambia. Figure 5 shows a screenshot of the portal with ETD metadata harvested from CBU and UNZA IRs.

D. Summary

This section has outlined some practical and actionable technology-centric approaches that relevant stakeholders can be undertaken to work towards increasing the visibility of scholarly output. In particular, the case examples discussed illustrate the feasibility of these technological initiatives.

VI. CONCLUSION AND FUTURE WORK

This paper illustrates the extent of the low online visibility of scholarly research output in Zambia. HEIs IRs were empirically analysed by extracting digital object resources and corresponding metadata. Due to the noticeably low numbers of pre-prints in the IRs, emphasis was placed on ETDs. The findings highlight the low visibility of research and, additional factors that might ultimately affect visibility of research. The paper also describes technological initiatives that could potentially lead to increased visibility of scholarly output. While technology is a major contributing factors for increased visibility of research, working towards increasing the online visibility of research requires a multi-faceted approach that should also involve changes in institutional culture and research communication practices [10].

Ongoing work the author is involved with includes understanding barriers associated with electronic publishing and, the potential effectiveness and feasibility of using subject repositories. As part of future work, machine learning and crowdsourcing could be potentially employed to automatically verify, validate and re-classify digital objects that are not properly tagged.

REFERENCES

[1] National Assembly of Zambia, "The Higher Education 2013," 2013. [Online]. Available: http://www.parliament.gov.zm/node/3097

[2] Higher Education Authority, "Public HEIs." [Online]. Available: http://www.hea.org.zm/index.php/public-heis

[3] Higher Education Authority, "Registered Private HEIs." [Online]. Available: http://www.hea.org.zm/index.php/registered-private-heis2

[4] C. Kulyambanino, "Faculty Productivity at The University of Zambia: Exploring Research and Publication," Ph.D. dissertation, University of Zambia, 2016. [Online]. Available: http://dspace.unza.zm:8080/xmlui/handle/123456789/4899

[5] A. Akandelwa, "Author Collaboration and Productivity at the University of Zambia, 2002-2007," *African Journal of Library Archives and Information Science*, vol. 19, pp. 13–23, 2009.

[6] Y. Ahmed, C. Kanyengo, and A. Akandelwa, "Mapping Postgraduate Research at the University of Zambia: a review of dissertations for the Master of Medicine Programme." *Medical journal of Zambia*, vol. 37, no. 2, pp. 52–57, 2010.

---

[11]https://zajlis.unza.zm

[12]http://www.netd.ac.za
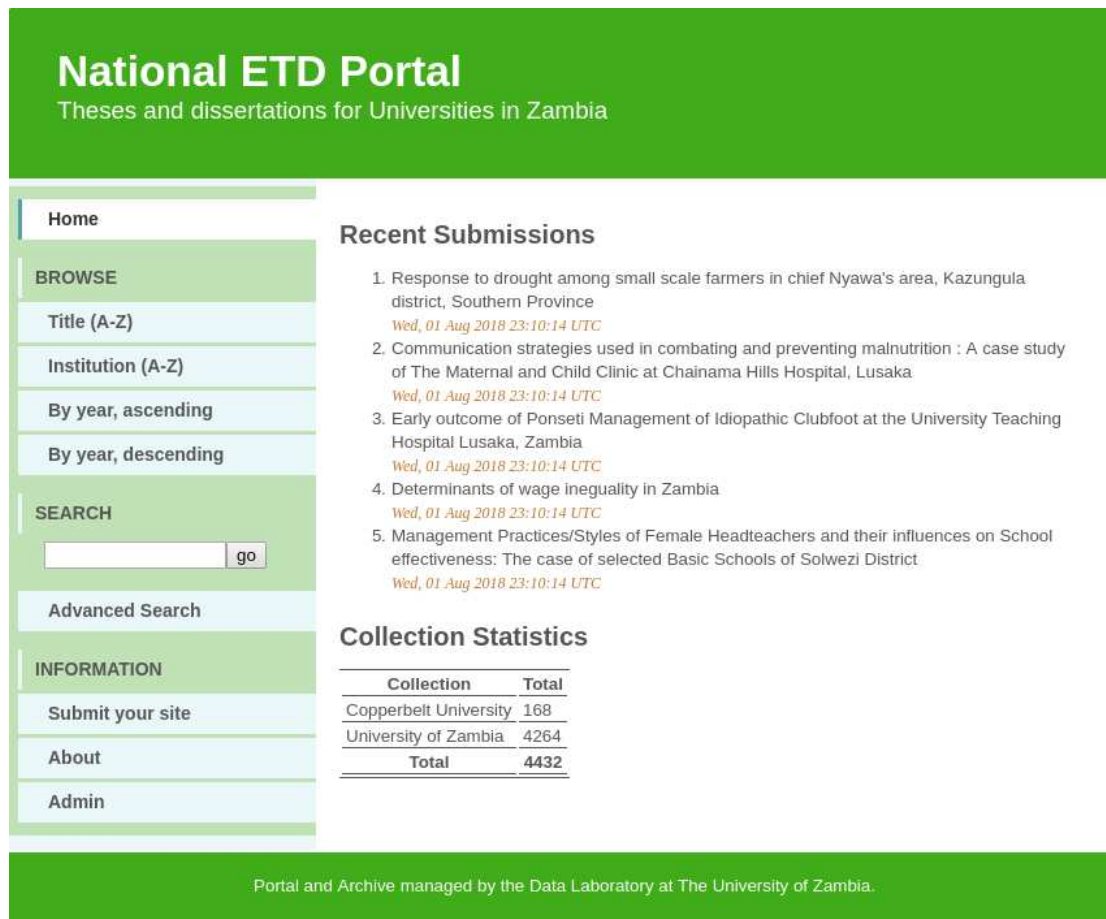[13]http://union.ndltd.org
[14]http://lis.unza.zm/portal

Fig. 5. OAI-PMH Downstream Service for Collecting and Disseminating Electronic Theses and Dissertations in Zambia.

[7] L. Abrahams, M. Burke, and J. Mouton, "Research Productivity-Visibility-Accessibility and Scholarly Communication in Southern African Universities," *The African Journal of Information and Communication*, no. 10, pp. 20–36, 2010.

[8] L. Czerniewicz and K. Wiens, "The online visibility of South African knowledge: Searching for poverty alleviation," *South African Journal of Information and Communication*, no. 13, 2013. [Online]. Available: http://hdl.handle.net/10539/19274

[9] L. Czerniewicz, S. Goodier, and R. Morrell, "Southern knowledge online? Climate change research discoverability and communication practices," *Information, Communication & Society*, vol. 20, no. 3, pp. 386–405, mar 2017. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/1369118X.2016.1168473

[10] H. Trotter, C. Kell, M. Willmers, E. Gray, and T. King, *Seeking Impact and Visibility: Scholarly Communication in Southern Africa*. African Minds, 2014. [Online]. Available: https://open.uct.ac.za/handle/11427/2310

[11] W. Y. Arms, *Digital Libraries*, 2nd ed. Cambridge, Massachusetts: The MIT Press, 2000. [Online]. Available: http://www.cs.cornell.edu/wya/DigLib/

[12] C. Gutteridge, "GNU EPrints 2 Overview," in *11th Panhellenic Academic Libraries Conference*, 2002. [Online]. Available: http://eprints.ecs.soton.ac.uk/6840/

[13] R. Tansley, M. Bass, D. Stuve, M. Branschofsky, D. Chudnov, G. McClellan, and M. Smith, "The DSpace institutional digital repository system: current functionality," in *2003 Joint Conference on Digital Libraries, 2003. Proceedings.* IEEE Comput. Soc, 2003, pp. 87–97. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1204846

[14] C. Lagoze, S. Payette, E. Shin, and C. Wilper, "Fedora: an architecture for complex objects and their relationships," *International Journal on Digital Libraries*, vol. 6, no. 2, pp. 124–138, dec 2005. [Online]. Available: http://www.springerlink.com/index/10.1007/s00799-005-0130-3

[15] J. Willinsky, "Open Journal Systems," *Library Hi Tech*, vol. 23, no. 4, pp. 504–519, dec 2005. [Online]. Available: https://www.emeraldinsight.com/doi/10.1108/07378830510636300

[16] M. Kökörčen\'y and A. Bodnárová, "Comparison of digital libraries systems," in *Proceedings of the 9th WSEAS international conference on Data networks, communications, computers.* World Scientific and Engineering Academy and Society (WSEAS), 2010, pp. 97–100. [Online]. Available: http://www.wseas.us/e-library/conferences/2010/Faro/DNCOCO/DNCOCO-16.pdf

[17] C. Lagoze, H. Van de Sompel, M. Nelson, and S. Warner, "Open Archives Initiative-Protocol for Metadata Harvesting-v. 2.0," 2002. [Online]. Available: http://www.openarchives.org/OAI/openarchivesprotocol.htmlhttp://www.nii.ac.jp/irp/archive/translation/oai-pmh2.0/OpenArchivesProtocol.htm

[18] LibreCat, "LibreCat/Catmandu data processing toolkit." [Online]. Available: http://librecat.org

[19] S. Weibel, J. Kunze, C. Lagoze, and M. Wolf, "Dublin Core Metadata for Resource Discovery," 1998. [Online]. Available: http://www.hjp.at/doc/rfc/rfc2413.html

[20] C. Lagoze, H. Van de Sompel, M. L. Nelson, S. Warner, R. Sanderson, and P. Johnston, "Object Re-Use & Exchange: A Resource-Centric Approach," *0804.2273*, apr 2008. [Online]. Available: http://arxiv.org/abs/0804.2273

[21] The University of Zambia Press, "UNZA Press." [Online]. Available: https://www.unza.zm/units/press

[22] Directorate of Research and Graduate Studies, "Dissemination of Research Findings." [Online]. Available: http://graduate.unza.zm/research/dissemination-of-research-findings