

What to Curate? Describing the Data Curation Practices of selected Philippine Academic Libraries

¹Maria Pretty Lay T. Abdala, ²Roilingel P. Calilung, ³Venus B. Oruga

¹Bulacan State University

²University of the Assumption

³University of the East

ARTICLE INFO

Received: 01 Aug 2023

Revised: 17 Sept 2023

Accepted: 27 Sept 2023

ABSTRACT

The government, agencies, and scholarly communities have all emphasized the importance of managing research data. A massive amount of data production is causing a data sudden influx. The implications and efficiency of scientific activities had broadened in accessing research data. This event resulting to libraries planning to build their institutional repositories and make this data to be part of. Making data more accessible and reusable is the goal of the data management practice known as “data curation”. Data management techniques include curation, preservation, archiving, and storage layers. Data curation includes the extraction, transformation, and loading of the data. The ability to perform traditional library duties well has given librarians the chance to take the lead as a data curator. However, it is necessary to look into the current data curation methods used by selected Philippine academic libraries and to give a thorough explanation of social and technical factors and connections that influence those practices. This study aims to describe the data curation practices of selected Philippine academic librarians. Quantitative research design, particularly the descriptive method was employed in the study. The research questions were answered using a researcher-made survey instrument. The research instrument was validated by experts in the field of specialization. Academic libraries whose head are officers of the 17 Philippine Librarians Association, Inc. (PLAI) regional councils would be the study’s respondents. Data gathered were analyzed and interpreted using descriptive statistics, primarily frequency distribution, percentage, mean, and standard deviation.

Keywords: Data Curation, Data Curation Practices, Data Management, Challenges

INTRODUCTION

Academic libraries have traditionally been in charge of curating and disseminating data, especially in light of the recent development of institutional repositories. According to Yoon and Donaldson (2019), the optimal approach to implementing data services is to assess researchers' needs, choose a repository environment, and create training materials for both librarians and researchers. Librarians, information professionals, faculty, and researchers work together to curate data. It necessitates the use of a broad range of data-related expertise, starting with research planning and progressing through the phases of long-term management and data reuse for new purposes (Karasti et al., 2006; Weber et al., 2012; Punzalan and Kriesberg, 2017, as cited in Shajitha, 2020).

The government, agencies, and scholarly communities have all emphasized the importance of managing research data. A massive amount of data production is causing a data sudden influx. The implications and efficiency of scientific activities had broadened in accessing research data. This event resulting to libraries planning to build their institutional repositories and make this data be part of. The demand for and interest in data services in libraries is growing rapidly. As a result, academic libraries, are trying to develop these services.

Even though a vast amount of research data is produced by current research activities, there has been a loss of valuable research data due to a lack of knowledge about documentation, metadata rendering, versioning, file format selection, data cleaning, secure storage options, data deposit templates, and making research data easily accessible through research data repositories (T R & Gala, 2021). These are the challenges that the libraries faced while establishing their data repositories.

RESEARCH QUESTIONS

The study aims to describe the data curation practices of selected Philippine academic librarians. Specifically, it answers the following questions:

1. How may the data curation practices of academic librarians be described in terms of:
 - a. Data curation processes
 - b. Data curation services
 - c. Application tools used for data curation processes
2. What are the challenges experienced by librarians in data curation?
3. What measures can possibly be suggested to improve the delivery of data curation services in academic libraries?

LITERATURE REVIEW

Data curation is the continuous and proactive management of data throughout its lifecycle of interest and relevance (Horowitz, 2019). Data curation is a subset of the broader context of scholarly communication, in which research libraries, publishers, and scholars have always served as the primary actors and stakeholders (Palmer et al., 2013). It entails data collection, selection, identifying and correcting erroneous data, heterogeneity treatment, categorization and classification, metadata creation, and preservation. Lee and Stvilla (2017) noted that metadata management is one of the essential components of research data curation.

Data curation is a developing practice in the library and information centers. Libraries are the experts in storing and managing data in repositories. If no value is added to the data stored in the repository, it becomes junk. The curation activity creates value for raw research data in order for it to be processed and reused (Tiwari, 2018). Johnson et al. (2018) found that documentation, chain of custody, and secure storage were the most important data curation activity for researchers, however, they were not satisfied with the result of the activities. Libraries may provide better tools and best practices to researchers in order to improve the effectiveness of these data curation activities.

According to the study of Yoon and Donaldson (2019), academic libraries are currently driving the growth of data curation services. Academic libraries' active participation in data curation was well evidenced in the study conducted. The breadth and depth of data curation services are provided differently across libraries, while library data curation services address the entire life cycle (T R & Gala, 2021; Yoon & Donaldson, 2019). Understanding current data curation practices is essential for delivering effective data curation services and achieving the goals of institutional repositories, which include but are not limited to sharing, accessing, controlling, and preserving knowledge and data. In the study of Masinde et al. (2021), data was curated for value addition, discovery, sharing, and reusing. Furthermore, the library generated metadata for the likely purposes of privacy, consistency, simplicity, and clarity in relationships.

METHOD

The research design, tool, participants, data collection, statistical analysis of the data and ethical considerations are all presented in this section.

Research Design

Quantitative research design, particularly the descriptive method, was employed in the study. Calderon (2006) defined descriptive research as a purposive process of gathering, analyzing, classifying, and tabulating data about prevailing conditions, practices, processes, trends, and cause-effect relationships and then making adequate and

accurate interpretation about such data with or without or sometimes minimal aid of statistical methods. Using descriptive research, the study described the data curation practices of selected Philippine academic libraries with the view of suggesting measures to improve data curation services in academic libraries.

Participants

The participants of the study are academic libraries whose head are officers of the 17 Philippine Librarians Association, Inc. (PLAI) regional councils.

Research Instrument

The study made use of a researcher-made questionnaire as the main data gathering tool. The questionnaire consists of two parts. Part 1 collected information about the data curation practices of academic libraries in terms of data curation process, tools for data curation, and services. Part 2 gathered data on the challenges experienced by academic libraries in data curation.

Validation of the Research Instrument

The questionnaire was validated using face validation process. To improve the questionnaire, validators' suggestions were considered. The Cronbach alpha was computed using JAMOV to assess the questionnaire's internal consistency and reliability after it had undergone pilot testing and been finished. The calculated Cronbach alpha is 0.772, which is good, as seen in Figure 1.

	Cronbach's a	Verbal Interpretation
Scale	0.772	Good

Fig. 1. Scale Reliability Test using JAMOV

Data Collection

After getting the approval from the association President to conduct the study, the validated questionnaire was transformed into a Google form and shared to the 17 PLAI regional councils through the council president and public relations officer.

Data Analysis

Data gathered were analyzed and interpreted using descriptive statistics, primarily frequency distribution, percentage, mean, and standard deviation.

Norm for interpretation. To properly interpret the results of the computation on the degree of awareness, the following norms for interpretation were utilized.

Range Interval	Descriptive Rating/Interpretation
3.25 – 4.00	Highly Aware
2.49 – 3.24	Aware
1.73 – 2.48	Unaware
1.00 – 1.72	Highly Unaware

RESULTS

Table 1. Data Curation Practices of Academic Librarians in Terms of Data Curation Process

Data Curation Process	Mean	SD	Interpretation
Preparation for data curation	3.08	0.759	Aware
Understanding needs of data curation	3.58	0.640	Highly aware
Data receiving and gathering	3.42	0.759	Highly aware
Data cleansing and appraisal	3.25	0.721	Highly aware
Data documentation and metadata	3.42	0.640	Highly aware
Long term data preservation	3.25	0.595	Highly aware
Data publishing in repository	3.33	0.623	Highly aware
Data access and reuse	3.42	0.640	Highly aware
Re-evaluation of data	3.25	0.721	Highly aware
Analyzing the data usage	3.17	0.552	Aware
Continuing practices	3.00	0.577	Aware
TOTAL	3.28	0.657	Aware

Table 1 shows academic librarians' data curation practices in terms of the data curation process. Academic librarians are highly aware of most data curation processes, with a good level of awareness on data curation preparation (mean = 3.08; SD = 0.759), data usage analysis (mean = 3.17; SD = 0.552), and continuing practices (mean = 3.00; SD = .577). Academic librarians are generally aware of the data curation process.

Table 2. Data Curation Practices of Academic Librarians in Terms of Data Curation Tools by Types

Data Curation Tools by Types	Mean	SD	Interpretation
Animation	3.17	0.687	Aware
Data Repository Software (e.g., Hydra, Dataverse, DSpace)	2.12	0.953	Unaware
Metadata Schemas (e.g., Dublin Core (DC), VRA Core, Premis)	3.00	0.816	Aware
Metadata Schemas used in Supplementary Space (e.g., Darwin Core, Geographical Metadata)	2.83	0.687	Aware
Identifier Schemas (e.g., HTTP URI, Permanent local URL)	3.17	0.552	Aware
Controlled Vocabularies (e.g., MeSH, LCHS, RDF Schema, NASA Thesaurus)	3.17	0.687	Aware
Creating and Editing Metadata (e.g., WordPad, Notepad++)	3.17	0.687	Aware

Editing Images and Videos (e.g., SnagIt Photoshop for Images, Handbreak for Audiovisual)	3.00	0.408	Aware
Cleaning Data (e.g., Open Refine, Data Cleaner)	3.00	0.577	Aware
Storing Data (e.g., Dropbox, Google Drive)			
Identifying and Validating Data Files (e.g., DROID, PRONOM)	2.67	0.623	Aware
Transferring Data (e.g., BagIt, WeTransfer)	3.33	0.623	Highly aware
Indexing Data for Searches (e.g., Apache Scholar)	2.42	0.493	Unaware
Tracking and Measuring Data (e.g., Altmetric)	2.75	0.721	Aware
Internet web applications (e.g., EZID service, Google refine)	2.67	0.623	Aware
	2.58	0.640	Aware
	2.83	0.799	Aware
TOTAL	2.84	0.659	Aware

Table 2 depicts academic librarians' data curation practices in terms of the data curation tools they use. Although academic librarians are aware of various data curation application tools (mean = 2.84; SD = 0.659), it has been reported that they are still unaware of data repository software (e.g., Hydra, Dataverse, DSpace) and tools used in identifying and validating data files (e.g., DROID, PRONOM), with computed mean scores of 2.12 (SD = 0.953) and 2.42 (SD = 0.493), respectively. While academic librarians are well-versed in data storage tools (e.g., Dropbox and Google Drive), the top three tools that they are also well-versed in are animation, identifier schemas, and controlled vocabularies, with a computed mean score of 3.17.

Table 3. Data Curation Practices of Academic Librarians in Terms of Data Curation Services

Data Curation Services	Mean	SD	Interpretation
Free consultation and assessment	2.67	0.623	Aware
Extended consultation services	2.58	0.640	Aware
Dataverse set-up services and dataset and data file ingest	2.33	0.471	Unaware
Ongoing dataverse administration and curation services	2.42	0.522	Unaware
Custom services for existing dataverse	2.42	0.522	Unaware
TOTAL	2.48	0.555	Unaware

Table 3 depicts the data curation practices of academic librarians in terms of the data curation services they provide. Academic librarians are unaware of the following data curation services: dataverse set-up services and dataset and data file ingest (mean = 2.33; SD = 0.471), ongoing dataverse administration and curation services (mean = 2.42; SD = 0.522), and custom services for existing dataverse (mean = 2.42; SD = 0.522). However, academic librarians are aware of two data curation services: free consultation and assessment and extended consultation services, with mean scores of 2.67 (SD = 0.623) and 2.58 (SD = 0.640).

Table 4. Challenges Experienced by Academic Librarians in Data Curation

Data Curation Services	F	%	Ranking
Collection, duplication, and inconsistency of data	4	5	4-6
Data management	6	3.33	1
Data privacy	2	10	13
Free data curation tools	3	6.67	7-12
Lack of data curation tools	5	4	2-3
Lack of familiarity on data curation tools	5	4	2-3
Limited budget	4	5	4-6
Limited technological facilities	3	6.67	7-12
Slow connection to cloud	3	6.67	7-12
Staff development	3	6.67	7-12
Unavailability of data	3	6.67	7-12
Unskilled staff	3	6.67	7-12
Unstable Internet connection	4	5	4-6

Table 4 depicts the difficulties that academic librarians face when it comes to data curation. Academic librarians' top three challenges are data management (F = 6), free data curation tools without admin and technical support, and a lack of data curation tools (F = 5) to assist them in data curation.

DISCUSSION AND RECOMMENDATIONS

The management of data over its entire lifecycle, from creation or ingestion through the point at which it is archived or rendered unnecessary and is removed, is known as data curation. Data undergoes several stages of change along the way, and the main goal of its curation is to make sure that data is securely preserved and can be consistently and effectively retrieved.

It is clear from the results that academic librarians are aware of the numerous tasks involved in data curation. They are familiar with several application tools, but they do not however know how to identify and validate data files or use data repository software. The specialized services associated with data curation are not known to academic librarians. The findings are supported by the claims of Tiwari (2018) that there are so many organizations and libraries who provide data curation services, but the data curation is still in the primary stage of development. Heidorn (2011) stated that if libraries do not actively engage in the task, then society may choose to create a new type of institution to curate digital data. This only serves to highlight the significance of data curation and the expertise required to retrieve and validate data before disseminating it to users. The large majority of academic libraries are still having difficulty with data curation or managing their data. Based on the findings, data management tops the list of difficulties faced by academic librarians when it comes to data curation. All disciplines involved in the

administration of data as a valuable resource fall under the umbrella of data management. In providing efficient data curation services and attaining its goals, which include but are not limited to sharing, accessing, regulating, and preserving information and data, having a comprehensive understanding of data curation processes is vital.

As a proposal, academic library organizations may have webinars focusing on data management best practices that can be recorded and accessed later for review. These webinars can also be webcast for broader university audiences. The acquired knowledge can be ensured that data is more accurate, secure, and accurate, which will improve decision-making. Universities may also be able to offer release time and tuition reimbursement for attending a course on data processes in person or online, possibly through a reputable institutional provider.

If one or more of the aforementioned suggestions were implemented, such encouraging gestures may signify a dedication to professional growth, offer additional motivation, and increase data management skills in academic libraries. As a result, these suggestions might contribute to the development of an academic library workforce with "confident data management."

Additionally, it may be fruitful for future research to examine how librarians in various academic library positions provide data curation and management information to library users as well as whether different sizes of these institutions have different training requirements and resource availability. In order to uncover aspects and intricacies of current data curation policies where a lack of knowledge or confusion may exist, the current research study could possibly be enhanced and supplemented with qualitative in-depth interviews with academic library employees.

CONCLUSION:

Although there have been many studies on matters relating to data curation in higher education, few have specifically looked at these issues. The results of this nationwide survey have practical implications for data management in academic libraries in order to successfully address the growing issues of data curation in academic settings. In the first instance, the outcomes can help with determining the workforce training requirements for academic libraries in order to better assist the management and curation of data. The dissemination of data information to user communities also depends on raising knowledge of data policies and enlisting prospective stakeholders across campus to work together on data curation issues. Additionally, these results can stimulate suggestions for how academic libraries can enhance their data curation offerings generally, leading to service enhancements.

It can be concluded that there are many difficulties in data curation, however these difficulties should be accepted because data curation offers more benefits than threats.

ACKNOWLEDGMENTS

We are immensely thankful for the chance provided to us by the International Library and Information Science Society (I-LISS).

We would like to express our sincere gratitude to our family, friends, and colleagues in the profession for their unwavering encouragement and support throughout.

Most importantly, we want to praise and worship the Almighty God who gave us the capacity and wisdom to complete this study.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors if applicable.

About the Author(s)

Maria Pretty Lay T. Abdala, RL, is an Associate Professor from Bulacan State University and also the University Librarian of the said institution. She obtained her Master's in Library and Information Science (MLIS) in 2011. She is also part of the teaching tasks force of the Bachelor of Library and Information Science (BLIS) under the College of Information and Communications Technology (CICT). She is also an active accreditor of the Accrediting Agency of Chartered Colleges and Universities in the Philippines (AACUP). Currently, she is the Regional President 2022-2023 of PLAI-Central Luzon.

A holder of Master of Education in Library and Information Science from Philippine Normal University, Roilingel P. Calilung is presently the Director of Libraries and BLIS Program Chair of the University of the Assumption in Pampanga. Currently, Mr. Calilung is also the Executive Vice President of the Philippine Librarians Association, Inc. (PLAI) and the Vice President of PLAI Central Luzon Region Librarians Council, the former president of the Association of Librarians in Public Sector (ALPS), Inc. and the founding president of the Librarians Association of Metro Pampanga (LAMP).

Venus B. Oruga is the Director of Libraries at the University of the East. She is also a Professor for the Bachelor of Library and Information Science (BLIS) and Master in Library and Information Science (MLIS) programs. She is the Vice President for Luzon of the Philippine Librarians Association, Inc. (PLAI) and the Vice President of PLAI National Capital Region Librarians Council.

Appendix A. An example appendix (if applicable)

Survey instrument : <https://forms.gle/rvKSxgVZcCLRaorg7>

REFERENCES:

- [1] C.S. (2020). Digital curation practices in institutional repositories in South India: a study. *Global Knowledge, Memory and Communication*, 69(8/9). <http://dx.doi.org/10.1108/GKMC-10-2019-0125>
- [2] Heidorn, P. B. (2011). The Emerging Role of Libraries in Data Curation and E-science. *Journal of Library Administration*, 51(7-8), 662-672. doi:10.1080/01930826.2011.601269
- [3] Horowitz, B.M. (2019). Policy issues regarding implementations of cyber attack: resilience solutions for cyber physical systems. In Lawless, W. (Ed.). *Artificial intelligence for the internet of everything*. <https://doi.org/10.1016/C2018-0-00012-2>
- [4] Johnson, L.R., et al. (2018). How Important Are Data Curation Activities to Researchers? Gaps and Opportunities for Academic Libraries. *Journal of Librarianship and Scholarly Communication*, 6(General Issue), eP2198. <https://doi.org/10.7710/2162-3309.2198>
- [5] Lee, D. J., & Stvilia, B. (2017). Practices of research data curation in institutional repositories: A qualitative view from repository staff. *PloS one*, 12(3), e0173987. <https://doi.org/10.1371/journal.pone.0173987>
- [6] Masinde, J. et al. (2021). Research librarians' experiences of research data management activities at an academic library in a developing country. *Data and Information Management*, 5(4). <https://doi.org/10.2478/dim-2021-0002>
- [7] Palmer, C.L. et al. (2013). Foundations of data curation: the pedagogy and practice of "purposeful work" with research data. *Archival Journal*. <https://www.archivejournal.net/essays/foundations-of-data-curation-the-pedagogy-and-practice-of-purposeful-work-with-research-data/>
- [8] Tiwari, A. (2018). Data curation: an opportunity for the libraries. *International Conference on Emerging Trends in Librarianship: Roles of Libraries in Learning Environment*. https://www.researchgate.net/publication/329609112_Data_Curation_An_opportunity_for_the_libraries
- [9] T R, M. & Gala, B. (2021). Data Curation Activities in Research Data Repositories: Best Practices. *International Conference on Statistical Tools and Techniques for Research Data Analysis (ICSTRDA) 2021*, School of Library and Information Science, Central University of Gujarat, Gandhinagar, Gujarat. https://www.researchgate.net/publication/348873527_Data_Curation_Activities_in_Research_Data_Repositories_Best_Practices
- [10] Yoon, A. & Donaldson, D.R. (2019). Library capacity for data curation services: a US national survey. *Library Hi Tech*, 37(4). <https://doi.org/10.1108/LHT-12-2018-0209>