Exploring the Architecture of Research Data Management

Pauline Ruguru Njagi*, Gitau Njoroge**

ARTICLE INFO

ABSTRACT

The study focused on model for effective research data management practices including metadata creation, design, storage, security, preservation, retrieval, sharing, and reuse. Data storage in a format that can be easily accessed, processed, and analyzed requires a functional architecture as datasets are often fragile and susceptible to storage malfunctions and advancing technology. The study used descriptive quantitative research design. Closed-ended questionnaire was the instrument of data collection. A total of 35 participants selected purposively were engaged to provide insight on study topic. The study used content validity to establish the degree to which the measure represented the paradigm of interest. Test retesting was done to establish questionnaire's reliability which yielded a Cronbach Alpha result of 0.78 ascertaining reliability before questionnaires administering. Data was analyzed using statistical analysis for social sciences and result presented in pie-chart. The finding informs of widespread use or plan to adopt the "Repository" data architecture model reflecting a widespread understanding among participants of its applicability for enabling research data management practices in academic libraries. However, the study recommends for further studies on repository model standards in academic libraries and blackboard functionality and effectiveness in research data management.

1. Introduction

In the past two decades, research data management has gained increasing importance due to the emerging demand for a broader and higher-quality range of data services that meet patron needs at various points in the research process (Wong & Chan, 2021). The scientific community is now placing more emphasis on exchange of open data. The shift of the scientific data paradigm and the rapid rise of the open access movement are driving advances in research data management, establishing universities as essential hubs for creating data management services and implementing research data management services to better meet the standards for data openness (Zhou, 2018).

*   Kenyatta University (rugurupauline@gmail.com) (First Author)
** Kenyatta University (Gitau.njoroge@ku.ac.ke) (Corresponding Author)

Research data is the information gathered, observed, or developed for analysis in order to construct the original study. The data may comprise of variables like data from surveys, seismic simulations; data from labs; and data derived or compiled form testing algorithms or text mining. Before sharing data, challenges including metadata compilation, data navigation, and copyright protection must be handled, frequently with the help of a library.

Research data differ among disciplines and take different forms, including textual, qualitative, quantitative, images, recording, verbal communication, experimental readings, codes, and simulations hence different types of hosting models (Tripathi, Shukla, & Sonkar, 2017). Also, the upsurge in the generation and use of massive datasets as part of the research process fuels the need for the adoption of data architecture model to allow the storage of data in a form that can allow easier access, processing, and analysis (Cox & Pinfield, 2014). Research data can be found in a variety of digital file formats, such as text, numbers, images, and video. The essence of managing research data is to allow harvesting by researchers for knowledge advancement and meet funding and regulatory requirements. Suppressing data generated with public funds is viewed as undemocratic, and restricting access to a public asset is unacceptable. Also, without sharing data, it is hard to verify study findings, which is a core principle of good science (Cox & Pinfield, 2014).

Developed states championed to formulate and implement RDM policies to allow adoption of functional research data model specially journal publishers and research grant commissions to manage and share research data (Kinde, Addis, & Abebe, 2021). Clear policies for open data and data sharing have been created by the National Science Foundation, the Medical Research Council (MRC), the Arts and Humanities Research Council (AHRC) in the UK, and many other research funding agencies. Similar rules for data availability and sharing have been put in place by private organizations like the Gates Foundation, Ford Foundation, and Sloan Foundation (Zhou, 2018). Data support services are provided by more than half of US academic libraries.

Over 30 universities in the UK are working on data management initiatives with support from organizations like the Digital Curation Centre and the Joint Information Systems Committee (JISC). China has conducted a great deal of research in a number of disciplines on research data management service models. This covers studies on data literacy instruction, comprehensive case studies of tangible services, assessments of data management policies, evaluations of research data management system platforms, and analysis of research data management services based on network research (Zhou, 2018).

In Africa, South Africa championed expressing her commitment to openness by signing the Berlin declaration on openness to make scholarly output visible, accessible, searchable, and useable by a potential community of researchers (Kahn et al., 2014). The hosting of a workshop by the library and Information Association of South Africa (LIASA) in 2014 in cooperation with the UK digital curation center allowed librarians to evaluate the changing shift in the research data management landscape. Ng'eno (2018), study of the agricultural research institute in Kenya observed that some developing states have adopted strategies to enhance RDM services though, the study noted inadequacies in technical proficiencies and RDM architecture as significant contributors to incomplete, inaccurate, and loss of data hindering important activities such as wide data distribution and reuse.

2. Problem Statement

The alteration of the scientific data paradigm, as well as the quick expansion of the open access movement, have accelerated the development of research data management. As a result, universities have emerged as important hubs for developing data management services. While developed states understands research data management as a multifaceted process that includes handling and monitoring data at every stage of its lifecycle, not only managing it, a contrary scenario exists in developing countries. There is knowledge gap of research data management particularly covering publishing, sharing, and reusing data in addition to describing, storing, and preserving data over time.

In Africa, institution of higher learning have shown commitment to openness by signing the Berlin Declaration to adapt the evolving research data management (RDM) landscape though significant challenges exist. Ng'eno's (2018) study of the agricultural research institute in Kenya highlighted that while some developing states have adopted strategies to enhance RDM services, there are still notable inadequacies in technical proficiencies and RDM architecture. These deficiencies contribute to incomplete, inaccurate, and lost data, which hinder critical activities such as wide data distribution and reuse, thereby impeding the overall effectiveness of research efforts in the region.

2.1 Research Questions

What model could be appropriate for implementation in academic libraries for research data activities.

2.2 Research Objectives

To establish the research data management architecture used by academic libraries.

2.3 Study Limitation

The scarce literature on research data management architecture impacts on the study ability completely represent the diversity of existent systems. Also, small sample size limits how broadly the results may be applied. These variables could affect the study's capacity to offer a thorough analysis or make generalizations regarding research data management procedures.

2.4 Literature Review

Research takes various dimensions, including statistical, investigational results, consultation recording, transcriptions, physical records or files, and terabytes of data on shared servers hence need for domain-specific expertise to enhance adoption of data- model to help in the translation of research data into metadata, disseminate, and archive valuable results (Cox & Pinfield, 2014). Also, gathering, sorting, analyzing, classifying, and storing research data requires integration within the framework of scientific research. Research data management entails giving researchers access to processed,

high-value data, individualized guidance, and support across the whole data life cycle (Zhou, 2018). The main ways that metadata services help researchers are in producing metadata that complies with standards, enhancing dataset interoperability, raising the possibility of data discovery, and providing more comprehensive and in-depth descriptions of the data. Metadata services can be approached in two ways. The first is developing customized training programs that concentrate on particular metadata standards. These programs are labor-intensive but flexible. The second strategy makes use of metadata tools or systems (like Morpho) that produce metadata records automatically or that progressively assist users in accordance with project or discipline-specific requirements (Zhou, 2018).

According to Yoon and Schultz (2017), specialized metadata can be made accessible through adoption of models that allow storage of dataset in repositories. Geological Survey (2017), informs that RDM products should be stored in appropriate model to allow subjection to rigorous testing to ascertain adherence to the mandatory procedures mostly stipulated in the data management plan. Morgan, Duffield, and Walkley Hall (2017) advocates the need to adopt models that allow quality data capture for processing, organizing, and structuring of data files, validation to enhance flexible conversion, and transfer to intended destination. Andrikopoulou, Rowley, and Walton (2021), point that data curation process depends on adopted model to enhance appraisal including selection, digitizing and transcribing, validating and cleaning, anonymizing data, describing, managing, and data storage. Concurring, Ng'eno, and Mutula (2022), point that functional RDM model helps in quality control and assurance measures to provide consistent naming of data, search, and retrieval in data repositories.

A number of academic storage systems are operating at high levels, such as the DSpace at the MIT Library, the DataSpace at Princeton academic, the DataStar at Cornell University, the E-Data at Purdue University, and the HMDC at Harvard University. These platforms provide scientific researchers with storage and sharing services for research data (Zhou, 2018). Oxford University has a two-tier data management storage system that helps scientific establishments manage and maintain their data while also satisfying the needs of researchers for local data management. The Fudan University social science data platform was introduced in 2014 by Harvard University's Dataverse Network to provide tools for online analysis and sharing, services for universities, research institutes, and government organizations to store, publish, and exchange research data. It was founded on the open-source software DSpace with set standards for data submission, organization, preservation, sharing, and utilization. Consequently Zhou (2018) proposed the need to comprehend and assess external storage solutions that are appropriate for scientific researchers in addition to creating and promoting an internal storage system that satisfies user needs.

Zhou (2018), Pointed the importance of research data management is to gather, compile, and assess the data that has been saved in order to find connections with other relevant information. Equally, carry out secondary development to increase the value of the data and make it easier for other scientists to utilize by providing channels for sharing that have been specified. Cox and Pinfield (2014), uses the lewis-Corrall archetypal to explain RDM activities hierarchically including policy-making, training, and mapping of the potential role of information professionals which should be considered in choosing RDM model. Also, Cox, and Pinfield (2014) proposes use of DCC Lifecycle Curation Model, often associated with record administration to guide the adoption of RDM model

though, to some extent fails to solve challenges associated with RDM scale complexity. Also, professional information researchers proposed a nine-area pyramid to map RDM activities, which is more implicit and strategic as compared to Lewis model. The nine-level pyramid archetypal is significant due to the incorporation of national policy and partnership with educational providers. However, the archetypal fails to address intra-organizational collaboration and librarians' roles which need a multi-professional approach.

Pinfield, Cox, and Smith (2014), proposed an RDM data-centric architecture with drivers such as storage, jurisdiction, and technologies drives. Tripathi, Shukla, and Sonkar (2017), point that institutions need to consider functional RDM model that allow seamless access, browsing, consulting, and built-in for future academic work and research activities. Alhussain (2017), proposes blackboard citing the model possibility of information exchange, effective communication, and comprehensive framework for comparing success across multiple information systems or usage of a single system, given its multidimensional and interrelated nature. Consequently, some studies often promote blackboard as a Learning Management System (LMS) that supports thousands of institutions worldwide in addressing educational challenges and driving innovation (Alhussain, 2017). Concurring Zhou (2018), point that data warehousing and technology platforms are included in the category of RDM architecture, which also includes storage systems for managing and storing research data. The efficient handling of research data, especially for academic and data-intensive companies, depends on this architecture. It is impossible to organize, preserve, and secure data effectively without a specialized platform.

Tripathi, Shukla, and Sonkar (2017) used the National Data Service (N.D.S.) a U.S. initiative and data providers association computing infrastructure providers and publishers to map up research activities including deposit, use, reuse, data analysis, to describe an RDM model that enhance support, promote, and strengthen research endeavors while allowing analysis of research data policies formulated and implemented in India. Yu, Deuble, and Morgan (2017), used a consultative leadership approach to explain RDM prototype that could support library services based on the research lifecycle including Vaughan et al. (2013), five-stage architecture comprising of constructs including; ideas development, funding, proposal, conducting, and dissemination of metadata. Unfortunately, the guide supported a few RDM activities such as locating data sources, preparing a data management plan, describing data, and navigating repository options for a comprehensive RDM model.

Yu, Deuble, and Morgan (2017), proposes the adoption of research lifecycle structure with listed RDM related activities across three research project phases. The structure has several stages stipulating activity including preparation of a data management plan, conducting ethical clearance, and training in the pre- research location. Also, the structure guide on procedures for research data collection and analysis; metadata generation; data storage and access during the research; publishing research data, and on- going curation in the post-research stage. According to Gries et al. (2018), the repository model is an open access data repository that share a standard framework for data deposition, discovery, and reuse by offering a consistent experience for both producers and users across repositories. Also, the model uses multiple metadata standards or export metadata in multiple specifications as employ the specification that best fits its data entities and curation procedures. Equally, Gries et al. (2018), point that repository model is automated hence increase the quality and complexity of metadata

making it easier to codify and incorporate minute details of the data that might be missed or ignored during manual compilation. Also, the repository model has explicitly described and regulated vocabularies, authoritative definitions, resolvable URIs, and unique identifiers attached. At several universities, research data management systems have recently merged tailored discipline management platforms and a variety of collaboration platforms as demand has increased. As a result, hierarchical or cooperative system structure like the DATA-PASS platform group of the US Data Management Alliance have been developed (Zhou, 2018).
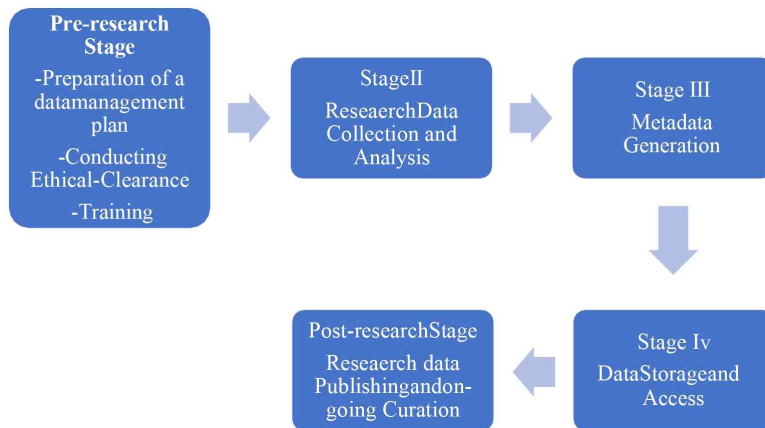
Fig. 1. Illustrating a Research Data Model
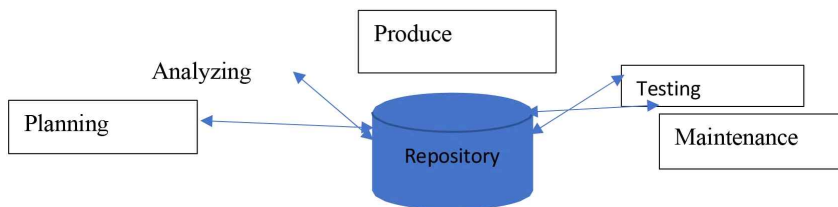Source: Yu, Deuble, and Morgan, 2017.

Fig. 2. Illustrating a Repository Architecture
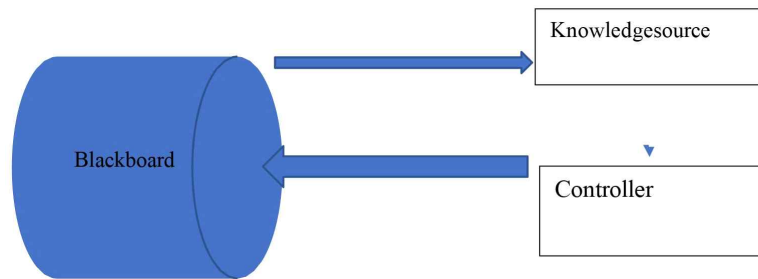Source: Researchers, 2024.

Fig. 3. Illustrating a blackboard Architectural model/style
Source: Researcher, 2024.

3. Study Objective

To establish the research data management architecture model used by academic libraries.

4. Research Questions

4.1 Methodology

The study employed a descriptive quantitative research method to gather quantifiable data. This design was chosen because it enabled a holistic understanding of the study topic and provided an opportunity to collect varied and diverse data. Data collection was conducted using closed-ended questionnaires, which proved effective for measuring participants' preferences, intentions, and opinions on the study topic. Additionally, questionnaires are flexible to administer and facilitate the analysis of responses, as noted by Watson (2015). A total of 35 participants were purposively sampled. The reliability of the questionnaires was confirmed using the Cronbach Alpha of 0.78. Also, content validity approach was used to establish the degree to which the study measure represented the paradigm of interest. The collected data was analyze using the statistical package for social sciences and result presented through a chart for understanding.

The information supplied shows the outcomes of a reliability test for a collection of items using Cronbach's Alpha.

Table 1. Illustrates the Case Processing Summary

		N	%
Cases	Valid	19	61.3
	Excluded[a]	12	38.7
	Total	31	100.0

Source: Researcher, 2024.

Table 2. Represent the Reliability Statistics Test Results

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.785	.883	58

A total of 58 products were tested for reliability. The Cronbach's Alpha coefficient of 0.785 indicates the reliability of internal consistency. A scale of 0 to 1, with higher numbers indicate greater dependability. A rating of 0.785 indicates a moderate to good level of internal consistency among the elements in this scenario. Based on standardized items, the Cronbach's Alpha was 0.883. This coefficient is frequently greater than the usual Cronbach's Alpha since it took into account the variance of each item as well as the covariance between items. When examining standardized items, the higher score (0.883) suggests increased reliability.

4.2 Study Findings

The findings show participants considerable preference for the repository model over the blackboard architecture for research data management (RDM) activities.

Table 3. Decision wise on Data Architecture Model

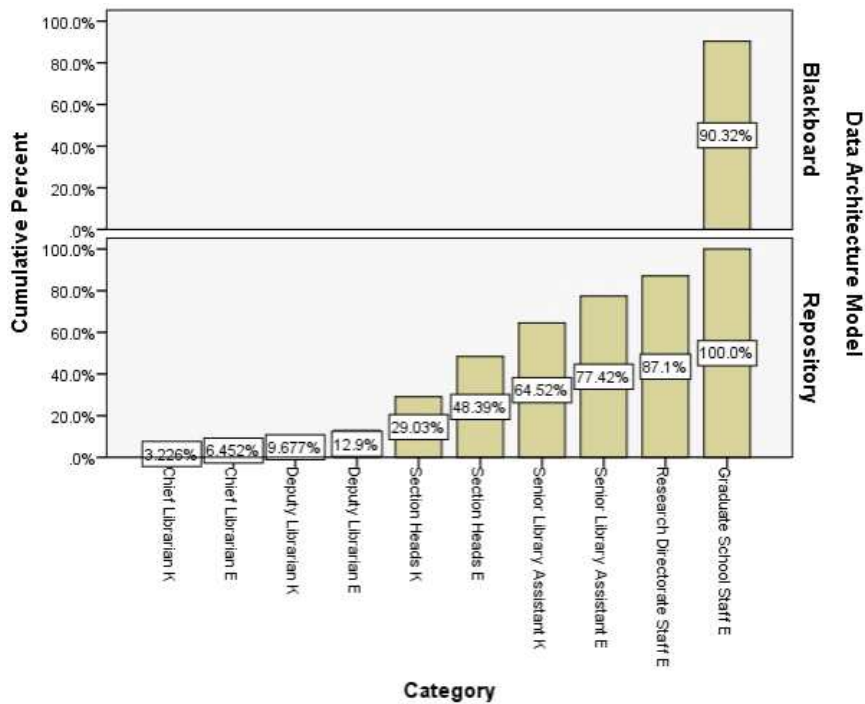|		Data Architecture Model					
		Blackboard			Repository		
Category	Chief Librarian K	0	0.0%	0.0%	1	100.0%	3.2%
	Chief Librarian E	0	0.0%	0.0%	1	100.0%	3.2%
	Deputy Librarian K	0	0.0%	0.0%	1	100.0%	3.2%
	Deputy Librarian E	0	0.0%	0.0%	1	100.0%	3.2%
	Section Heads K	0	0.0%	0.0%	5	100.0%	16.1%
	Section Heads E	0	0.0%	0.0%	6	100.0%	19.4%
	Senior Library Assistant K	0	0.0%	0.0%	5	100.0%	16.1%
	Senior Library Assistant E	0	0.0%	0.0%	4	100.0%	12.9%
	Research Directorate Staff E	0	0.0%	0.0%	3	100.0%	9.7%
	Graduate School Staff E	1	25.0%	3.2%	3	75.0%	9.7%

Source: Researcher, 2024.

Fig. 4. Showing Decision Wise for Research Data Architecture
Source: Researcher, 2024.

The study sought to establish the type of model libraries have adopted or intend to implement for effective research data management. The study findings reveal participants perception on different types of architecture to host and allow research data management activities. According to the finding, blackboard was less preferred with only one category of participants graduate school staff at 90.32% though the cumulative support is at 3.23%. The blackboard architecture refers to a collaborative and interactive educational platform where users can share material and engage in debates. The blackboard paradigm may be seen as less suitable for thorough and methodical handling of research data. The blackboard paradigm, which is commonly linked with learning management systems, may be perceived as less capable of handling the complexities of research data, metadata, and the different requirements of academic research workflows. As a result, academic libraries may prefer models created expressly for research data curation. Concurring Alhussain (2017), point that blackboard is not popular in hosting research data management practices as it often promoted as a Learning Management System (LMS) used by thousands of institutions worldwide in addressing educational challenges and to drive innovation.

The participant strongly supported the repository architecture including chief librarian's E/K 3.2%, 6.5%, deputy librarian K/E 9.7%, 12.9%, section head K/E 20%, 48.4%, senior library assistant K/E 64.5%, 77.4%, research directorate staff 87.1%, and graduate school at 100%. The cumulative support for the repository is 96.77% demonstrating that university libraries recognize and respect the repository-based architecture for hosting research data. A repository architecture is a centralized and structured system for storing, organizing, and distributing digital content, including research

data. The repository concept is adaptable and can house a wide range of research data, including datasets, articles, multimedia files, and other digital artifacts. This adaptability corresponds to the variety of data generated in academic study. Because of its emphasis on data protection and accessibility, academic libraries may choose the repository approach (Gries et al., 2018). Version control, metadata standards, and permanent identifiers are common elements of repositories, ensuring the long-term preservation of research. Concurring Pinfield, Cox, and Smith (2014); and Tripathi, Shukla, and Sonkar (2017), proposed a model that allow seamless access, browsing, consulting, and built-in for future academic work and research activities. Also, Gries et al. (2018), support the finding by informing that the repository model is an open access data repository that share a standard framework for data deposition, discovery, and reuse by offering a consistent experience for both producers and users across repositories. Also, the model uses multiple metadata standards or export metadata in multiple specifications as employ the specification that best fits its data entities and curation procedures.

5. Discussion

The findings revealed a broad agreement among participants regarding the best data architecture for research data management (RDM) activities in academic libraries, as evidenced by the overwhelming preference for the repository model (96.77%) and the small percentage of participants who expressed interest in the blackboard model (3.23%). A repository is a centralized and structured system for storing, organizing, and distributing digital content, including research data, version control, metadata management, access control, and preservation capabilities critical for protecting the integrity, accessibility, and lifespan of research data. The minority (3.23%) expressing interest in the blackboard paradigm suggests contemplation for an alternative method. The blackboard approach refers to a collaborative and interactive educational platform where users can share material and engage in debates. This decision may indicate a desire for a more engaged and dynamic atmosphere for research collaboration. Given the overwhelming preference for the repository model, academic libraries should invest in solid repository infrastructure. Implementing or improving systems that support metadata standards, version control, and persistent identification for effective research data management.

6. Conclusion

Given the overwhelming preference for the repository model, academic libraries should invest in solid repository infrastructure. Implementing or improving systems that support metadata standards, version control, and persistent identification for effective research data management is part of this. Effective RDM requires familiarity with repository functionality and best practices. When implementing repository models, academic libraries ought to take into account interoperability standards in order to enable smooth integration with other research infrastructure and systems. This guarantees that research data may be found, shared, and utilized again on a variety of platforms with ease. Consequently,

the study recommends further studies on how repository standard for research data management in academic libraries. Also, studies should be done to explore the blackboard model functionality and effective in research data management.

References

Alhussain, T. (2017). Assessing information quality of blackboard system. *International Journal of Computer (IJC), 25*(1), 1-7.

Cox, A. M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of librarianship and information science, 46*(4), 299-316.

De Silva, P. U., & Vance, C. K. (2017). Scientific Scholarly Communication: Moving Forward Through Open Discussions. In *Scientific Scholarly Communication* (pp. 1-15). Springer, Cham.

Elsayed, A. M., & Saleh, E. I. (2018). Research data management and sharing among researchers in Arab universities: An exploratory study. *IFLA journal, 44*(4), 281-299.

Gries, C., Budden, A., Laney, C., O'Brien, M., Servilla, M., Sheldon, W., ... & Vieglais, D. (2018). Facilitating and improving environmental research data repository interoperability. *Data Science Journal, 17*, 22-22.

Kinde, A. A., Addis, A. C., & Abebe, G. G. (2021). Research data management practice in higher education institutions in Ethiopia. *Public Services Quarterly, 17*(4), 213-230.

Morgan, A., Duffield, N., & Walkley Hall, L. (2017). Research data management support: sharing our experiences. *Journal of the Australian Library and Information Association, 66*(3), 299-305.

Ng'eno, E. J., & Mutula, S. M. (2022). Research data management in Kenya's agricultural research institutes. In *Handbook of Research on Academic Libraries as Partners in Data Science Ecosystems* (pp. 334-361). IGI Global.

Pinfield, S., Cox, A. M., & Smith, J. (2014). Research data management and libraries: Relationships, activities, drivers, and influences. *PLoS One, 9*(12), e114734.

Tripathi, M., Shukla, A., & Sonkar, S. K. (2017). Research Data Management practices in university libraries: A study. *DESIDOC Journal of Library & Information Technology, 37*(6), 417.

Vaughan, K. T. L., Hayes, B. E., Lerner, R. C., McElfresh, K. R., Pavlech, L., Romito, D., ... & Morris, E. N. (2013). Development of the research lifecycle model for library services. *Journal of the Medical Library Association: JMLA, 101*(4), 310.

Watson, R. (2015). Quantitative research. *Nursing standard, 29*(31).

Wong, G. K., & Chan, D. L. (2021). Designing library-based research data management services from the bottom up. In *Future Directions in Digital Information* (pp. 55-68). Chandos Publishing.

Yoon, A., & Schultz, T. (2017). Research data management services in academic libraries in the US: A content analysis of libraries' websites.

Yu, F., Deuble, R., & Morgan, H. (2017). Designing research data management services based on the research lifecycle–A consultative leadership approach. *Journal of the Australian Library*

and Information Association, 66(3), 287-298.

Zhou, Q. (2018). Academic libraries in research data management service: Perceptions and practices. *Open Access Library Journal, 5*(6), 1-4.

[About the authors]

Pauline Ruguru Njagi is a dedicated information science professional currently pursuing a PhD in Information Science at Kenyatta University. She holds a Master's degree in Information Science and a Bachelor's degree in the same field. As a prolific scholar, Pauline has published four journal articles covering themes such as institutional repositories, collaborative partnerships in research data management, the legal framework governing research data management, and demystifying open access for improved research discoverability. She has also contributed a book chapter on the role of libraries in achieving Sustainable Development Goal 4 (SDG 4) by bridging the universal literacy gap. Pauline research interests encompass research data management, open access, and data mining, reflecting her engagement with contemporary trends in information science. Professionally, she serves as a Senior Library Assistant at Murang'a University of Technology. In addition to her role there, she teaches at three public universities: University of Embu, Karatina University, and United States International University-Africa. Pauline contributions to the field are marked by her commitment to advancing knowledge and improving information access and management. Her work continues to impact the academic and professional communities significantly.

Gitau Njoroge is currently the Chief University Librarian and Senior lecturer in the Department of Library and Information Science at Kenyatta University. He is also the Eifl open access country coordinator ended in 2023. He holds a Ph.D. degree in Library and Information Science from Moi University specializing in the utilization of ICT in library processes. Gitau has a Master's degree in Library and Information Science from the University of Wales, Aberystwyth, and a degree in Education Science specializing in chemistry and physics from Kenyatta University. In addition, Gitau holds a Certificate in Modern Information Resource Management from DSE-ZED, Berlin/ University of Potsdam, Germany. Gitau was formally the University Librarian at Strathmore University. Gitau has undertaken a professional course on library database management. He has an interest in the field of Library management, research data management, and information retrieval systems. Throughout the years Gitau has been involved in teaching both postgraduate and undergraduate students in several universities. His research interests are in the field of Library management in general but in particular in information retrieval systems, information literacy, knowledge management, open educational resources including open journal publishing, research data management, and digital institutional repositories. Gitau has been actively involved in the training of lecturers and students on scholarly trustworthiness and using information responsibly. Currently running an Eifl-sponsored project on online journal publishing using OJS in 23 universities in Kenya.