DCPAPERS

DCMI 2026

Proceedings of the International Conference on Dublin Core and Metadata Applications 2026

ISSN: 1939-1366

# Finding Identities : Machine Learning Metadata Analysis on LGBTQ+ Knowledge Representation in the University of the Philippines Diliman Academic Libraries

**Jessie Rose M. Bagunu** [1], **Miriam Charmigrace Q. Salcedo** [1], **Michael D. Amandy** [1], and **Maria Maura S. Tinao** [2]

[1] Library, School of Library and Information Studies, University of the Philippines Diliman, [2] Faculty, School of Library and Information Studies, University of the Philippines Diliman

**Abstract**

This study investigates the gap between the rich conceptual depth of LGBTQ+ knowledge representation and the limited metadata used to describe identities within the Philippine academic landscape. A mixed-methods approach was employed and focused on a corpus of over 1,400 bibliographic records across five library units at the University of the Philippines Diliman. The study concludes that the future of library metadata must evolve from static, rigid filing systems into dynamic and socially responsive maps of human identity. There is an urgent need to move beyond exact-match keyword searches and similarity-based discovery be embrace which allows users to explore the conceptual relationships between diverse works.



## 1. Introduction

Metadata is the primary mediator of visibility and accessibility of repositories in academic libraries. However, it is never a neutral technical stance as it is a reflection of the values, biases, rules and cultural assumptions from the time from which it was created. This study investigates the gap between the rich conceptual depth of LGBTQ+ knowledge representation and the limited metadata used to describe identities within the Philippine academic landscape. In many knowledge repositories, the standardized language used to organize research functions as a mechanism of silence were adopted from classification and descriptors that fail to mirror the vibrant and evolving reality of queer identities. When metadata remains overly generalized or Western-centric, it goes beyond making research material harder to find it effectively erases the identity of the work and the community it represents, creating a barrier to the true discovery and representation of identities.

## 2. Related Literature

This research is grounded in the "invisible substrate" of information science (Bates, 1999) [1], recognizing that library classification is a reflection of institutional power rather than a neutral tool (Berman, 1993) [2]. By drawing on critical cataloging scholarship (Drabinski, 2013) [3], the study interrogates how "descriptive poverty" and outdated terminology marginalize queer identities, often collapsing a diverse spectrum of experiences into monolithic categories. This creates a "metadata silence" that leads to the "symbolic annihilation" of marginalized communities (Adler, 2017) [4].

**Poster**

**Available** Aug 01, 2026

**Correspondence to**
Jessie Rose M. Bagunu
jmbagunu@up.edu.ph

In the Philippines, this gap is evidenced by the barriers to representing Filipiniana LGBT materials within university collections (Inciong, 2014) [5]. To counter this, the study aligns with "reparative cataloging" efforts and the use of inclusive, linked data vocabularies like the Homosaurus to enhance discoverability (Kunesh & Romines, 2025; Ogletree et al., 2015) [6], [7]. By integrating Natural Language Processing (Greenberg, 2010) [8] through a usable metadata framework (Crystal & Greenberg, 2005) [9], this research proposes a shift from rigid, top-down taxonomies to a fluid, bottom-up representation of identity that restores the "power of names" to the communities being described (UBC Researchers, 2024) [10].

## 3. Methodology

This research employed a mixed-methods approach to bridge the gap between traditional technical cataloging and contemporary user-centered information retrieval. The study focused on a corpus of over 1,400 bibliographic records across five library units at the University of the Philippines Diliman.

Central to the methodology was a three-stage computational pipeline designed to analyze and enrich these records. The first stage, Preprocessing, utilized Rapid Automatic Keyword Extraction (RAKE) and Term Frequency-Inverse Document Frequency (TF-IDF) to generate high-impact tags directly from titles and abstracts, capturing the natural language of the authors. The second stage, Exploration, applied K-Means clustering to group these records based on their semantic content, visually mapping the disconnect between existing "library silos" and the actual thematic intersections of the research. The final stage, Bridge-building, employed Cosine

Similarity to develop a semantic-based search prototype, enabling a system that suggests related materials based on conceptual proximity rather than shared classification codes. This quantitative pipeline was complemented by qualitative inquiry, including interviews with librarians and user feedback, ensuring the findings remained grounded in the practical realities of library service

## 4. Findings and Discussions

Through this deep computational analysis, the research revealed a profound disconnect between the richness of academic content and the poverty of its metadata. The analysis uncovered a pervasive silence where the vast majority of records lacked the specific descriptors necessary for modern discovery. While the physical collections contain a wealth of information on human rights, local politics, and intersectional health, these nuanced themes are frequently collapsed under the single, monolithic subject heading of "Gender."

By applying machine learning techniques, the study was able to uncover latent conceptual structures hidden within the library's silos. This process demonstrated that AI can effectively reveal thematic connections across disparate departments, transforming the

collection into an interconnected and visible digital knowledge system that reflects the actual complexity of the research.

## 5. Conclusion

The study concludes that the future of library metadata must evolve from static, rigid filing systems into dynamic and socially responsive maps of human identity. To achieve this, a hybrid approach is essential—one that harnesses the scalable power of machine learning to enrich records while maintaining the ethical oversight and cultural empathy of human librarians. There is an urgent need to move beyond exact-match keyword searches and embrace similarity-based discovery that allows users to explore the conceptual relationships between diverse works. By integrating localized vocabularies and community-informed descriptors into the metadata pipeline, knowledge repositories can transform their catalogs from mere inventories into instruments of social inclusion. Finally, by reimagining how we find identities, we ensure that the stories of marginalized communities are not just safely stored away, but are actively found, recognized, and celebrated.

## 6. About the Authors

**Jessie Rose M. Bagunu** is a College Librarian at the University Library, University of the Philippines - Diliman and currently is the Head Librarian at the University of the Philippines School of Library and Information Studies Library or #UPSLISLib. Her research interest includes information needs and behavior, personal learning networks, metadata and classification as well as on the well-being of senior librarians and members of the elderly population, which she actively supports and advocates on their behalf.

**Miriam Charmigrace Q. Salcedo** is the Reference Instruction Librarian at the University of the Philippines Diliman's School of Library and Information Studies (SLIS) Library. Her work centers on bridging the gap between historical preservation and modern accessibility. Through her research in social media, digital democratization, and user engagement, she leads initiatives that transform library services. Committed to community outreach, she is a passionate advocate for equitable information access, ensuring that valuable resources remain available and accessible to the public.

**Michael D. Amandy** is a library staff member at the University of the Philippines School of Library and Information Studies (UP SLIS) Library and an LIS student at Manuel S. Enverga University Foundation. Passionate about information management and queer studies, he is currently working toward becoming a licensed professional librarian in the Philippines.

**Maria Maura S. Tinao** has been with the University of the Philippines since 1993. She is a full-time faculty at the School of Library and Information Studies (SLIS) at the University of the Philippines Diliman. A certified IBM Data Science and Data Analyst Professional, her expertise lies in data science, machine learning, predictive analytics, and data analysis. Her research spans breast cancer detection, sarcopenia risk predic-

tion, and AI adoption among the aging population, demonstrating her commitment to projects that bridge technology and meaningful societal impact.

## References

[1] M. J. Bates, "The invisible substrate of information science," *J. Am. Soc. Inf. Sci.*, vol. 50, no. 12, p. 1048, 1999.

[2] S. Berman, *Prejudices and Antipathies: A Tract on the LC Subject Heads Concerning People*. Scarecrow Press, 1971.

[3] E. Drabinski, "Queering the Catalog: Queer Theory and the Politics of Correction," *The Library Quarterly*, vol. 83, no. 2, pp. 94–111, 2013.

[4] M. Adler, *Cruising the Library: Perversities in the Organization of Knowledge*. Fordham University Press, 2017.

[5] J. B. C. Inciong, "Representation of Filipiniana LGBT books in selected university libraries in Metro Manila," University of the Philippines Diliman, 2014.

[6] T. Kunesh and J. Romines, "Changing the subject: The Homosaurus in Emory University's library catalogue," *PINES Bibliographic Services Specialist*, 2025.

[7] A. Ogletree *et al.*, *RDA Overview and Metadata Activities*. in DC-2015: Metadata and Ubiquitous Access to Culture, Science, and Digital Humanities: Proceedings of the International Conference on Dublin Core and Metadata Applications. 2015.

[8] J. Greenberg, "Metadata and Digital Information," in *Encyclopedia of Library and Information Science*, 3rd ed., Marcel Dekker, Inc., 2010, pp. 3610–3623. doi: 10.1081/E-ELIS3-120044415.

[9] A. Crystal and J. Greenberg, "Usability of a metadata creation application for resource authors," *Library and Information Science*, vol. 27, pp. 177–189, 2005.

[10] e. a. UBC Researchers (Watson B., "The Power of Names: Exploring Queer Cataloguing Conventions in Libraries," UBC School of Information, 2024.

[11] The Queer Metadata Collective *et al.*, "Best Practices for Queer Metadata (1.0)," Zenodo, 2024. doi: 10.5281/zenodo.12580531.