IST-Africa

IST-Africa 2021 Conference Proceedings
Miriam Cunningham and Paul Cunningham (Eds)
IST-Africa Institute and IIMC, 2021
ISBN: 978-1-905824-66-3

Effective Ingestion of Digital Objects in Institutional Repositories Using Subject Repositories

Moonga HABUKALI, Mathews MBEWE, Nozyenzi MWALE, Mathews MWEWA, Nandi SIKAZINDU, Lighton PHIRI
University of Zambia, P.O. Box 32379, Lusaka, Zambia
Tel: +260 211 295220, Fax: +260 211 253952,
Email: {14089394,14127644,14008343,13048911,14100967}@student.unza.zm, lighton.phiri@unza.zm

Abstract: The online visibility of scholarly research output from the Global South is cited as being low. This is despite the increased number of Higher Education Institutions (HEIs) setting up institutional Repositories (IR), to facilitate the collection, long-term preservation and effective access to scholarly research output. One of the most common challenges faced by HEIs is ensuring regular ingestion of digital objects in IRs. In order to address this challenge, we propose the use of subject IRs by HEIs in order to facilitate the increased online visibility of scholarly research output. This paper outlines a case study conducted at The University of Zambia (UNZA), in order to demonstrate the feasibility of using subject IR at relatively large HEIs. A situational analysis was conducted to determine the relative effectiveness of ingesting digital objects at UNZA, and additionally, a subject repository was implemented in order to demonstrate the feasibility of implementing subject repositories. The results suggest that it is feasible to implement usable subject repositories, presenting opportunities to decentralise the ingestion of digital objects at large HEIs. This has the obvious benefit of facilitating effective self-archiving of digital objects and the subsequent increased online visibility of scholarly research output.

Keywords: Digital Libraries, Institutional Repositories, Scholarly Research

1. Introduction

The visibility of research output in the Global South, and in particular the African continent, is cited as being low compared to the rest of the world [1, 2, 3, 4]. This is despite the increase in the number of Higher Education Institutions (HEIs) being established. HEIs typically use Institutional Repositories (IRs) to make their research output available and a number of HEIs, including those in the Global South, have established fully functional IRs [5]. However, despite the existence of IRs in most HEIs, very little research output is made available on the platforms. This is the case for The University of Zambia (UNZA), which has grappled with the problem of ensuring that research output produced by faculty staff and postgraduate students is made available in its IR.

Prior to ingestion into IRs, the metadata associated with the digital object needs to be prepared, making the ingestion process time consuming and error prone. Timely and accurate ingestion of digital objects is in fact cited as one of the major challenges with IRs [1]. These challenges are further amplified by the complex workflows associated with the ingestion process, coupled with challenges associated with self-archiving [6, 7].

In order to address these challenges, a viable solution could potentially involve decentralisation of the ingestion process, through the use of subject repositories. This paper

discusses a case study conducted to outline the potential positive effect of using subject repositories during the ingestion of digital objects in IRs.

The main objective of the study outlined in this paper was to investigate the feasibility of using subject IRs. The main contributions of this paper are as follows:

- A strategy for comprehensively evaluating IR ingestion practices, using a content analysis approach of IR digital objects;
- Results from empirical evaluation of usability studies conducted to determine the outcome of using subject repositories;
- A demonstration of the feasibility and efficacy of using subject repository for the increased online visibility of scholarly research output.

The remainder of this section is organised as follows: Section 2 is a synthesis of related work linked to this study; Section 3 describes the methodology followed to conduct this study; Section 4 describes and discusses the study results; and, finally, Section 5 comprises concluding remarks.

2. Related Work

2.1 Institutional Repositories

Institutional Repositories (IRs) are information management systems used by Higher Education Institutions (HEIs) to showcase their scholarly research output. IRs serve the function of collecting, managing, archiving and making available digital content corresponding to scholarly research output. Scholarly research output in IRs are represented as digital objects, an entity that is composed of metadata and bitstreams. The bitstream is the content that is consumed by the end users - for instance PDF documents - while the metadata is representational information that provides additional contextual information about the digital objects [8, 9].

IRs are considered a specialised form of Digital Libraries (DLs) and thus exhibit the same features and functionalities. Fundamentally, DLs will be implemented such that they facilitate the storage of digital objects through a storage layer, the management of digital objects through a service layer and access to digital objects through a user interface layer [10]. Subject repositories, unlike IRs, are not used to store generic output produced by HEIs, but rather content specific to a specialised domain. Platforms such as ArXiv [11] and RePEc [12] provide examples of large-scale subject repositories for archiving science-based and economics-based research output, respectively. In addition to such large-scale repositories, HEIs can decentralise the processing of digital objects through the use of subject IRs. In this work, subject IRs are proposed as a potentially viable solution to guaranteeing effective ingestion of digital objects.

2.2 Repository Software Tools

There are a number of free and open source repository software tools that are available to be used to implement IRs. While the fundamental concepts of DL are incorporated into the design and implementation of such tools, the tools are generally tailored for specific domains [13]. For instance, Omeka [14] is generally used for building digital collections that are primarily used to showcase photographs, and DSpace [15] and EPrint [16] are typically used to build document archives.

In this work, DSpace was used to set up the subject repository due to its popularity and, more importantly, because the main IR at The UNZA is implemented using the same software tool, facilitating the seamless integration between the subject repository and the main IR.

2.3 Interoperability Protocols in Repositories

Interoperability in DLs enables them to work with other external services, a crucial ingredient for openness [17]. In DLs, interoperability makes it possible for aggregate services such as the Networked Digital Library of Theses and Dissertations (NDLTD) [18] and Open Access Theses and Dissertations (OATD) [19] to be implemented.

There are a number of protocols that are implemented within IRs. For instance the Sword protocol is used to facilitate remote deposit of digital content [20], Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) enables harvesting of digital object metadata [21] and the Open Archives Initiative Object Reuse and Exchange (OAI-ORE) for harvesting digital object bitstreams [22]. In this work, the OAI-PMH and OAI-ORE protocols are used to make it possible for The UNZA IR to harvest metadata and bitstreams from the subject IR. Furthermore, it is possible for the subject IR to harvest metadata and bitstreams from the main IR.

3. Methodology

This study used a case study research design, with The UNZA used as the study setting. Digital objects were harvested from The UNZA IR, as outlined in Section 3.1. Semi-structured interviews were conducted with staff responsible for ingested digital objects in the IR at The UNZA, as described in Section 3.2. A subject repository was implemented for the Department of Library and Information Science at The UNZA, as outlined in Section 3.3 and, a usability study was conducted in order to evaluate the usability of the subject repository, as explained in Section 3.4.

3.1 Ingested Digital Objects

Digital object metadata were harvested from The UNZA IR using the OAI-PMH protocol [21]. The OAI-PMH ListRecords verb was used to extract Dublin Core encoded metadata, while the ListSets verb was used to identify hierarchical structures associated with the digital object metadata. The harvested metadata was then analysed by focusing on the IR hierarchical structures, in order to determine the communities and collections associated with the digital objects. In addition the analysis focused on the ingestion date in order to determine the recency of digital objects.

The UNZA IR is presently used to archive a diverse range of digital objects types and as such, a further more detailed analysis was conducted in order to determine the representation of the various digital object types.

3.2 Ingestion Workflow

In order to understand the current workflows associated with ingestion of digital objects into the repository, two (2) members of staff responsible for managing the IR at The UNZA were interviewed. The interview sessions were recorded and eventually transcribed, in order to extract relevant information. In essence, the interview sessions were aimed at understanding the procedures followed during ingestion of digital objects, the type of materials ingested into the repository and policies followed during the ingestion process.

3.3 Subject Repository

A subject repository, shown in Figure 1, was set up and configured for the Department of Library and Information Science at The UNZA. DSpace, a free and open source Digital Library software platform [23], was used to implement the subject repository.

In order to facilitate the synchronisation of digital content between the subject repository and the main UNZA IR, the subject repository was configured with a data provider and a harvester. Figure 2 shows how metadata and bitstreams are synchronised by using the OAI-PMH [21] and OAI-ORE protocols [22]. The data provider allows external

third-party services to harvest metadata and bitstreams using the OAI-PMH and OAI-ORE protocols, while the harvester enables the repository to harvest digital content from external repositories.

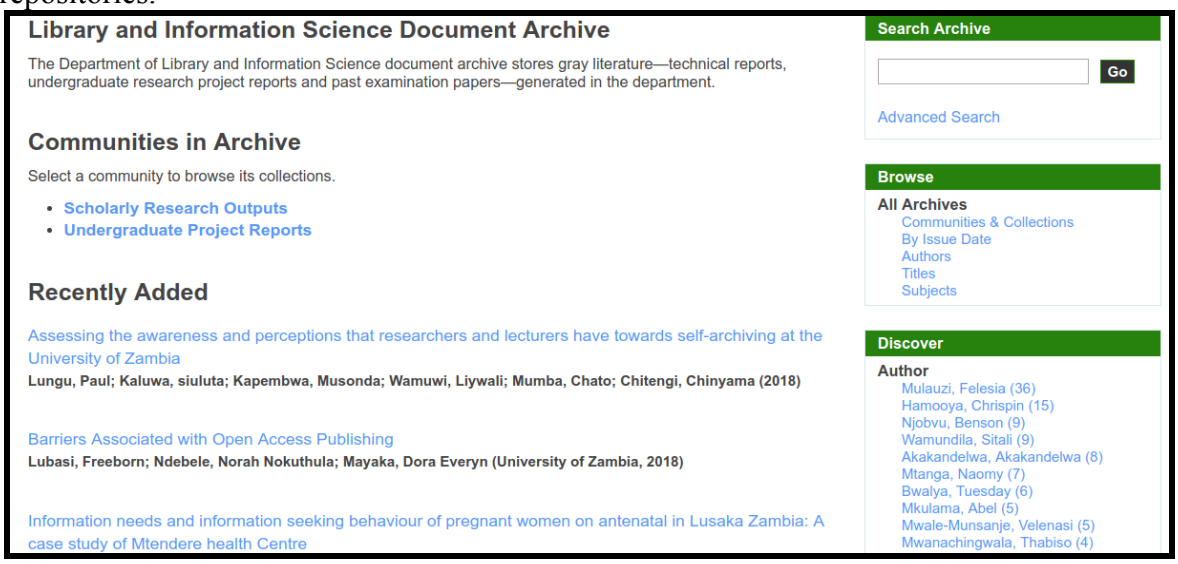

Figure 1: Department of Library and Information Science Subject Repository at The University of Zambia

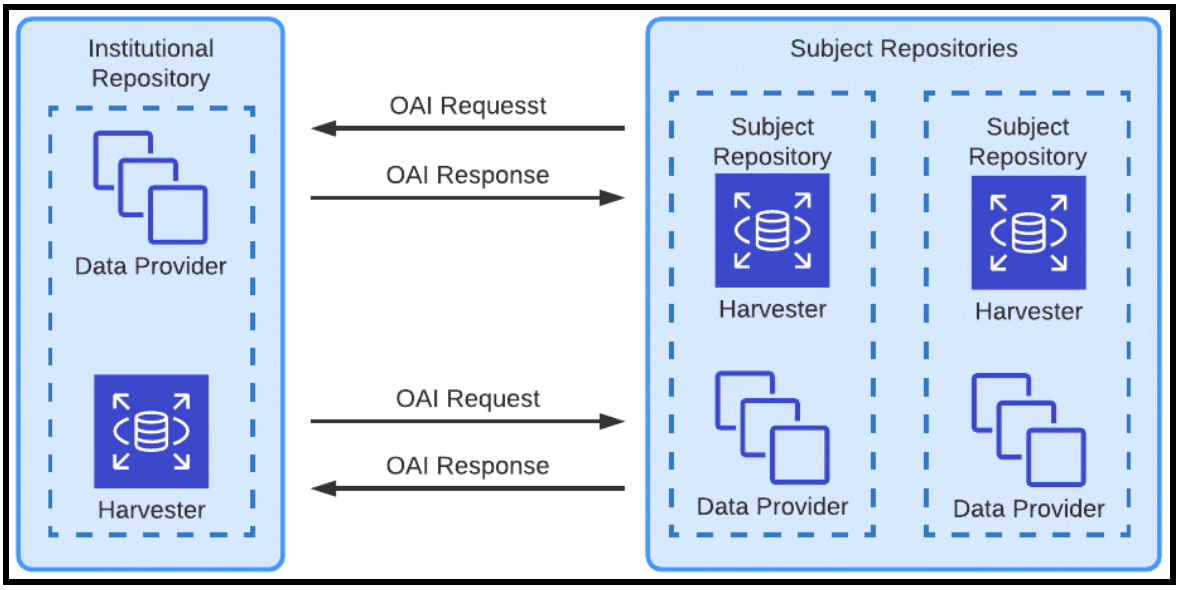

Figure 2: OAI-PMH Data Provider and Harvester Configuration for Subject IRs and Main IR

3.4 Subject Repository Usability

The System Usability Scale (SUS) [24] was used to evaluate the usability of the subject repository, in order to assess the perceived subjective views of the subject repository by individuals that had no knowledge of repositories. Purposive sampling was used to recruit participants from fourth year students in the School of Education at The UNZA. The study participants ingested a single digital object into the subject repository and subsequently filled out the SUS questionnaire.

4. Results and Discussion

4.1 Repository Content Analysis

A total of 5636 digital object metadata was harvested from The UNZA IR. The distribution of the metadata, by community, is shown in Figure 3. The harvested metadata was filtered to include scholarly research output generated by faculty staff—pre-prints and post-prints—

and postgraduate students—Electronic Theses and Dissertations (ETDs).

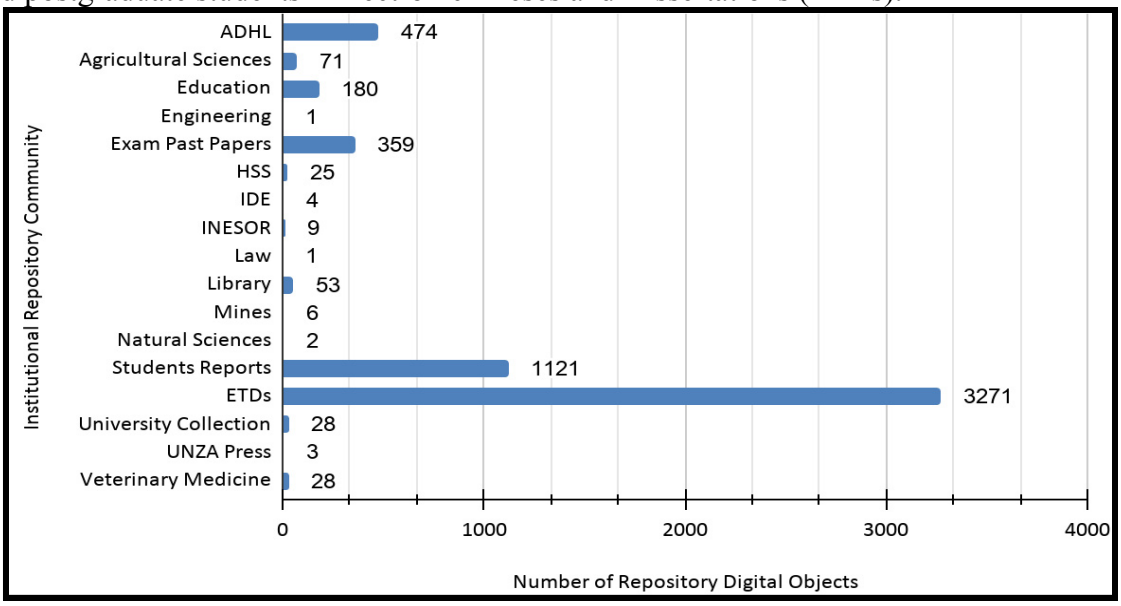

Figure 3: Distribution of the 5636 Digital Objects in the 17 Communities in The UNZA IR

There were a total of 380 faculty-generated scholarly research output and 3271 ETDs. While there is an uneven distribution of faculty-generated scholarly research output by discipline, this is not the case for ETDs as there is a policy in place that mandates that ETDs be ingested into the IR. The even distribution of ETDs is shown in Figure 4.

Prior studies conducted [2] have attributed the centralised ingestion of digital objects into the IR as being one of the main reasons for the low number of digital objects in the repository and, this is mostly as a result of fewer members of staff dedicated to ingesting digital objects into the IR. A decentralised approach that, in part, leverages subject repositories could potentially address this problem as the individual disciplines would be able to effectively ingest content into the IR using focused self-archiving techniques. Subject repositories could also potentially make it easier to integrate discipline-specific controlled vocabularies into the ingestion workflow.

The argument for decentralising ingestion of content is further supported by the significantly low number of digital objects ingested per year. The results indicate that ingestion of content is irregular, with a further analysis indicating that significant time gaps exist between publication and subsequent ingestion of digital objects.

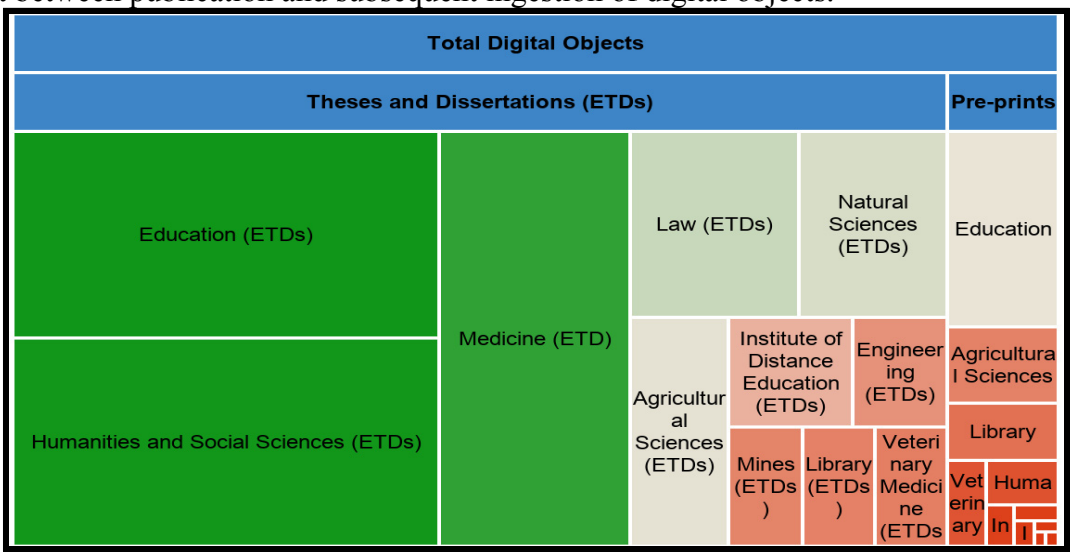

Figure 4: Proportion of Scholarly Research Output in The University of Zambia Institutional Repository

4.2 Current Ingestion Workflow

The ingestion workflow procedure involves a series of steps that include: scanning/digitising, cataloguing and, finally, review and approval of the metadata.

- Scanning and digitisation—when the document is judged fit for publishing, the library requests both a hard and electronic copy of the material so as to be archived on the IR. If the material has no electronic copy, the material is digitised
- Cataloguing—the descriptive metadata associated with the digital objects are prepared
- Review and approval of metadata—the prepared metadata is verified for consistency and correctness

The major challenges with the current ingestion workflow is that it is time consuming, making it difficult for the few members of staff to effectively ingest digital objects in the IR in a timely manner. With a decentralised approach, additional staff from the various disciplines could be incorporated into the workflow, thus facilitating effective ingestion of digital objects.

4.3 Subject Repository Usability

A total of 40 individuals participated in the usability study, with 36 of them performing the experiment tasks and completing the questionnaire. 97% of the participants had more than two years of experience using computing devices, making them eligible to participate in the study.

The computed mean SUS score was 65.5, which corresponds to a "Good" adjective rating, using the scale proposed by Bangor [25]. Figure 6 shows the distribution of the SUS scores, using predefined bands [24]. It is evident that the vast majority of the participants—83%—rated the subject repository as being "Ok", "Good", "Excellent" or "Best Imaginable". The results suggest that subject repositories have the potential to facilitate effective ingestion of digital objects, due to their perceived usability. More importantly, potential users of subject repositories would only need basic computing experience for them to effectively use the platforms.

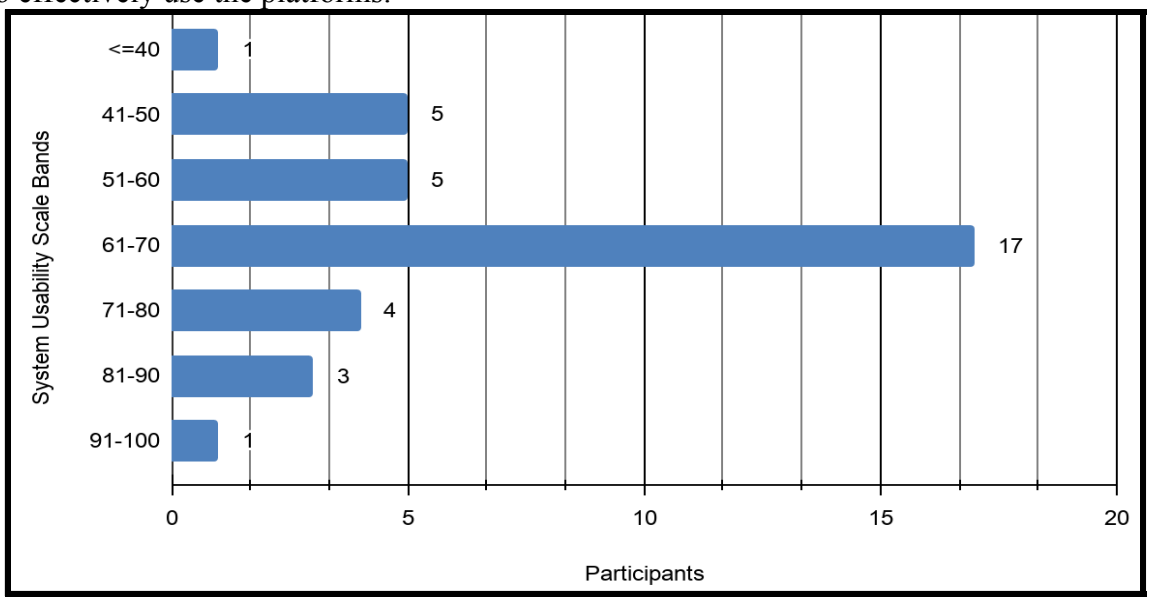

Figure 6: Frequency Distribution for the SUS Scores from the Usability Study

5. Conclusions

This paper outlined a case study aimed at assessing the feasibility of using subject IRs for facilitating effective ingestion of digital content. Specifically, the paper presented a potential solution for addressing challenges associated with timely ingestion and recency of

the ingested digital objects. The empirical analysis of digital objects ingested into The UNZA IR demonstrates the extent of these challenges. The implementation of the subject IR at The UNZA demonstrates the feasibility of implementing usable subject IRs.

The obvious benefits of subject IRs is that self-archiving is decentralised, making it relatively easier for faculty and students to upload research output. With such a system, HEIs are better placed to ensure that research output is made available in IRs in a timely manner. In addition, large HEIs generally comprise of different units, each with unique ingestion requirements; for instance, controlled vocabularies are generally different. As such, the use of subject IRs makes it possible to take into account unique requirements for different units.

As IRs scale to store the diverse types of scholarly research output produced by HEIs, it is vital that techniques for effective ingestion of digital objects are exploited and while there are numerous solutions for addressing this deluge in scholarly research output, the use of subject repositories presents opportunities for exploiting self-archiving. In essence, an ideal solution would incorporate a number of techniques that could collectively address the problem.

References

[1] L. Phiri, "Research Visibility in the Global South : Towards Increased Online Visibility of Scholarly Research Output in Zambia," In Proceedings of the 2nd IEEE International Conference in Information and Communication Technologies, 2018.

[2] H. Trotter, C. Kell, M. Willmers, E. Gray, and T. King, Seeking Impact and Visibility. 2014.

[3] 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.

[4] I. J. Ezema and O. B. Onyancha, "Open access publishing in Africa: Advancing research outputs to global visibility," African Journal of Library, Archives and Information Science, vol. 27, no. 2, pp. 97–115, 2017.

[5] "OpenDOAR: Browse by Country - V2.sherpa". Accessed on: March 1, 2021. [Online]. Available: https://v2.sherpa.ac.uk/view/repository_by_country

[6] S. Harnad, "The self-archiving initiative," Nature, vol. 410, no. 6832. pp. 1024–1025, 2001.

[7] J. Kim, "Faculty self-archiving: Motivations and barriers," Journal of the American Society for Information Science and Technology, vol. 61, no. 9, pp. 1909–1922, 2010.

[8] W. Y. Arms, Digital Libraries, 2nd ed. Cambridge, Massachusetts: MIT Press, 2000.

[9] W. Y. Arms, "Key Concepts in the Architecture of the Digital Library", D-Lib Magazine, vol. 1, no. 1, Jul. 1995.

[10] W. Y. Arms, C. Blanchi, and E. A. Overly, "An Architecture for Information in Digital Libraries," D-Lib Magazine, vol. 3, no. 2, Feb-1997.

[11] "arxiv.org e-Print archive". Accessed on: March 1, 2021. [Online]. Available: https://arxiv.org

[12] "RePEc: Research Papers in Economics". Accessed on: March 1, 2021. [Online]. Available: http://www.repec.org

[13] M. Kökörčený and A. Bodnárová, "Comparison of digital libraries systems," in Proceedings of the International Conference on Data Networks, Communications, Computers, 2010, pp. 97–100.

[14] "Omeka". Accessed on: March 1, 2021. [Online]. Available: https://omeka.org

[15] "DSpace - A Turnkey Institutional Repository Application". Accessed on: March 1, 2021. [Online]. Available: https://duraspace.org/dspace

[16] "EPrints Services". Accessed on: March 1, 2021. [Online]. Available: https://www.eprints.org/uk

[17] H. Suleman, "Interoperability in Digital Libraries," E-Publishing and Digital Libraries: Legal and Organizational Issues, Information Science Publishing, 2010, pp. 31–47.

[18] "NDLTD Union Archive of ETD Metadata". Accessed on: March 1, 2021. [Online]. Available: http://union.ndltd.org/portal

[19] "OATD - Open Access Theses and Dissertations". Accessed on: March 1, 2021. [Online]. Available: https://oatd.org

[20] J. Allinson, S. François, and S. Lewis, "Sword: Simple web-service offering repository deposit," Ariadne, no. 54, 2008.

[21] C. Lagoze and H. Van de Sompel, "The Open Archives Initiative: Building a Low-Barrier Interoperability Framework," in Proceedings of the first ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '01), 2001, pp. 54–62.

[22] 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.
[23] M. Smith, M. Barton, M. Branschofsky, G. McClellan, J. H. Walker, "DSpace: An Open Source Dynamic Digital Repository," D-Lib Magazine, vol. 9, no. 1, Jan. 2003.
[24] J. Brooke, "SUS: a retrospective," Journal of Usability Studies, vol. 8, no. 2, pp. 29–40, 2013.
[25] A. Bangor, P. Kortum, and J. Miller, "Determining what individual SUS scores mean: Adding an adjective rating scale," Journal of Usability Studies, vol. 4, no. 3, pp. 114–123, 2009.