2022

Library Services and Linked Data at Makerere University: Prospects of a Research-Led University

Onan Mulumba
Makerere University, onan.mulumba@mak.ac.ug

Library Services and Linked Data at Makerere University: Prospects of a Research-Led University

Onan Mulumba

Makerere University, Kampala – Uganda

onan.mulumba@mak.ac.ug

Abstract

Makerere University is the oldest and most prestigious university in East Africa. As one of the best universities on the African content, Makerere is aligned to three main activities; Teaching, Research, and Community Outreach, and these are inclined to 5 core values; allegiance to the institution, integrity, customer responsiveness, professionalism, and openness to diversity. Makerere University has a library system, hosting both electronic and print resources. The digitization of print resources has created a system, through which delicate material is conserved. The files and their metadata, using Dublin Core, are stored in the institutional repository and in the integrated library system (Virtua). The library system further generates library usage data which can be used in making acquisitions decisions. For most research output files, the data is available and would be archived alongside the digital files. Other than the National Biodiversity Data Bank which is transitioning to online, the rest of the data remains stored on authors' personal computers and other storage tools. With support from the government and other development partners, Makerere university strives for a shift from a more academic-driven to a research-led university. This calls for more support from the university administration and other key units, to provide access points to linked data. This paper aimed at establishing the status of linked data at Makerere university, challenges encountered in data management, and prospect strategies. The researcher used an exploratory research methodology to generate key findings, from existing literature and personal experiences from experts. The findings reveal that Makerere university researchers archive their data in external repositories and that, while the Makerere Institutional Repository has the potential to host linked data, this service has not been exploited basically due to lack of skilled labor force, awareness, financial resources, and policy. However, this service, when incorporated in the institutional research agenda, can be implemented and also enhance data linking in the NBDB. It is anticipated that the results from this study will guide Makerere and other institutions in Uganda to embrace the value of linked data and therefore ignite its implementation.

Keywords: Linked Data, Research data management, Open repositories, Open Data,

Introduction

Makerere University was established in 1922 and is currently one of the most prestigious universities in Africa (Makerere University, 2020a). Makerere University is the oldest university in East Africa and manifests as one of the best universities on African content. Makerere is aligned to three main activities; Teaching, Research, and Community Outreach, and these are inclined to 5 core values; allegiance to the institution, integrity, customer responsiveness, professionalism, and openness to diversity (*Makerere University Annual report 2017*, 2017). Constructed in 1959 is the main library which is part of the university library system and located on the main campus. Besides many other sections, the university library comprises the Information and Communication Technologies (ICTs) section, the digitization section, the technical services section, and the reference and circulation section, which are directly in charge of making resources accessible. However, the collection that Makerere university library boost includes monographs and electronic journals, and has nothing to do with data files, whether in electronic or print, linked or unlinked (Makerere University Library, 2019). While the currently developed strategic plan for Makerere University is aimed at transforming the University into a "research-led" institution with a multi-faceted research agenda and enhanced engagement with the industry and business sector, it is silent on how the research data will be managed and/or linked (Makerere University, 2020c). Additionally, among the policy framework for the university, the data security policy only focuses on the security of administrative data and not research data.

A recent study by Mulumba et al. (2017) highlights the lack of linked data at Makerere university, mainly stemming from lack of awareness and skilled manpower, and also due to internal structures like lack of a data management policy at the institution. However, Mushi et al. (2020) connote, in their study about the identification and implementation of RDM services at the university of Dodoma library, that the inexistence of RDM policy shouldn't be an excuse for not implementing RDM skills and services. In a study about the role, academic libraries can play in developing Research Data Management Services at Makerere University Ssebulime (2017) established that "the extent to which research data are being managed across the entire institution is not yet fully aligned with the international RDM standard practices which aim at addressing the reproducibility crisis". Furthermore, it was observed that researchers at Makerere University generate large volumes of data which they manage, control, and store using removable devices, a risky practice.

National Biodiversity Data Bank (NBDB)

Established in 1990, the National Biodiversity Data Bank (NBDB) aimed at providing immediate reaction to the information needed by conservationists to have prompt access to information and data regarding the nation's

biodiversity, and to illuminate dynamic procedures in Uganda and globally, that would influence biodiversity and the general environment (Wildlife Conversation Society, 2018). With its initial funding from USAID, it was agreed to be hosted at Makerere University. With support from the Wildlife Conservation Society (WCS) and USAID, Makerere University, on 5th September 2019, launched a web application to enhance biodiversity research in Uganda (Wamai, 2019). It is reported that presently, the NBDB contains about 8,000 species entries, 140,000 georeferenced species locality records, a gazetteer file with about 7,000 entries that are places of biological recording, data on protected areas, data on administrative units, data on recorders/observers that have contributed data or whose data have been computerized, and citations (Wildlife Conversation Society, 2018). The web application is still in its infant stage and sometimes not able to retrieve desired results from the search query. Furthermore, there are no strategies highlighted to make the data linked, thus keeping it geographically localized.

Linked data

With the use of RDF, several institutions and organizations are using the semantic web to make data accessible (Sharma et al., 2018). It is further noted that due to the scale and complexity of data Librarians should pay more attention to the global trends and thus prioritize the integration of library data to the web. This would enable access and reusability of data by other professionals.

From a technical perspective, Bizer et al. (2009) define Linked Data as "data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can, in turn, be linked to from external data sets". On the other hand, linked data is perceived as a "well-established standard for publishing and managing structured information on the Web, gathering and bridging together knowledge from different scientific and commercial domains" (Po et al., 2020).

Linked data life cycle

Rodríguez (2012) elicits a cyclic four staged Linked data cycle including stages such as produce, publish, consume, and feedback/update. These four stages, however, can each undergo a validation process. Another study presents six stages starting from data awareness through use cases (Hausenblas, 2011). However, a more recent study to deconstruct the data life-cycle in digital humanitarianism brings forth a more robust data life cycle unraveling categories of organizations and people (Roth & Luczak-roesch, 2020). While this cycle exhibits the four main stages; data acquisition, processing, storage, and use. However, it combines two types of life cycles; a cycle that presents data sets from a particular knowledge domain, and a life cycle that presents data from different knowledge domains, in the context of "big data". Therefore, the participation of the actors tends to vary at each stage in the

two categories of the data life cycle. On the other hand, Ngomo et al. (2014) assert that "all datasets published as Linked Data share a uniform data model" and what differentiates them is the mode of presentation. In their study, they present a link that life cycle with 8 cyclic stages from Manual revision and authoring to storage and querying, as shown in figure 1.

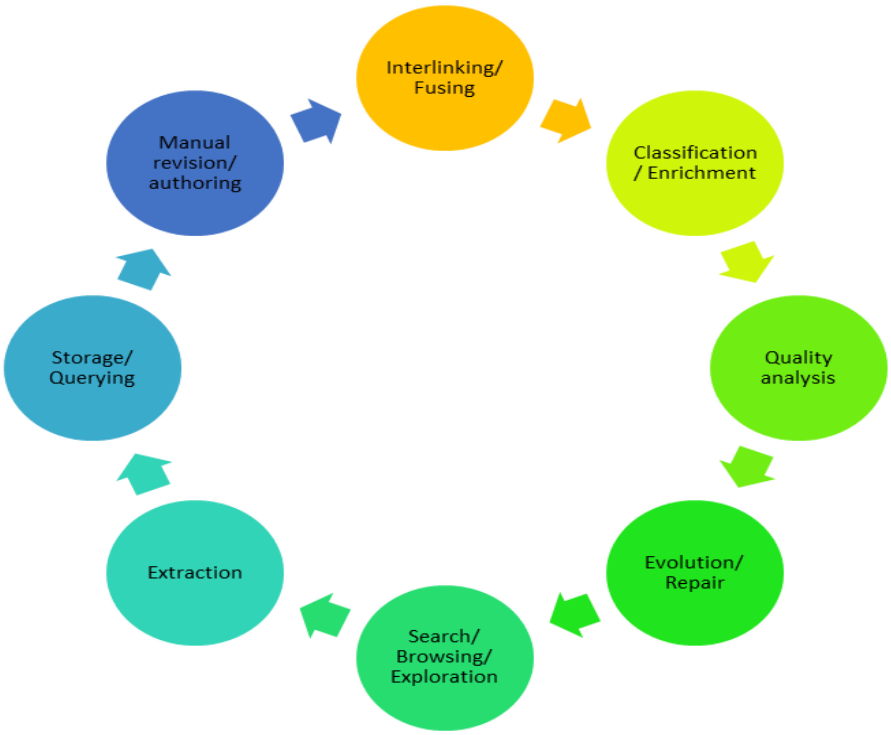

Fig 1. The Linked Data Life-Cycle (Ngomo et al., 2014)

This linked data life-cycle is comprehensive and applicable in the situation of Makerere University. If all resources are in place, Authoring/ manual revision would provide for generation of data, Interlinking/fusion would create a connection for the data, Classification/ enrichment would allow for processing, the Quality analysis would provide for technical review, Evolution/ repair would create room for adjustment, Searching/ browsing/ exploration would give room for accessibility, Extraction would allow for data recall, and Storage/ querying would provide for use and reuse, or feedback that would require the author to review and make adjustments. In the case of the latter, the cycle continues.

Research statement

Makerere University ranks 5th best in Africa, according to the Times Higher Education (THE) 2020 rankings released in September 2019 (Lumu, 2019). Furthermore, a lot of research is undertaken at Makerere University and data is collected. This is supported by the government monetary investment in form of the Research and Innovations Fund (RIF) equivalent to 30 billion Uganda shillings annually, in addition to external sources of funds through donations, grants, scholarships, and partnerships. Consequently, the research data

generated is never archived and linked, through the institutional repository to allow for access and reuse by other experts.

Objective

The main objective of this study was to establish the status of linked data at Makerere University and the prospect of linking the data through the Makerere Institutional Repository. The study was accomplished through the following specific objectives.

- Establishing the status of linked data at Makerere University.
- Instituting strategies of linking the data at Makerere University
- Identifying the underlying challenges and prospect solutions.

Methodology

Due to the vagueness of the existing situation about linked data at Makerere university, the researcher used an exploratory research method to gain an understanding of how data is managed at Makerere university, the challenges that exist, and the prospective strategies to steer the data linking process. The researcher focused on secondary data from already existing situations, key informants especially from the information management and digitization team, and from their own experience as an expert in information science. The researcher further explored some of the external open data repositories where Makerere university researchers archive their data and these include Mendeley data, Zenodo, Figshare, UK data service, Open Science Framework, and database. However, the researcher selected Mendeley data as a focus for this study.

Data linking at Makerere University

Makerere University produces a lot of administrative and research data every year. However, most of the data remains redundant and eventually disposed of after its intended purpose. There are many projects and collaborations within the university and some of the data generated through these projects and partnerships is stored and linked in external open data repositories or at partner institutions. Mendeley web contains a gateway and one of the open repositories through which some data generated at Makerere university is linked to the web (data.mendeley.com). There are 1495 data sets directly linked through Mendeley data, as shown in Fig 2.

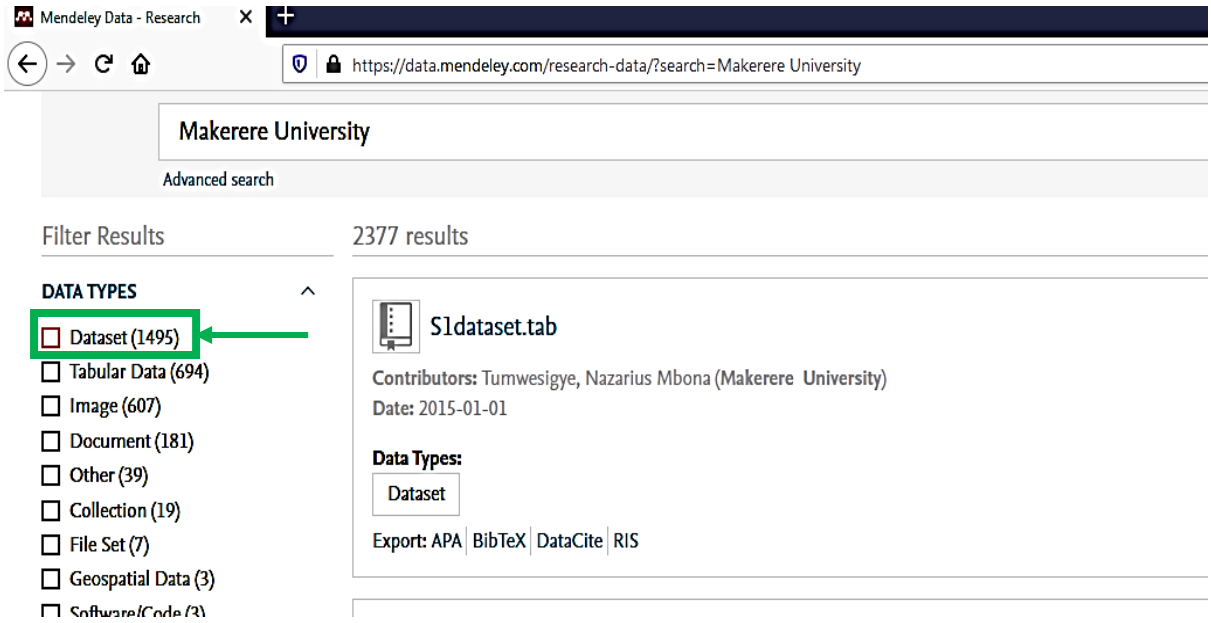

Fig 2: Makerere University visibility on Mendeley data

Some data sets are archived at other open data repositories but linked and accessible through Mendeley as shown in figure 3.

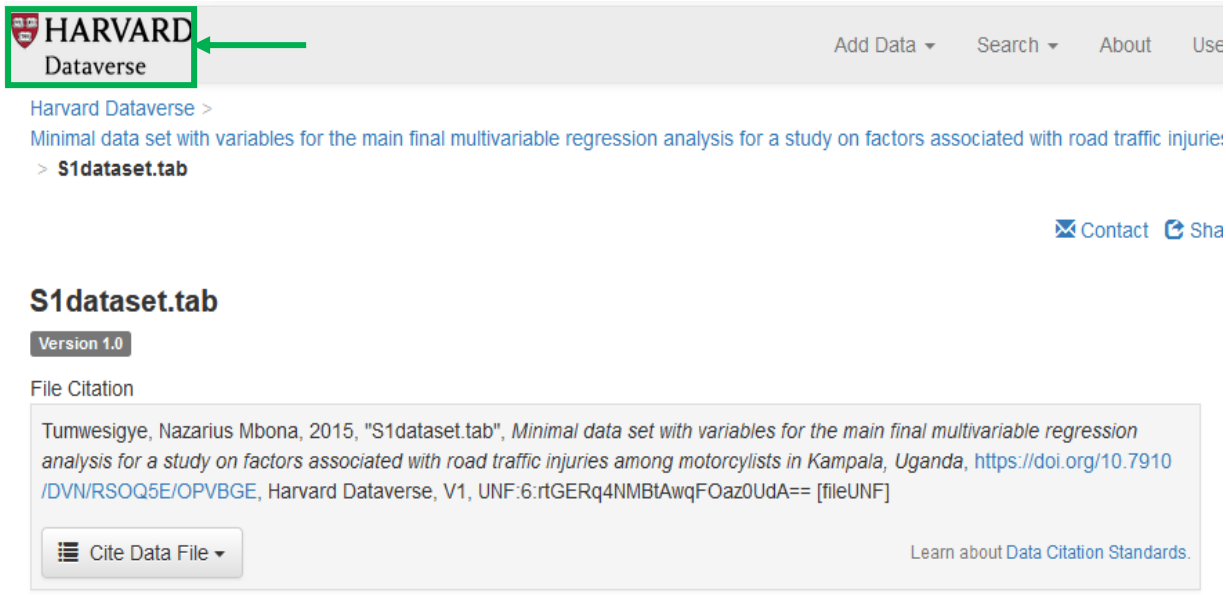

Fig 3: Makerere university research data file in external repositonnnnnnry at Harvard dataverse

Attempting to link the data sets already archived in other external open repositories the experts will have to generate metadata from the existing files and submit it for review and approval in the Makerere Institutional Repository. Some of the key metadata fields would capture details like data authority, title, publisher, URI, and year of publication, as exemplified in figure 4.

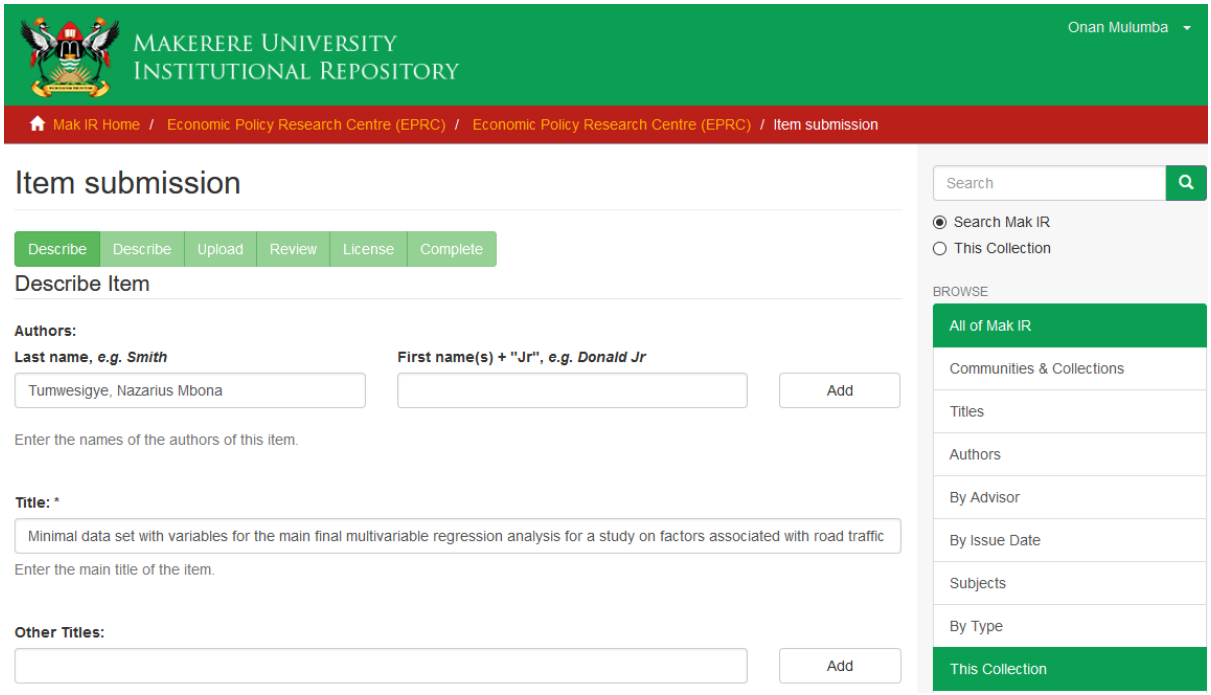

Fig 4: Metadata form in MakIR for item submission

During the process of capturing the metadata, the systems automatically generate a URI, alongside which the submitter captures the URI from the data source, as a separate entity. After a complete submission, the entry is reviewed, for quality control, and eventually approved for archival. The final entry as shown in figure 5 contains 2 URI, one, a system-generated for the entry in MakIR and the other providing a direct link to the source of the data file.

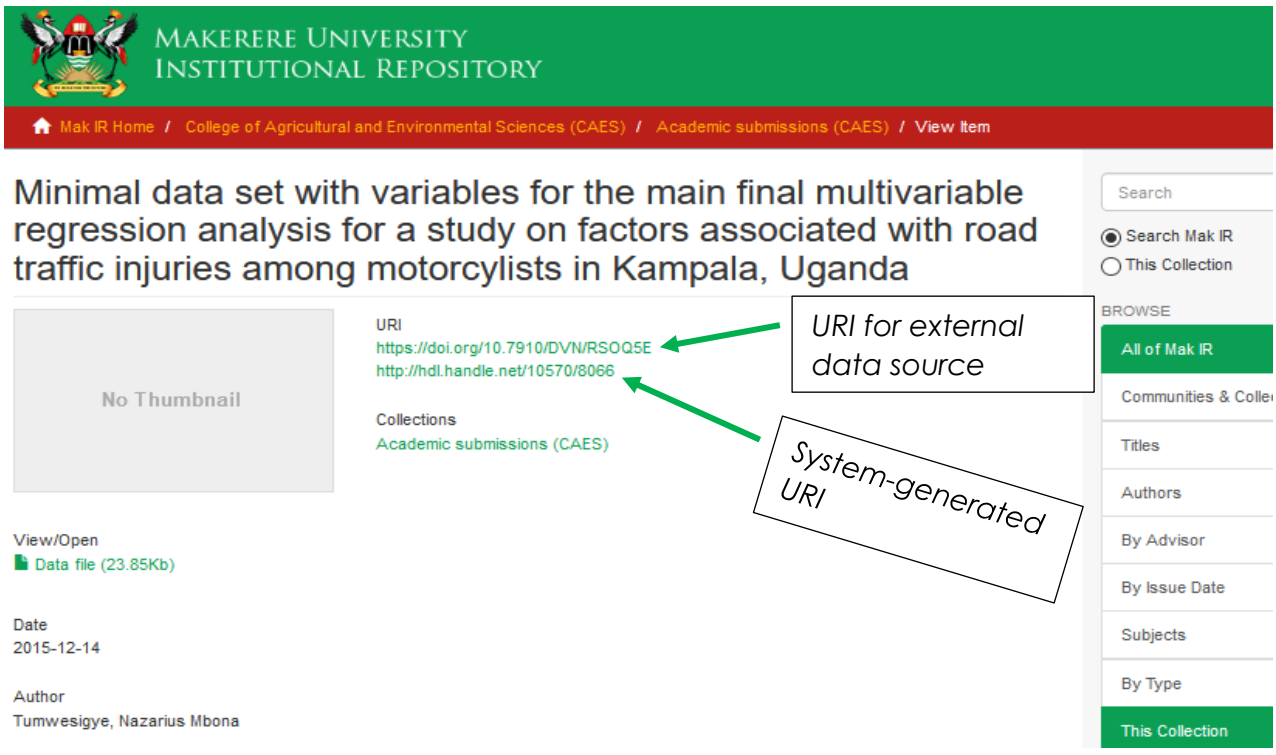

Fig 5: Linked data file in MakIR

The system-generated URI can be used to link a data file originally archived in MakIR to external data repositories, as shown in figure 6. This URI is clickable and will directly divert a searcher to the source of the data (MakIR).

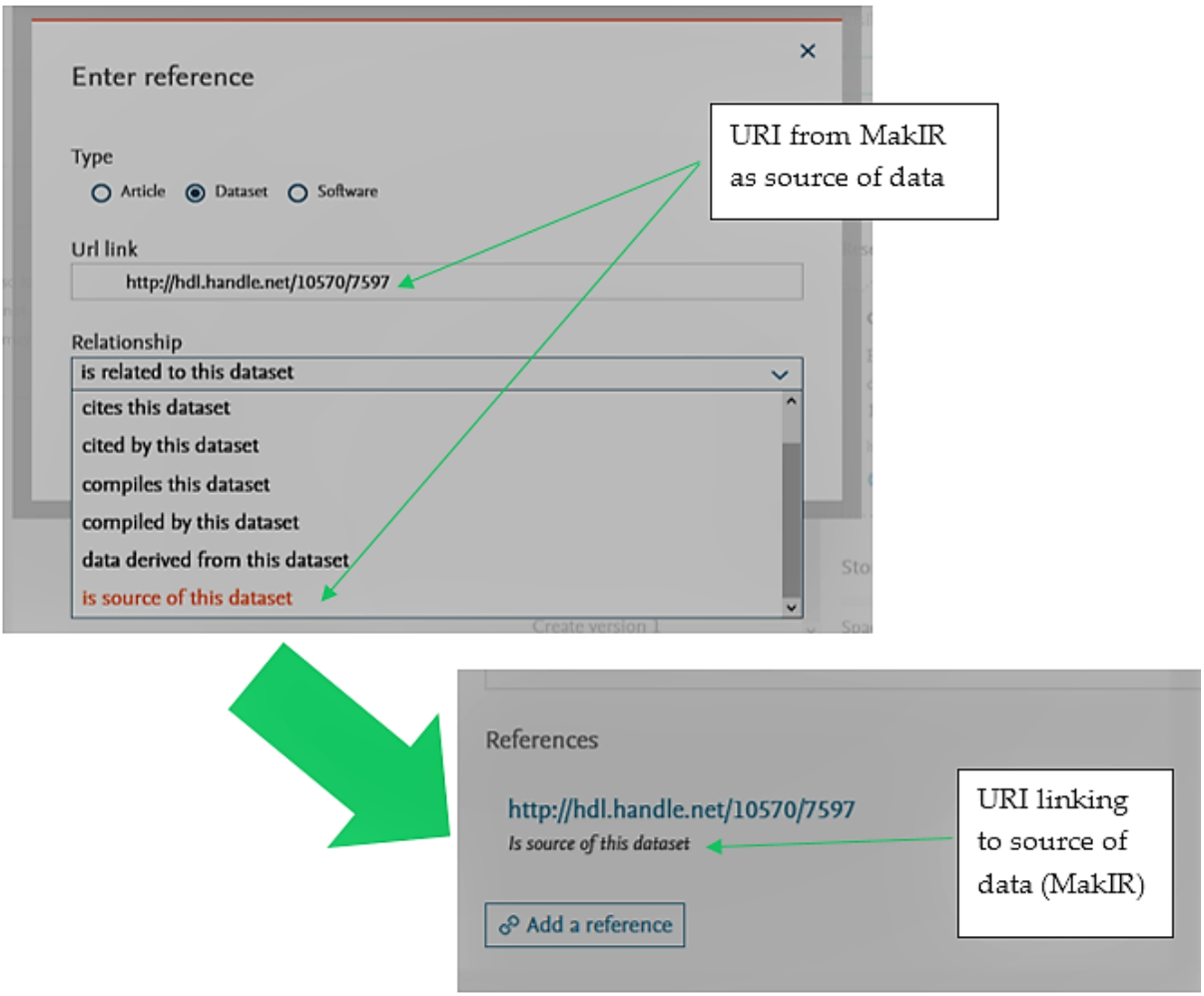

Fig. 6: Data linking of file from MakIR in Mendeley data repository using URI

Data linking remains a key concern at Makerere University despite strategies to transform it into a more research-led university. Given the university's vast experience, Makerere established a multi-disciplinary research agenda. The disciplines in limelight include Education for Development, Food, Nutrition and Value Addition, Sustainable Environment Development, Appropriate Technology, Good Governance, Equity, Service Delivery, Methodological Studies, Health, Natural Resources Utilization and Conservation, Biotechnology, Economics, and Staff Development (Makerere University, 2020b). Additionally, the 2020-2030 strategic plan will see Makerere's research is driven by the fourth industrial revolution, redirecting resources towards emerging technologies in the fields of "robotics and artificial intelligence, big data, quantum computing and the Internet of Things, as well as Nano and biotechnology" (Makerere University, 2020c). Therefore, more resources are oriented towards these

research themes and a lot of research data is generated in the process. Archiving and linking of all generated data in the Institutional repository ought to be considered as a key strategy by the Makerere University Directorate of Research and Graduate Training. This would save resources wasted on research activities replicating already existing inaccessible data.

It is noted that as science becomes more collaborative, computational, and data-intensive, the data management needs coupled with funding directives requiring data management planning will influence the direction of research at Makerere University (Ssebulime, 2017).

Challenges

Several impediments face data management and sharing at Makerere. Mainly these relate to financial constraints to pay for data archiving and sharing, especially through commercial services. Other challenges include creating meaningful datasets and metadata, lack of technical skills in organizing and sharing data by compliance needs, and lack of skills in data archival in repositories (Ssebulime, 2017). Additionally, the same study also notes the need to create awareness about research data management at Makerere University. Many times, the lack of a data management policy affects activities relating to sharing and linking of data. Many researchers and institutions exhibit fear for violating the law, in case of underlying restrictions, or the absence of guiding principles.

Strategies

Makerere University, is a leading national and regional university ought to extra measures to champion research through increasing access to internally generated data. Firstly, the institution should develop data management awareness plans for the staff and student communities. These can be fostered through either face-to-face training or electronically. Brochures or handouts can also be produced and distributed for reference. Secondly, the university administration should put in place a policy to govern data management and linking. The policy may require research submissions to be accompanied by data files and the creation of linked data. Thirdly, the university skill up the technical personnel to manage data. However, this should depend on the type of data generated because the management of big data varies from that of subject-specific conventional data, as supported by Roth and Luczak-roesch (2020). In addition to all other strategies, the university should dedicate a budget to support data linking activities and associated initiatives.

Conclusion

A lot of research data is generated at Makerere University and within the country. The biodiversity data is from the country is archived in the National Biodiversity Data Bank but the rest of the data remains stored on personal

computers and other storage devices. However, the existence of externally linked data structures provides an opportunity for Makerere to adopt and move alongside other institutions, which have fully implemented linked data services. Makerere university library and DRGT need to champion the process, create awareness, lobby for budget allocation, institute a policy, and acquire a skilled labor force, so as the overcome the prevalent challenges. Additionally, steps should be taken to ensure that the biodiversity data in the NBDB is linked to allow for wider discoverability and access.

References

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data - The story so far. *International Journal on Semantic Web and Information Systems*, *5*(3), 1–22. https://doi.org/10.4018/jswis.2009081901

Hausenblas, M. (2011). *Linked data life cycles*. Slideshare. https://www.slideshare.net/mediasemanticweb/linked-data-life-cycles

Lumu, D. (2019, September 18). Makerere University ranked fifth best in Africa. *New Vision*. https://www.newvision.co.ug/news/1507233/makerere-ranked-fifth-africa

Makerere University. (2020a). *Historical Background*. Makerere University. https://www.mak.ac.ug/about-makerere

Makerere University. (2020b). *Research agenda*. Research. https://mak.ac.ug/research/research-agenda

Makerere University. (2020c). *Unlocking the Knowledge Hub in the Heart of Africa: Strategic plan 2020 - 2030*. Makerere University.

Makerere University Annual report 2017. (2017). https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi3z5Ge8bvqAhUBylUKHd56AG8QFjACegQIAhAB&url=https%3A%2F%2Fpdd.mak.ac.ug%2Fwp-content%2Fuploads%2F2018%2F11%2FMakerere-Anual-Report-2017.pdf&usg=AOvVaw3Z7lBt-PNi7wkvVqpIlanl

Makerere University Library. (2019). *About us*. University Services. https://mulib.mak.ac.ug/index.php?option=com_content&view=article&id=1&Itemid=103

Mulumba, O., Kinengyere, A. A., & Akullo, W. N. (2017). The Horizon of Information Sharing at Makerere University : Multimedia, Audiovisual, and the Linked Open Data. *American Journal of Data Mining and Knowledge Discovery*, *2*(4), 114–120. https://doi.org/10.11648/j.ajdmkd.20170204.14

Mushi, G. E., Pienaar, H., & Deventer, M. van. (2020). Identifying and implementing relevant research data management services for the library at the University of Dodoma, Tanzania. *Data Science Journal*, *19*(1), 1–9. https://doi.org/10.5334/dsj-2020-001

Ngomo, A.-C. N., Auer, S., Lehmann, J., & Zaveri, A. (2014). Introduction to Linked Data and Its Lifecycle on the Web. In M. Koubarakis, G. Stamou, G.

Stoilos, I. Horrocks, P. Kolaitis, G. Lausen, & G. Weikum (Eds.), *10th International Summer School 2014* (pp. 1–99). Springer. https://doi.org/10.1007/978-3-642-23032-5

Po, L., Desimoni, F., Ding, Y., & Groth, P. (2020). *Linked Data Visualization Techniques, Tools, and Big Data Synthesis Lectures on Data, Semantics, and Knowledge Synthesis Lectures on Data, Semantics, and Knowledge* (Y. Ding & P. Groth (Eds.)). Morgan & Claypool Publishers. https://doi.org/10.2200/S00967ED1V01Y201911WBE019

Rodríguez, J. M. A. (2012). *My Linked Data Lifecycle*. Research. http://www.josemalvarez.es/web/2012/04/06/my-linked-data-lifecycle/

Roth, S., & Luczak-roesch, M. (2020). Deconstructing the data life-cycle in digital humanitarianism. *Information, Communication & Society*, *23*(4), 555–571. https://doi.org/10.1080/1369118X.2018.1521457

Sharma, K., Marjit, U., & Biswas, U. (2018). Efficiently processing and storing library linked data using apache spark and parquet. *Information Technology and Libraries*, *37*(3), 29–49. https://doi.org/10.6017/ital.v37i3.10177

Ssebulime, J. B. (2017). *The role academic libraries could play in developing Research Data Management services: a case of Makerere university library* [University of Pretoria]. https://repository.up.ac.za/bitstream/handle/2263/74961/Ssebulime_Role_2017.pdf?sequence=1&isAllowed=y

Wamai, M. (2019). *Mak Launches Online National Biodiversity Data Bank*. Research News; Makerere University. https://news.mak.ac.ug/2019/09/mak-launches-online-national-biodiversity-data-bank

Wildlife Conversation Society. (2018). *National Biodiversity Data Bank*. Bational Biodiversity Data Bank. http://nbdb.mak.ac.ug/