Implementing Persistent Identifier Infrastructure for Effective Management of ETD Repositories

A Case Study from Chartered Institute of Personnel Management, Sri Lanka

Kamani Perera[1], Ravindu Sachintha[2], Anushka Earskin[3] and Indika Wijayasriwardana[4]

[1,2,3,4]Chartered Institute of Personnel Management (CIPM), Sri Lanka

Author Note

Correspondence concerning this article should be addressed to Kamani Perera, Chartered Institute of Personnel Management (CIPM), No. 43, HR House, Vijaya Kumaranathunga Mawatha, Colombo 5, Sri Lanka. Email: kamani@cipmlk.org

Abstract

Electronic Theses and Dissertations (ETDs) play a crucial role in scholarly communication by providing valuable insights into academic research and contributing to knowledge advancement across various fields. However, the management and accessibility of ETDs pose significant challenges for academic institutions, libraries, and researchers. This study emphasizes the importance of implementing a strong infrastructure for persistent identifiers (PIDs) to ensure the effective management of ETDs. PIDs uniquely and unambiguously identify digital objects, enhancing their long-term discoverability, accessibility, and citability. This research explores the role of PIDs in improving ETD discoverability, ensuring data integrity, and fostering collaboration within the scholarly community. By reviewing existing literature and evaluating various PID systems such as DOI (Digital Object Identifier) and ARK (Archival Resource Key), the study identifies best practices for implementing PID systems tailored to ETD repositories. The results demonstrate that implementing PID infrastructure significantly enhances ETD discoverability and long-term preservation, promotes interdisciplinary research, and streamlines the citation process. Ultimately, the adoption of PIDs supports the broader dissemination of knowledge and fosters a more interconnected academic ecosystem.

Keywords: Chartered Institute of Personnel Management (CIPM), Electronic Theses and Dissertations (ETDs), Open Scholarly Infrastructure, Persistent Identifiers (PIDs), Sri Lanka

Implementing Persistent Identifier Infrastructure for Effective Management of ETD Repositories

A Case Study from Chartered Institute of Personnel Management, Sri Lanka

Electronic Theses and Dissertations (ETDs) have become integral components of scholarly communication, offering valuable insights into academic research and contributing to the advancement of knowledge across diverse fields. As the digital landscape continues to evolve, the management and accessibility of ETDs present both opportunities and challenges for academic institutions, libraries, and researchers alike. In this context, it is important to establish a vibrant infrastructure for persistent identifiers (PIDs) to proper management of ETDs. Persistent identifiers are essential tools that enable the unique and unambiguous identification of digital objects, ensuring their long-term discoverability, accessibility, and citability. In the context of ETD repositories, implementing a comprehensive PID infrastructure is paramount to streamline access, enhance interoperability, and facilitate scholarly communication. In this research study, it is described the significance of implementing a persistent identifier infrastructure for the effective management of ETD repositories. By examining the role of PIDs in improving discoverability, ensuring integrity, and fostering collaboration within the scholarly community, this study seeks to provide insights into best practices and strategies for implementing PID systems tailored to the unique needs of ETD repositories.

Problem Statement

It is long felt need that CIPM needs to implement ETD repository to share, reuse and preserve their ETDs along with other research documents in digital platform and provide PIDs to provide global recognition to this treasured wealth of knowledge in the field of human resource management (HRM) in Sri Lanka and beyond. Being a nations leader in HRM, CIPM faced many challenges when conducting national level research projects, international symposiums, publishing bi-annual journal and conducting

graduate level professional courses including a research component. Thus, the institute has taken a firm decision to develop such digital platform to upload their research publications and theses for present and future use as well as long term preservation of their invaluable scholarly assets.

Objectives

- To enhance discoverability
- To improve accessibility and long-term preservation
- To foster interoperability and integration among different ETD repositories
- To enable accurate and reliable citation of ETDs

Benefits for PIDs

Benefits that can accrue by implementation of PIDs (Persistent Identifiers) such as DOIs (Digital Object Identifier), ORCID IDs (Open Researcher and Contributor Identifier), and ROR IDs (Research Organization Registry) are vast for any academic or professional education organization. PIDs help to identify outputs, contributors and organizations in the research network. At this juncture, DOI is assigned to identify outputs, ORCID ID for contributors and ROR ID is for organizations. In CIPM, it is planned to obtain DOIs for their research products such ETDs, Monographs, symposium articles and journal articles while ORCID IDs have obtained by the individual researchers/contributors and ROR ID has already taken for CIPM. Research data can be easily findable, accessible, interoperable and reusable (FAIR) by integrating PIDs.

Examples are given below -

ROR ID for CIPM

Name of organization: Chartered Institute of Personnel Management Sri Lanka (Inc.)

ROR ID: https://ror.org/05g7w4342

In this context, it is crystal clear that PIDs are essential component for linking research outputs, discovery and access and thereby facilitate easy citation. It helps to verify data accuracy, ensure data can be shared and reused without transformations, facilitating integration from multiple sources and enabling new discoveries (Kim, 2023).

Open Scholarly Infrastructure

PIDs can be considered as a major component in scholarly systems, particularly within open scholarly infrastructure. Further, it makes the reliable link among people, places, and things on the web, creating extensive 'PID graphs' that map connections between entities in scholarly landscapes (Cousijn et al., 2021). Moreover, PIDs make objects/articles discoverable, accessible, usable, interoperable, and assessable and that helps to enhance the identification, discovery, and retrieval of scholarly work, presenting many openings for revolution in scholarly communications. In brief, PIDs ensure entities are easy to find by giving them unique and reliable identifiers. They remain accessible because PIDs direct users to the entity even if its location changes. PIDs also enhance usability by referencing specific versions or states of an entity. They promote interoperability by establishing trust through transparency and provenance. Lastly, PIDs make entities assessable by linking them within a connected network, allowing for easier evaluation and analysis (CERN, 2020). These features offer many opportunities for innovation in scholarly communications, particularly in identifying, discovering, and retrieving scholarly entities (Ananthakrishnan et al., 2020). The possibility of PIDs to change future open research ecosystems is significant that more and more research organizations are demanding their use whenever possible (Bosman et al., 2021). The implementation and use of PIDs in scholarly systems are primarily managed by the system administrators, and most scholars interact with these features only inactively.

However, many users also play a crucial role in contributing to the PID data that make up the PID graph, often being the original sources of this data (Dappert et al., 2017).

Scholarly Awareness of PIDs

PIDs are mainly used for citing purposes and its accuracy is depend on reproduction by others. It is evident that there is a minimal interaction with PIDs. Thus, indirect research areas should be considered, particularly the capacity of scholars' who discover work by recognizing scholarly objects, and their capability and citation behaviors (Macgregor, et al., 2023). According to the existing literature, in the initial stages of bibliometrics, it is found recurrent and frequent errors in citation in academic articles, presented by authors during drafting. Journals were not properly identified, incorrect author names and name mis-orderings etc. were common (Garfield, 1974, 1990). There were mistakes connected with a lack of citation confirmation (e.g. faulty citations are copied from a faulty source by authors) (Broadus, 1983) and this has been called 'referencing misbehavior' (Liang et al., 2014). Moreover, there are instances that disparities in quality of citing and the pattern of conduct across specific subjects by scholars, deepening the problem (dos Santos et al., 2022).

The existing literature revealed that 40% of all experimented citations were inaccurately cited (Key & Roland, 1977), 48% of all articles were containing one or more errors, the 70% of errors found on article titles and author names (Asano et al., 1995). Similarly, Logan (2022) pointed out that overall citation error rate of 40% were from renowned journal titles (between 1991 and 2019). In this context, it is disclosed that scholars incapability of citing scholarly sources accurately.

It is evident that lack of Digital Literacy, Web Literacy (Alexander et al., 2017) and Meta Literacy (Mackey & Jacobson, 2017) capabilities were common between few scholarly groups (Ong, 2021). This has proved by Basilotta-Gómez-Pablos et al. (2022) by analyzing 56 articles to study the digital competency and found that **"low or medium–low"** digital skill was leading. In the same vein, Judd

(2018) stated that though the young researchers were ‘digital natives’ and likely to be digitally capable or digitally knowledgeable, that concept has been broadly disproved by current evidence. Similarly, this has been proved by Greer and McCann (2018) conducting an experiment among final year undergraduates and found that they “do not understand URLs”, “official and unofficial URLs” and failure to identify digital sources on the web. Thus, it is important to carry forward Digital Literacy, Web Literacy and Meta Literacy from undergraduate level to doctoral level or beyond that to maintain the PID competencies.

Methodology

A comprehensive review of existing literature, scholarly articles, reports, and best practices related to persistent identifiers, electronic theses and dissertations, repository management, and digital preservation were conducted. This review provides a foundational understanding of the current state-of-the-art, challenges, and opportunities in implementing persistent identifier infrastructure for ETD repositories. Various persistent identifier systems and frameworks such as DOI (Digital Object Identifier), ARK (Archival Resource Key), were evaluated to identify the most suitable system for the repository.

Results and Conclusion

Implementation of persistent identifier infrastructure led to a significant increase in the discoverability of ETDs within repository systems and external databases. By assigning unique identifiers to each document, ETDs become more easily searchable and identifiable, thereby improving their visibility among scholarly community. Further, it ensured the long-term accessibility and preservation of ETDs. Stable links provided by persistent identifiers enabled continued access to ETDs over time, even as technologies evolved and platforms changed. In this context, ETDs remained accessible to future generations. It promoted collaboration and knowledge sharing across institutional boundaries, facilitating interdisciplinary research and fostering a more interconnected scholarly ecosystem.

Persistent identifiers streamlined the citation process for ETDs, promoting accuracy, reliability, and proper attribution to authors. Stable references provided by persistent identifiers were easily included in scholarly publications, citations, and bibliographies, supporting transparency, reproducibility, and scholarly communication. At this juncture, attention is made to maintain the accuracy of the citation when uploading ETDs and other research documents to the CIPM Research Nexus. Moreover, it is planned to conduct regular PID Literacy programs to make aware the scholarly community on the importance of PIDs and how to maintain the accuracy of citations.

Conclusion

The implementation of persistent identifier infrastructure has proven to be instrumental in enhancing the management of ETD repositories. By providing unique PIDs for ETDs, this infrastructure has improved the discoverability, accessibility, interoperability, and citability of scholarly works, thereby facilitating broader dissemination of knowledge and fostering collaboration within the academic community. PIDs support the long-term preservation and accessibility of valuable research outputs for the benefit of future generations.

References

Alexander, B., Becker, S. A., Cummins, M., & Giesinger, C. H. (2017). *Digital Literacy in Higher Education, Part II: An NMC Horizon Project Strategic Brief* (pp. 1–37). The New Media Consortium. https://www.learntechlib.org/p/182086/

Ananthakrishnan, R., Chard, K., D'Arcy, M., Foster, I., Kesselman, C., McCollam, B., Pruyne, J., Rocca-Serra, P., Schuler, R., & Wagner, R. (2020). An Open Ecosystem for Pervasive Use of Persistent Identifiers. *Practice and Experience in Advanced Research Computing*, 99–105. https://doi.org/10.1145/3311790.3396660

Basilotta-Gómez-Pablos, V., Matarranz, M., Casado-Aranda, L.-A., & Otto, A. (2022). Teachers' digital competencies in higher education: A systematic literature review. *International Journal of Educational Technology in Higher Education*, *19*, 8. https://doi.org/10.1186/s41239-021-00312-8

Bosman, J., Frantsvåg, J. E., Kramer, B., Langlais, P.-C., & Proudman, V. (2021). The OA Diamond Journals Study. Part 1: 42 Findings. *Zenodo*. https://doi.org/10.5281/zenodo.4558704

Broadus, R. N. (1983). An investigation of the validity of bibliographic citations. *Journal of the American Society for Information Science*, *34*(2), 132–135. https://doi.org/10.1002/asi.4630340206

CERN. (2020). *Why use persistent identifiers?* https://web.archive.org/web/20221001165705/https://sis.web.cern.ch/submit-and-publish/persistent-identifiers/why-pids

Cousijn, H., Braukmann, R., Fenner, M., Ferguson, C., van Horik, R., Lammey, R., Meadows, A., & Lambert, S. (2021). Connected Research: The Potential of the PID Graph. *Patterns*, *2*(1), 100180. https://doi.org/10.1016/j.patter.2020.100180

Dappert, A., Farquhar, A., Kotarski, R., & Hewlett, K. (2017). Connecting the Persistent Identifier Ecosystem: Building the Technical and Human Infrastructure for Open Research. *Data Science Journal*, *16*(0), Article 0. https://doi.org/10.5334/dsj-2017-028

dos Santos, E. A., Peroni, S., & Mucheroni, M. L. (2022). The Way We Cite: Common Metadata Used Across Disciplines for Defining Bibliographic References. In G. Silvello, O. Corcho, P. Manghi, G. M. Di Nunzio, K. Golub, N. Ferro, & A. Poggi (Eds.), *Linking Theory and Practice of Digital Libraries* (pp. 120–132). Springer International Publishing. https://doi.org/10.1007/978-3-031-16802-4_10

Greer, K., & McCann, S. (2018). Everything Online is a Website: Information Format Confusion in Student Citation 99 Behaviors. *Communications in Information Literacy*, *12*(2), 151-165. https://doi.org/10.15760/comminfolit.2018.12.2.6

Kim, S. (2023). Introduction to PIDs: What they are and how to use them. *Washington University in St. Louis*. Retrieved from https://becker.wustl.edu/news/introduction-to-pids-what-they-are-and-how-to-use-them/

Liang, L., Zhong, Z., & Rousseau, R. (2014). Scientists' referencing (mis)behavior revealed by the dissemination network of *Measuring the concept of PID literacy* referencing errors. *Scientometrics*, *101*(3), 1973–1986. https://doi.org/10.1007/s11192-014-1275-x

Logan, S. W. (2022). Reference Accuracy in Research Quarterly for Exercise and Sport: A 30-Year Follow-Up to Stull et al. (1991). *Research Quarterly for Exercise and Sport*, *93*(2), 401–411. https://doi.org/10.1080/02701367.2020.1853019

Macgregor, G., Lancho-Barrantes, B. S., & Pennington, Diane Rasmussen (2023). Measuring the concept of PID literacy: User perceptions and understanding of PIDs in support of open scholarly infrastructure. *Open Information Science*, 7(1). http://dx.doi.org/10.1515/opis-2022-0142

Mackey, T. P., & Jacobson, T. E. (2017). Reframing Information Literacy as a Meta literacy. *College & Research Libraries*, *72*(1), 62–78. https://doi.org/10.5860/crl-76r1

Ong, L.T. (2021). Information technology literacy: The crucial factor in aged second-career academics' sustainability. *SHS* 205 *Web of Conferences*, *124* (5), 06005. https://doi.org/10.1051/shsconf/202112406005