Data Literacy: A catalyst for improving research publication productivity of Kyambogo University Academic Staff

Robert Stalone Buwule, Mbarara University of Science and Technology, Mbarara, Uganda, rbuwule@must.ac.ug
Eliz Nassali State, Kyambogo University, Kampala, Uganda
Mukiibi Edward, Kyambogo University, Kampala, Uganda

Abstract

Objective: The aim of this study is to explore how data literacy can influence the research and publications productivity of Kyambogo University academic staff.

Methods: The study employed a literature review to collect detailed information. It observed lessons, and studied patterns of the phenomenon to explore data literacy initiatives that can be used by Kyambogo University academic staff to improve their research publications productivity and also to mitigate the accrued challenges.

Results: The paper explored eight initiatives through which data literacy skills could enhance the research productivity of Kyambogo University academic staff. These were awareness and promoting freedom of using open data, engaging students in data literacy activities, pedagogical reflection, datafication of single and joint academic staff publications, visualization of data, storytelling, ethical use of data, and preservation of research data.

Conclusion: While this paper relies on the context of the Kyambogo University academic staff, the authors posit that these data literacy skills can be embraced by universities in developing economies; especially those struggling with poor research and publications productivity. The paper further identifies areas where universities in developing economies, in conjunction with their libraries, can improve the academic staff pedagogy and compliance to eScience through polishing their data literacies.

Received: November 1, 2022 Accepted: July 17, 2023 Published: November 6, 2023

Competing Interests: The authors declare that they have no competing interests.

© 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See https://creativecommons.org/licenses/by/4.0.
Introduction

Research conducted by the university academic staff and students contributes to national development (Bernanke 2011). Yet, if this research is not well managed, it remains in repositories and library shelves without triggering any societal impact. Kyambogo University (2022) is one of the public universities in Uganda established in 2003 to advance research, knowledge creation, and development of skills in Science, Technology, and Education. It is situated on Banda Hill, in Nakawa Division, East Kampala, and has over 30,000 students and 425 permanent academic staff. Since the university’s inception in March 2003, extensive research has been conducted. But the University’s research publication productivity of academic staff is still low. Yet one of the criteria for appointment and promotion of academic staff is based on production of publications from reputable publishers (Kyambogo University 2014).

Science, Technology, Engineering, Mathematics and Social Sciences (STEMS) research involves the use of research data (Kafel, Creamer, and Martin 2014). It is actually not only in research but as well as teaching, the entire instruction process, publication and promotion of eScience. Data literacy triggers Data Driven Decision Making (DDDM). This is a systematic process of collecting, evaluating, and using data to inform policy formulation and standardize education practice (Mandinach 2012). DDDM improves curriculum and instruction and can lead to greater teacher effectiveness, student equity, and performance (Poortman, Schildkamp, and Lai 2016). Teaching and conducting research in the pure, applied, and social sciences requires the use of data that requires some level of statistical efficiency, causal interpretation, and empirical examples (Angrist and Pischke 2017). Thus, academic staff of any university must be oriented towards developing empirical data literacy skills if they are to remain relevant to their employers (Angrist and Pischke 2017). Extant literature has shown how data literacy has improved research publication productivity. For example, Shrivastava and Gupta (2019) report on how sharing of data improved the visibility of the research publications in India; other scholars have also indicated that data literacy can enhance research publications productivity (Raffaghello and Manca 2023; Ramkumar 2018).

Data literacy can be defined as the knowledge that links research data generated from the research labs and the field all through to research publication and data repositories. Data literacy enables academic staff to apply statistical knowledge while writing a variety of individual, joint, and corporate publications (Carlson et al. 2011). Data literacy is one of the skills that is causing the role of the librarians to evolve. Librarians are expected to offer this service through discussing these skills with academic staff, students, and researchers. Data literacy is discussed in terms of effective instructional content, methods, and formats (Martin 2014).

Integration of Data Literacy among academics

Data literacy (DL) was created from traditional research in the fields of statistics and numeracy (Gould 2017). It was employed in mathematical operations and environments, but later contemporary society adopted it. This followed the widespread use of statistics and numeracy in support of journalism, public policies, business,
library, and many other spheres. DL is a new and an under-researched construct in Uganda and particularly at Kyambogo University. There isn’t much research done or literature available on the topic, though it is emerging at an exponential rate. Data literacy competencies (knowledge and skills) can enable an academic staff to identify, access, analyze, and use open data from internal and external databases (Fotopoulou 2021). Beck et al. (2020) broadens this definition further by stating “Data literacy among academics is the ability to transform information into actionable instructional knowledge and practices by collecting, analyzing, and interpreting all types of data (assessment, behavioral, snapshot and longitudinal)”. It involves understanding what data means, reading graphs and charts, recognizing when data is misused, and finally drawing correct conclusions from data (Carlson et al. 2011). At policy formulation level, ‘data literacy enables academics to understand and use data to inform decisions and influence policy’ (Mandinach and Gummer 2013). Data literacy is diverse and has been understood to be part of assemblages of data science (Jasanoff 2017). Data literacy calls for the understanding of data standards, disciplinary types, practices and knowledge of publishable content.

Wolff et al. (2016) posit that there seems to be no conventional data literacy skills for academic staff. Each institution defines and assesses data literacy according to their context of beginner, intermediate, and advanced level (Hammer 2019). Beginner level is composed of basic data literacy skills, intermediate are the moderate skills, and advanced is for experts in data literacy. Below is a table showing the different data literacy skills for each level.

Table 1: Showing the different levels of Data literacy (Source: Hammer 2019)

<table>
<thead>
<tr>
<th>Beginner Level</th>
<th>Intermediate Level</th>
<th>Advanced Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Find data</td>
<td>• Data management plans</td>
<td>• Set up &amp; govern data repositories</td>
</tr>
<tr>
<td>• Evaluate data</td>
<td>• Curate data</td>
<td>• Visualize data via software</td>
</tr>
<tr>
<td>• Cite data</td>
<td>• Clean data</td>
<td>• Add metadata to datasets</td>
</tr>
<tr>
<td>• Organize data</td>
<td>• Spread sheets competent</td>
<td>• Analyse data via a tool such as R or Tableau</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Data modelling, segmentation and attribution</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Predictive analytics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Data Carpentry instructor</td>
</tr>
</tbody>
</table>

The campaign of requiring all academic staff have data literacy skills can be done based on three phases of the data literacy levels. The first phase is taking a deliberate effort to ensure that all academic staff meet the data literacy skills for beginners. This enables them to train all their students to obtain the same skills. The second phase is scaling up all academic staff interested in research and publication to meet the intermediate
data literacy skills. The third phase is identifying champions and sending them for modular courses to meet the advanced data literacy skills. Later they can come back to train others.

Integration of data literacy in the research culture of world-class universities has gone far in addressing the common research and publication bottlenecks of academic staff (Yu 2019). Data literacy develops a supportive research infrastructure that sustains quality teaching, learning, eScience, and community engagement (Mandinach and Gummer 2016b). Universities have taken a step further to mobilize academic staff members to learn the data literacy skills that have empowered the staff to identify, access, interpret, manipulate, and share the research data and communicate it through publications. In developing countries universities are still struggling to embrace this new concept of data literacy. A study conducted in New Zealand explored the perceptions of lecturers, school leaders, and candidates in the data literacy program. It was revealed that most participants struggled with many of the same aspects of mathematical and statistical literacy. Moreover, some participants reported lacking confidence, motivation, and enjoyment for quantitative data (Cowie and Cooper 2017).

The leading barriers to the adoption of data literacy in research and publication around the world included: reluctance to share research data, lack of knowledge on how to access open data, shortages in capacity building and training of staff members to read and understand data literacy and visualizations (Copeland, Yoon, and Zhang 2021). This is even worse in Sub-Sahara Africa where institutions are less data-driven and rarely incorporate analytical data approaches in the business processes. Few university lecturers have the data literacy skills, yet they interface with an increasing number of students. Further still they have a high workload and find it difficult to find time to pass on these skills to their peers. Staff with data literacy expertise are few within the existing workforce in many universities. There is often little time and funds allocated towards learning data literacy design and analytics. As such, these issues are not given priority (Mittelmeier et al. 2018).

In Uganda, for example at Kyambogo University, a study was carried out on the research productivity of academic staff. This study revealed that from 2003 to 2020, out of the 425 academic staff, only 199 staff had a total of 440 publications (Anonymous 2022). The total number of staff with PhDs by that time was 192. One prolific author had 32 publications followed by another with 16 publications. The Faculty of Science was the most productive with 110 publications. This was followed by the Faculty of Education with 106 publications. It was further noted that 172 (42%) articles were published in predatory journals. Most articles are published in the health discipline which is not the University's niche area nor part of the University's research agenda.

The above anomalies could be addressed through improving the data literacy skills of Kyambogo University academic staff. The University library could also play a pivotal role in advocating and guiding academic
staff to openly share research data. It could also offer user support to staff who were unable to access desired data. However, in this paper, there are eight categories of data literacy skills needed by academic staff with more than average research publications productivity. The paper further aims to highlight how equipping Kyambogo University academic staff and other Universities in developing economies with data literacy skills can improve research publications’ productivity.

**Theoretical underpinnings**

There are many theories that scholars use to explain and describe data literacy concepts. For example, Kuhlthau’s Information Search Process (ISP) (Kuhlthau 2010) which provides tenets for librarians to illustrate ways that librarians can help patrons with data searching, use, and manipulation (Jefferson, et al. 2020). However, this paper is underpinned by the Data Literacy for Teaching (DLFT) framework. DLFT was coined by Mandinach and Gummer (Mandinach and Gummer 2016a) outlines how academics can use research data to guide teaching and learning. Wayman et al. (Wayman et al. 2017) further expands the DLFT framework to include components of data literacy such as: collaboration competency, attitudes, and support. In the context of research and publication, this framework is applicable because these are the very components needed by the academics right from the stage of data collection to research publication. This theory is the critical link that was used to discuss the different sections of the paper. It shows how data literacy can impact academic staff research publications productivity (Obery et al. 2021).

**Methodology**

The study reviewed literature on data literacy and academic staff research publication (Moher, Stewart, and Shekelle 2015). Two minor scientific databases ERIC and Taylor & Francis were searched as they have a broad subject coverage and also provide both restricted and open access journals. Queries used to search included ‘Data Literacy,’ ‘Research Publications,’ ‘academic staff,’ and ‘University’. This search yielded 133 papers from Eric and 317 from Taylor and Francis giving a total of 450 articles. From these, 401 were overlaps and once eliminated, 49 papers were considered for the screening. The researchers read the abstract and excluded 28 papers that were not relevant for analysis. Only 21 papers were considered for review. The researchers read the full text of the selected articles and classified them according to a set of categories. The articles were discussed and deemed relevant to depict the landscape of how data literacy influences academic staff research publications (Tricco et al. 2015).

**Data Literacy skills for Academic staff**

The data literacy skills needed by Kyambogo University academic staff developed here are premised on the DLFT framework. It espouses the component of competency as a key data literacy tenet for impacting research and publication in a higher education context. Wolff et al. (2016) partly concurs with Mandinach and Gummer (2016) where the authors identified four categories of data literate citizens based on the skill...
level and understanding required to use and manipulate research data. While these categories may not be exhaustive, they provide a framework for conveying the varying depths of skill required by academics to engage with data. The four categories were:

i. Academics need skills to read, understand, interpret and critique data

ii. Academics need communication skills to tell stories from data

iii. Academics need data presentations skills that will trigger or influence policy makers to ask and answer real-world questions basing on the presented data and ultimately use data to make the right decisions

iv. Academics as scientists need skills that combine strong technical data skills with in-depth knowledge of the domain of the data

Cowie and Cooper (2017) broadens the above four categories to seven in the context of University academic staff. Maybee and Zilinski (2015) further added one more to make them eight key frameworks for data literacy:

(a) Awareness open data: Understanding data and its role in society

(b) Access: Understanding how to identify, locate and use datasets and databases

(c) Engagement: Evaluating, analyzing, organizing and interpreting existing data for decision-making.

(d) Pedagogical reflection: Single and joint efforts to design data literacy instruments.

(e) Visualization of data: Planning and managing how to transform data into graphical and pictorial forms, including organization and analysis, security protocols for data storage, sharing data, and data-driven documentation;

(f) Story telling: Oral communication, synthesizing, visualizing, and representing data

(g) Ethical Use: Identifying diversified data sources, in particular data from human and social activity, considering the risks and issues implicit in the use of such data

(h) Preservation: Being aware of long-term practices of storing, using and reusing data.’

Academics therefore need to be data literate and competent in all these categories. When one views them through the lens of the DLFT, one sees that category A, B, E, and G above, fall under Competency, category C relates to Collaboration, category H relates to Attitude and category D and F relate to Support. Below is a detailed discourse of how these data literacies can improve the academic staff research and publication productivity.
Awareness of open data

Academics should learn how to identify the different available data sources. They should have a sense of what datasets are available, the different sources such as data aggregators, data portals, government, and non-government web links. Awareness of where to find and access open data is a basic data literacy skill and is essential for academics to be able to harvest and store open data for use and analysis. It is now a popular trend to provide open data and make it accessible, sharable, and understandable without interruption (Pothier and Condon 2020). Actually, most major funding agencies require data sharing as part of receiving funding, so researchers interested in this kind of funding need to comply with these mandates. Awareness of open data sources will equip academics with the ability to interpret, analyze, and process it into articles that can be published in reputable journals and other credible publishing houses. This competency and ability is triggered by knowing sources of open data, identifying key points from it and using it to derive practical solutions for the society (Ridsdale et al. 2015).

Access, use, and manipulation of open data

Access, use, and manipulation of open data lends itself to a wide range of data literacy practices that cannot be limited to only using and interpreting data. It goes further to include, downloading open data, promoting the freedom of using open data, manipulating, and interpreting different data types like graphs and frequencies. All these competencies should empower the academic to engage with the data and interpret in a way they can communicate the data’s meaning to others. Sometimes these skills might go deeper to encompass critical media literacy like algorithmic and data-based media forms. At this level, a data literate academic should be able to generate, work, and partner with agencies to use open educational resources.

Another aspect that lends itself to use and manipulation of open data is data reuse. Data reuse is the secondary use of data to study a new phenomenon (Yoon 2017). Academic staff and their students can explore existing data which can be reused to study similar problems. Such data is retrieved from open databases which are accessible and have machine readable protocols. Governments are opening more access to open data, making it accessible as a way of promoting transparency and open governance (Copeland, Yoon, and Zhang 2021).

Engaging students in data literacy activities

After equipping the academic staff with a range of data literacy skills, University management should further task the academic staff to transfer the basic skills to the respective students. As stipulated in the DLF, the staff could use tool-based capacity-building training approaches which equip students with both basic and advanced practical and technical data literacy skills. Tool based capacity building training approaches should be applied including DataHub workshops with sessions on data cleansing operations, use of data pivot tables, manipulation of data using web tools, visualization, and open data portals among others. Through such
engagements, students can voice their concerns about the challenges of the data-intensive environments of the University. This would allow the staff and University management to create an stronger environment for data collection and management which could culminate in improved research and publication productivity.

**Pedagogical reflection**

The University academic staff association could present an ideological reflection on pedagogical datafication. This reflection inspires academic staff to design the data literacy instruments that outlines the structure and learning outcomes of datafication. The pedagogical reflection could further shape how academic staff think about data literacies and inform how to conceptualize critical data to be included in research publications. All these initiatives will need the support from the University Council, top Management, Senate, and other administrative organs. Single and joint pedagogical reflection of data literacy is a critical literacy skill for academic staff because it enables them to conduct data manipulation, analysis, and visualizations when lecturing students. It therefore goes without saying that pedagogical reflection will enhance learner-centered methods and tools (D'Ignazio 2017).

**Visualization of data**

Data visualization is the ability to graphically represent information and data through visual elements like charts, graphs, and maps. Visualization enhances data such that the users can see and understand trends, identify relationships, outliers, and patterns in data presented in the research publications. Data visualization enhances transparency and makes the data presented more understandable. Visualization allows the academics to enrich presentations with patterns, animations, maps, and other interesting graphics (Fotopoulou 2021). The common tools used for data visualization are Tableau, Federal Reserve Bank of St. Louis (FRBSL/FRED), Geographical Information Systems (GIS), Excel, and infographics (Batt et al. 2020). These tools have enhanced features for developing charts, graphs, maps, and patterns in data. Comments can be generated, observed, and conclusions drawn from presented data. Animated graphics can be derived from the data means, modes, and medians, where worksheets are used to import or export data.

For example, Tableau is an emerging and complementary tool of data analysis and visualization. Owing to the graphics and animations in these tools, visualization is used by both academic staff and students when working with clean data sets.

**Storytelling with data**

Storytelling with data is the process of creating briefs and narratives based on research data. It starts with understanding the audience and the data use to communicate. Spotting common experiences that connect to people, highlighting the scale of an issue that makes the story more powerful. Using the DLFT lens, storytelling with data is a competency of the academic to use data and numbers to support realities, produce thinking patterns, and shape institutional and cultural contexts. Therefore, academic staff need a deeper
understanding of how data can be used for storytelling, advocacy, and campaigning. This is done through identifying leading examples of data that can be used in political campaigns, reporting for the media and journalists and the general arts. Storytelling is very helpful when academics are writing and publishing policy briefs which can later be used to influence and form policy.

Fotopoulou (2021) reports on how workshop speakers data literacy and analytics skills would use storytelling to obtain research funding. Data can be used to tell very interesting stories that easily get the message across because data is objective, legitimate, and trustworthy. This means that storytelling with data is very helpful for academic staff seeking funding for their research and publication projects from donors and development partners. The academic staff need the acumen to select and analyze the right data to include in the story before it is communicated to the donor.

Storytelling with data triggers the emotional engagement of audiences. Emotional engagement is a key dimension of effective written communication which is very critical when researching and publishing for social change (Chan-Olmsted and Wolter 2018). The academic staff can intellectualize data by engaging with the data to ensure that the presentation is emotionally engaging the reader. The moment the link between the data and the public is created then research publications could go a long way in influencing the readers’ attitudes towards a desired societal change. Further still, the same data can be used to tell different stories, and from diverse perspectives. Through storytelling the researcher’s data becomes the whole public’s story (C. S. Beck 2005).

**Ethical use of data**

The ease of access to open databases should not compromise the accountable and responsible use of data. Academics at all times should use other researcher’s data ethically. Not all data or stories need to be told even though the academic staff may have access to a wealth of big data and the skills for analyzing. Some stories are traumatizing and therefore simply too difficult or too painful for audience engagement. Some research findings clash with local social-cultural contexts, so researchers need to be sensitive to all these concerns (Street 2003). Ethical use of data involves observing the right to privacy, acknowledging the source, transparency, avoiding falsification and fabrication of data among other concerns. When data is used ethically it enhances decision making and social justice, for the community and other policy makers who consume the research and publications produced by the academics (Raboin 2020). Van Geel et al. (2017) conducted a study to establish the ethical use of secondary school data and its influence on educators' decision making and found that the students could easily access, use, and interpret the data mainly because they were already oriented on how to use data ethically.
However, sometimes the academic staff might not know how to use data ethically. Copeland et al. (2021) conducted a survey and found that both the respondents and librarians didn't know how to use data ethically, and the librarians also didn't know how to teach the ethical use of data. It is no surprise in today’s world that not only libraries, but most modern organizations, strive to be data centric as this is a valuable and powerful business asset (Pothier and Condon 2020). This therefore requires university library staff to be data literate and taught the ethical use of data. If they now know how to use and manage data ethically, they will be able to train the academics in the ethical use of data to prevent loss of time and money.

**Preservation and long-term data literacy practices**

Skills in preservation and long term data literacy are important for academic staff in all departments of an institution (Klenke et al. 2020). These skills should not only be limited to librarians, data scientists, and data analysists. To meet modern research and publication targets, academics entering the publishing world should be literate in working with data and preserving it for a variety of purposes. Research indicates that more training and preparation are needed in this area (Sapp Nelson 2017).

**Other Technical Requirements**

The paper continues to raise technical requirements that Kyambogo University and other institutions of higher learning should put in place training on the eight data literacy skills discussed above to work successfully. These requirements are also discussed using the DLFT framework which underpins collaboration, competency, positive attitude, and management support:

- The University should show its support for this approach by designing and formulating a policy to govern the data literacy programs and practices on the campus. This policy should be crosscutting to regulate the generated research data (Luo 2022).

- For data literacy to improve the publication productivity of academic staff, the University administration should focus on budgeting, developing, and investing in data literacy infrastructure (Beck, et al. 2020). This infrastructure may include servers, routers, switches, firewalls, and research data management software among others.

- The University must harness three key factors that affect the business. These include human, technological, and financial resources. Building data-centric work force is a critical priority, even if the university has enabled data centric technology. Having staff that have not embraced data literacy is a barrier instead of a catalyst to increasing research publications (Pothier and Condon 2020).
• Building a data-centric work force, the University Library in liaison with the Directorate of Human Resources, should restructure an existing position and create a data librarian or data scientist whose chief role would be to offer instruction in data literacy.

• In tandem with the DLFT, the library schools should collaborate with universities and train data librarians (Martin 2017). Library Schools should play a key role in making data instruction more available and approachable. Progressive library schools are working towards integrating data literacy in the library school curriculum (Martin 2017).

• Universities should come up with a deliberate and intentional arrangement of training academic staff in data literacy so that they can obtain skills in accessing, analyzing, interpreting, sharing, and communicating with research data (Starobin and Upah 2014). The data literacy trainings should be integrated with credible and ethical publishing.

• Kyambogo University should conduct a follow-up study to investigate the efforts taken to put in place a data literacy culture and infrastructure.

Conclusion

This paper provides a practical approach on how academic staff can improve their research publications through learning data literacy skills. The discussion in this paper indicated how the data literacy theoretical and practical aspects can impact the research publication productivity of the university. Academics, together with their instructional programs, don’t only benefit from research and publication, but also benefit by infusing data literacy throughout the curriculum to design student-centered pedagogy. Data literacy helps to influence eScience, policies, and decision making by developing strong and dependable research publications.

Acknowledgements

We would like to express our sincere gratitude to the Kyambogo University Competitive Grants Scheme for granting the authors the opportunity to receive a competitive research grant award in 2020. More appreciation goes to the International Association of Social Sciences Information Service and Technology (IASSIST) which sponsored one of the authors to travel and present this paper at the 2022 IASSIST International Conference in Gothenburg, Sweden.

Competing Interests

The authors declare that they have no competing interests.
References


