Jiang, B. et al. (Eds.) (2025). Proceedings of the 33rd International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

Enhancing Research Data Visibility: Developing an Open Data Portal for an HEI in a Developing Economy

Jonathan Angelo BUENO, Lester Paolo DIAZ, Claudio III SANTIAGO, Zachi Keindri VILLANUEVA & Mary Jane ARCILLA*
De La Salle University Manila, 2401 Taft Ave., Manila, Philippines
*mary.jane.arcilla@dlsu.edu.ph

Abstract: This research project aims to create an open data portal for a higher education institution (HEI) in a developing economy to address the institution's current shortcomings in the visibility, accessibility, and management of research data. Existing platforms like Open Data Philippines and the electronic Freedom of Information (eFOI) portal fall short in addressing academic needs due to usability, proper data validation, and discoverability issues. Designed for resilience and usability, the platform enables the institution's dataset providers to upload, share, and access validated research data. Beyond addressing its institutional needs, the system also supports national and international efforts toward research transparency and data-driven innovation, aligned with Sustainable Development Goals (SDGs) for quality education, innovation, and international partnerships.

Keywords: open data, research visibility, information technology

1. Introduction

Research is changing quickly, and open access and data sharing are becoming increasingly important. Open data is characterized by its accessibility, interoperability, and reusability—meaning it must be freely available online, machine-readable, and accompanied by clear licensing for wide reuse (Terzić & Majstorović, 2019; Davies & Calderon, 2022). Terzić and Majstorović (2019) define open data as free digital information that is legally and technically curated to be reusable, redistributable, and accessible. The need for greater transparency, teamwork, and the possibility of new discoveries are the main forces behind this trend. Ensuring data quality—through accuracy, completeness, consistency, timeliness, and relevance—is essential for sound, data-driven decisions that underpin such collaborative research (Angelopoulos & Pollalis, 2020; Wang et al., 2023).

The higher education institution (HEI) is an IT-focused college with a student capacity of approximately 3,500 across its constituent schools, all of which generate academic research in addition to faculty-led initiatives. The recent establishment of a dedicated Research Office demonstrates alignment with the updated guidelines of the Commission on Higher Education (CHED) to enhance institutional research contributions. Among the initiatives introduced is the development of an open data portal, reflecting the HEI's commitment to cultivating a more rigorous and sustainable research environment.

This chapter examines the prospective advantages and challenges associated with the implementation of such a project. Drawing on stakeholder interviews and an analysis of existing practices, the study seeks to identify key pain points and requirements within the HEI's research ecosystem. The overarching objective is to design an open data portal that is user-centric, fosters collaboration, facilitates efficient data sharing, and maximizes the scholarly impact of the institution's research outputs.

1.1 *Background of the Study*

Open Data (OD), in its basic definition, is free digital data that is legally and technically curated to make it reusable, redistributable, and accessible to the public anytime (Terzić & Majstorović, 2019). The term ‘open data’ was officially coined in 1995 based on a report by an American agency advocating for data sharing and transparency between nations to answer global phenomena (Connor, 2023).

In the Philippines, the government has been steadily embracing OD initiatives throughout the years. In 2011, an Open Government Partnership (OGP) initiative was organized to uphold better governance. Since then, the country has developed platforms such as Open Data Philippines (ODPH), eFOI, and OpenStat. The Philippines was among the first eight OGP pioneers (Alampay, 2020), which aimed to increase government transparency and openness for improved public engagement and societal innovation.

Despite the ongoing efforts of the Philippines to adopt OD initiatives and platforms, there are still challenges that impede the full potential and impact of OD. Key issues include poor search engines (Francey, 2023), poor research data (Zuiderwijk et al., 2020), lack of data quality (Martínez et al., 2022), lack of a centralized repository (Thoegersen & Borlund, 2021), lack of metrics regarding the impact of research data (Cruz & Lee, 2019), and the potential risks of exposing sensitive information associated with ethical and legal compliance (Thoegersen & Borlund, 2021). These gaps highlight the need for academically oriented open data platforms that integrate user-friendly interfaces, robust data management practices, and analytics to better support collaboration, data visibility, and data-driven innovation within Philippine academic institutions.

2. **Methodology**

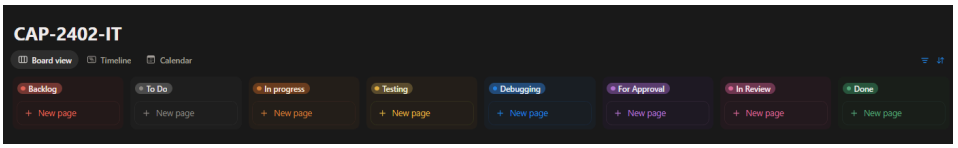

Figure 1. Kanban Board

To address the dynamic needs of the open data portal project, we adopted a hybrid methodology, combining components of Scrum and Kanban—known as Scrumban. This approach allowed for flexibility in managing the evolving requirements of the project while maintaining efficiency in delivery. Scrumban merges Scrum's iterative cycles with Kanban's continuous flow, ensuring that the team could adjust as new needs arose without compromising on the overall pace of progress.

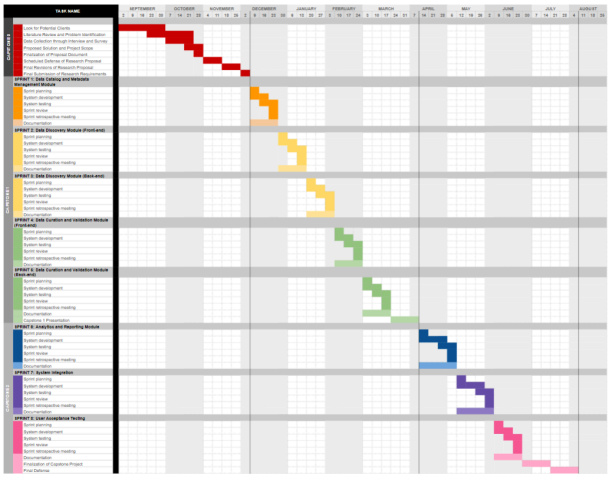

Figure 2. Gantt Chart

The project was structured into eight sprints, each spanning several weeks, and divided into three main phases. The first phase focused on requirement gathering and initial design, with feedback loops involving stakeholders and potential users to ensure the alignment of the system with their needs. Phase two transitioned into system development, where the design specifications from the first phase were translated into a functional product through a series of focused sprints. Each sprint targeted specific system modules, such as the Data Catalog, Metadata Management, and Data Discovery features. The final phase concentrated on refining the system based on user feedback collected during User Acceptance Testing (UAT), which highlighted areas for improvement in usability, functionality, and user experience.

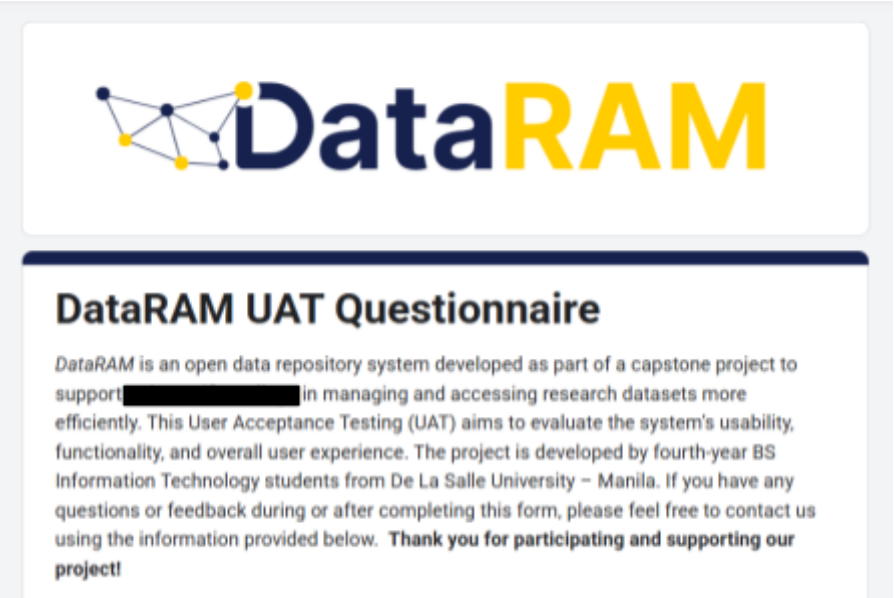

Figure 3. DataRAM UAT Questionnaire

To ensure that the developed system met the necessary user and technical standards, a comprehensive validation strategy was used. This included UAT involving various user groups, such as guests, external users, students, faculty, librarians, and administrators. The feedback gathered from these sessions, through a combination of Likert scale ratings and open-ended questions, was used to make iterative improvements.

In addition, formal interviews were conducted with key stakeholders to gain a deeper understanding of their needs and expectations, while online surveys were distributed to gather quantitative data on user preferences, usage patterns, and common challenges. This approach helped in continuously refining the platform and ensured that the final product was both user-friendly and effective in addressing the institution's data management and sharing needs.

3. Results and Discussion

3.1 *Validation and Testing*

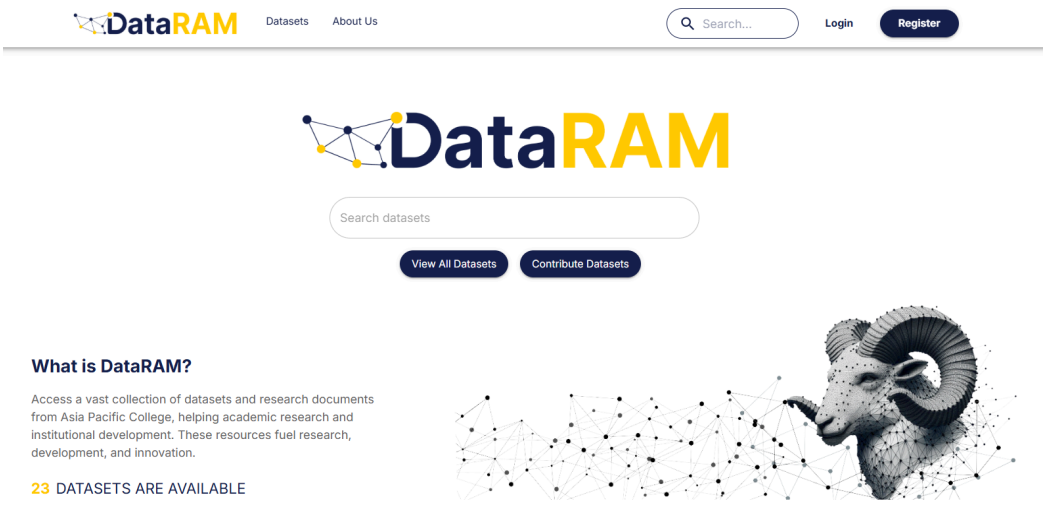

Figure 4. DataRAM Homepage

The User Management module was tested to ensure that the implemented role-based control, notification system, user authentication, user management functionalities, and contacting system work properly across different user scenarios. Test cases include users with different roles (internal users, external users, librarians, and administrators) to verify that they are assigned and restricted access appropriately.

The Data Catalog and Metadata Management module was tested to ensure that the dataset upload feature properly stores datasets in a centralized repository and links them with accurate and complete metadata like their categories, tags, SDGs, contact information, etc., using the metadata input functionality provided by both dataset providers and librarians. In addition, the features of the view dataset, edit dataset, delete dataset, and download were also tested.

The Data Discovery module was tested to ensure its core functionalities are working as intended. This includes searching, data visualization, download options, viewing datasets, the filtering system, the sorting system, search suggestions, dataset requests, bibliographic data tracking, and the contacting system.

The Data Curation and Validation module was tested to ensure that datasets in the system that are contributed by users undergo proper validation and curation before being published on the platform. This includes testing the functionalities such as the dataset upload, role-based access control, FAIR Data Quality Assessment, dataset management, notification system, data validator flagging system, metadata input, and dataset peer reviews.

The Analytics and Reporting module was tested to ensure that the system can generate analytics and reports that are accurate and meaningful. Test cases include data visualization, where key metrics are displayed, such as the total count of datasets, views, downloads, citations, and users. Charts and graphs are checked for responsiveness and accuracy. The system was also tested to confirm that the data behind the visual reports matches the actual datasets in the repository. The Bibliographic Data Tracking feature was tested by retrieving citation information via the OpenCitations API. Through the integration of metadata input, data requests, and categories, the system was tested to confirm that these attributes are correctly generated in the reports, as it helps identify popular dataset topics or monitor request volume per category.

3.2 *Results*

Having conducted a UAT, guest users confirmed that browsing, filtering, and searching felt intuitive with a clear indication of access restrictions. External users found the workflow for requesting private datasets to be straightforward and user-friendly and have reported having a smooth download experience. The students revealed that the datasets' submission

and approval process was clear and dependable. The faculty appreciated the transparency of the review and approval process, especially the visibility of status updates and dataset history. Librarians reviewed the metadata editor, quality-assessment reports, and publishing pipeline. While they provided feedback, their concerns were already addressed, affirming that the system satisfies their operational needs. While the system administrator has acknowledged that the system has the essential controls, they have some suggestions to improve the clarity and usability for future scaling.

4. Conclusion and Recommendations

4.1 *Conclusion*

This research aims to address the issue of having limited research data visibility at an HEI in a developing economy by developing DataRAM, a centralized open data portal. This served as a foundation for the institution to transform how they manage, share, and utilize their datasets, shifting from physically archiving them in their library to using a more accessible, digital repository that enables seamless data sharing and utilization across departments and stakeholders.

The root cause of the problem stems from several issues, including a lack of standardization, limited accessibility, poor data validation, and ineffective infrastructure. In response, DataRAM was developed as a comprehensive, multi-layered validation framework built around the FAIR principles. This includes key features such as automated FAIR Data Quality Assessment, metadata extraction from PDFs, and standardized metadata input to address the standardization gap and improve data visibility. Accessibility was enhanced through a centralized data catalog, advanced search and filtering tools, and download options to datasets. To ensure data quality, the system integrates validator flagging, peer-review processes, and ethical compliance checks that collectively strengthen dataset validation. Real-time data sharing was also accomplished through built-in bibliographic tracking, which records views, downloads, and citation metrics for each dataset. These analytics are displayed using interactive data visualizations, helping users quickly identify popular or highly cited datasets, improving both accessibility and relevance-based search. Lastly, by having role-based access control and API integration, this enables the system to be reliable and scalable.

4.2 *Recommendations*

The relevance and utilization of FAIR assessments can be advanced by calibrating the evaluation criteria to institutional standards, thereby ensuring that the assessments are precise, contextually meaningful, and aligned with the institution's vision towards research excellence. Due to the absence of standardized dataset formats and criteria, dataset validation becomes complicated, and introducing a dual-layered scoring framework is recommended. This will distinguish metadata completeness and content indicators, such as abstracts, keywords, methods, and more, alongside FAIR scoring to provide more nuanced validation.

The integration of AI further augments the system's metadata extraction capabilities. Rather than having restrictions to PDF, it can be extended to various file formats such as CSV, DOCX, and XML. With this, critical fields such as the abstracts, keywords, and conclusions can also be identified and extracted, which will reduce reliance on manual data entry and minimize the risk of human error. Overall, this will improve metadata completeness and enhance dataset discoverability.

Initiate partnerships with other open data repositories or academic institutions to further increase the visibility and impact of academic institutions' research output. Integrating the system with other external platforms via redirected links to their original legal sources, such as Dryad, can facilitate collaboration without violating copyrights and duplicating content.

While the current system was designed as an open data repository, the level of dataset openness is still subject to institutional privacy and ethical concerns. A comprehensive institutional policy must be finalized to clearly define which datasets should remain publicly accessible and which should require restricted access to ensure compliance with ethical and legal standards.

Acknowledgments

The group would like to express their sincere gratitude to Mr. Jojo Castillo, Ms. Jo Anne de la Cuesta, Dr. Elvin Valerio, Ms. Rhea-Luz Valbuena, and Ms. Maylene Mallari for their time, ideas, and valuable suggestions for the research. We also appreciate all UAT respondents, such as faculty members, students, librarians, administrators, and external users, for their valuable feedback and suggestions, which contributed largely to the development of the system.

References

Alampay, E. (2020). Open government data in the time of the pandemic. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3581408

Angelopoulos, M. K., & Pollalis, Y. A. (2020). Use of open data as a tool for successful lean management in public services: evidence from Greece. SPOUDAI-Journal of Economics and Business, 70(1/2), 117-127

Connor, P. (2023, May 16). Open data 101: The history and principles of Open Data – Part 1. Apolitical. https://apolitical.co/solution-articles/en/open-data-101-the-history-and-principles-of-open-data-part-1

Cruz, R. A., & Lee, H. J. (2019). Open governance and duality of technology: The open data designer-user disconnect in the Philippines. JeDEM - eJournal of eDemocracy and Open Government, 11(2), 94–118. https://doi.org/10.29379/jedem.v11i2.545

Davies, T., Calderon, A. (2022). Open Data. In: Harris, P., Bitonti, A., Fleisher, C.S., Binderkrantz, A.S. (eds) The Palgrave Encyclopedia of Interest Groups, Lobbying and Public Affairs. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-44556-0_102

Francey, A. (2023). Drivers of dissatisfaction with an open government data portal: A critical incident technique approach. SpringerLink. https://link.springer.com/chapter/10.1007/978-3-031-41138-0_18

Martínez, R., Pons, C., Rodríguez, R., & Vera, P. (2022). Quality Study of open government data related to COVID-19 in Latin America. Universidad de Antioquia. https://www.redalyc.org/journal/430/43076001003/html/

Terzić, R., & Majstorović, M. (2019). Open data concept, its application and experiences. Vojnotehnicki Glasnik, 67(2), 347–364. https://doi.org/10.5937/vojtehg67-19935

Thoegersen, J., & Borlund, P. (2021). Researcher attitudes toward data sharing in public data repositories: A Meta-evaluation of studies on researcher data sharing. Emerald Insight. https://www.emerald.com/insight/content/doi/10.1108/JD-01-2021-0015/full/html

Wang, J., Liu, Y., Li, P., Lin, Z., Sindakis, S., & Aggarwal, S. (2023). Overview of Data Quality : Examining the dimensions, antecedents, and impacts of data quality. Journal of the Knowledge Economy, 15(1), 1159-1178. https://doi.org/10.1007/s13132-022-01096-6

Zuiderwijk, A., Shinde, R., & Jeng, W. (2020). What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption. PLOS ONE. https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0239283