

Preserving the Enggano Language: A Digital Dictionary Approach

Research-in-progress

Cokorda Pramatha

Centre for Interdisciplinary Research on the Humanities and Social Sciences
Udayana University
Indonesia
cokorda@unud.ac.id

Gede Primahadi Wijaya Rajeg

Faculty Of Linguistics, Philology, And Phonetics
University of Oxford
United Kingdom
gede.rajeg@ling-phil.ox.ac.uk

Mary Dalrymple

Faculty Of Linguistics, Philology, And Phonetics
University of Oxford
United Kingdom
mary.dalrymple@ling-phil.ox.ac.uk

I Wayan Arka

School of Culture, History and Language, ANU College of Asia and the Pacific
The Australian National University
Australia
wayan.arka@anu.edu.au

Abstract

The Enggano language, spoken by approximately 1,500 people on Enggano Island in Indonesia, is at risk of extinction due to a shift towards Indonesian and a lack of formal educational use. This research-in-progress paper presents the development of an online digital dictionary designed to preserve and revitalize the Enggano language. Utilizing the Design Science Research Methodology (DSRM), the project engages the local community to ensure relevance and accuracy. The digital dictionary provides a comprehensive resource for both native speakers and researchers, incorporating modern linguistic tools and a participatory design approach. Future work includes community demonstrations, feedback collection, and integration of contemporary language usage, fostering a sustainable platform for active language preservation. This project contributes to broader efforts in endangered language revitalization by leveraging digital technologies to support cultural heritage preservation.

Keywords enggano language, digital dictionary, language preservation, community collaboration, lexical resources

1 Introduction

Indonesia has a vibrant linguistic landscape with over 652 local languages spoken alongside Bahasa Indonesia, the national language, making most Indonesians multilingual (Wongso Adi 2022). In 2016, UNESCO reported that 139 of these languages were at risk of extinction, representing nearly 17% of all languages spoken in the country (Yvonne 2022).

The Enggano language, an Austronesian language, is spoken by about 1,500 people in Indonesia. Enggano Island, part of the Barrier Islands chain, lies south of the Mentawai, Nias, and Simeulue islands off the west coast of Sumatra. Despite the threat to its survival with an increasing number of speakers shifting to Indonesian, Enggano remains actively used in the central villages of Meok, Apoho, and Malakoni (Hemmings and Arka 2023). This threat is exacerbated by the lack of formal and educational use of Enggano, where Bahasa Indonesia predominates, even in religious practices under Protestant Christianity (Jamilah 2022). This situation is particularly evident in Enggano schools, which do not use the local language and rely heavily on textbooks devoid of local linguistic or cultural content. This reliance hampers the promotion and integration of Enggano among the younger generation (Muhammad Dian Saputra 2022).

Digital dictionaries play a crucial role in preserving endangered languages by providing accessible, editable and comprehensive lexical resources that can be utilized by both native speakers and researchers (Pramartha et al. 2023). For instance, the Ve' rdd system allows community members to re-evaluate and edit grassroots dictionaries, integrating their input into a finite-state language description, which is particularly beneficial for languages like Skolt Sami that are seriously endangered (Khalid et al. 2020). Similarly, the evolution of the Nxa?amxcin lexical resource project into an online database highlights the importance of using standards such as the fifth major release of the Text Encoding Initiative (TEI P5) for data encoding and archiving, ensuring that the resources are portable and interoperable with other linguistic databases (Ewa et al. 2014). The Pangloss Collection further exemplifies how digital archives can provide free access to annotated recordings of spontaneous speech in endangered languages, making these resources available for both academic research and community use (Boyd et al. 2014). Additionally, the development of mathematical models for language coexistence, such as those in the bilingual competition model, highlights the theoretical underpinnings that can inform policies and strategies for language preservation, including the creation and maintenance of digital dictionaries (Wenqi 2023). Finally, the integration of documentation and revitalization efforts, as suggested by Sapién and Hirata-Edds, underscores the need for collaborative projects that support both the preservation and active use of endangered languages, with digital dictionaries serving as a foundational tool in these efforts (Racquel-María and Tracy 2019). By leveraging these digital resources, communities can ensure the survival and continued use of their languages, thereby preserving their cultural heritage.

The Indonesia National Agency for Language Development and Cultivation (*Badan Bahasa*) has been at the forefront of language preservation, publishing over 100 dictionaries since 1977 and transitioning to digital platforms with initiatives like the *Kamus Besar Bahasa Indonesia* (KBBI) and the *Aplikasi Kompilasi Kamus* (AKK) to accommodate local, specialized, and bilingual dictionaries (Selly Rizki 2023). The urgency of such projects is underscored by the endangered status of many Indonesian languages, as highlighted by the development of the first-ever parallel resource for 10 low-resource languages in Indonesia, which includes datasets, benchmarks, and lexicons to support Natural Language Processing (NLP) research and applications (Winata et al. 2022). Specific efforts to preserve languages like Muyu through an Android-based Muyu-Indonesian-English dictionary application demonstrate the practical benefits of digital dictionaries in facilitating language learning and preservation among younger generations (Beatus et al. 2021). The effectiveness of electronic dictionaries in educational settings is further supported by research showing significant improvements in students' vocabulary acquisition when using these tools (Sarmila et al. 2023). Additionally, the creation of a lexicon-based dictionary containing abusive words from major Indonesian tribes for hate speech detection on social media illustrates the broader applications of digital dictionaries in addressing contemporary social issues (Mardhiya et al. 2020). Collectively, these projects not only aim to preserve Indonesia's rich linguistic heritage but also leverage technology to enhance language learning, support NLP research, and address societal challenges.

Given the vibrant linguistic landscape of Indonesia, with over 652 local languages, and the threat of extinction facing 139 of these, including the Enggano language, there is an urgent need for an online digital dictionary dedicated to Enggano. Spoken by approximately 1,500 people on Enggano Island, this Austronesian language is increasingly endangered due to a shift towards Indonesian and the lack of formal educational use. The implementation of a digital dictionary for Enggano would provide

accessible, editable, and comprehensive lexical resources crucial for both native speakers and researchers. Such a tool would facilitate language learning and preservation among younger generations and help maintain the cultural heritage of the Enggano community. UNESCO has underscored the importance of digital preservation for endangered languages, specifically highlighting the intersection of technology and culture in initiatives like this one. These efforts are crucial for sustaining cultural heritage through modern technological tools (UNESCO (n.d)).

2 Research Methodology

Design Science Research Methodology (DSRM) is a systematic approach used to create innovative solutions for real-world problems that involves developing artifacts through a series of stages, such as understanding the problem area, developing a solution, reflecting on the solution, and formalizing it, often through iterative processes involving stakeholders (Delpont et al. 2024; Prammartha et al. 2017). The methodology aims to address the complexity inherent in problem and solution spaces by decomposing projects into smaller, coherent parts known as "echelons," which follow a hierarchical organizing logic (Tuunanen et al. 2024). DSRM is grounded in a design science paradigm, combining elements of design science research and action research to ensure that the solutions are theoretically founded and practically applicable (Buys 2024). The methodology is instrumental in developing artifacts such as managerial costing systems models and digital maturity measurement services, contributing to advancements in various fields.

In addition to the DSRM, the development of the Enggano Digital Dictionary places significant emphasis on community involvement during the data collection process. This includes direct collaboration with native Enggano speakers through structured workshops, interviews, and surveys. The community members contribute lexical entries and provide contextual examples, ensuring the dictionary reflects both traditional and contemporary language usage. The participatory design approach also enables iterative feedback loops, where community suggestions are integrated into successive versions of the dictionary. By incorporating digital and in-person engagement strategies, the project ensures a comprehensive and inclusive approach to language preservation, fostering long-term community ownership of the resource.

The provided image (Figure 1) illustrates the DSRM in the context of developing a digital dictionary artifact for the Enggano language. The process begins with the identification of the problem and the motivation behind addressing it. This step is followed by formulating clear objectives for the solution, defining what the digital dictionary should achieve. Central to the methodology is the design and development initiation phase, which sets the foundation for creating the artifact. The subsequent stage involves the actual design and development of the digital dictionary, where detailed work is carried out based on the previously defined objectives. Once developed, the artifact undergoes demonstration and evaluation to selected users to ensure it meets the set objectives and functions as intended. Feedback is gathered to make necessary improvements. The final stage is communication, where results and findings are shared with relevant stakeholders, ensuring the knowledge gained from the project is disseminated effectively. This methodology integrates a systematic approach, ensuring that the developed artifact is both theoretically sound and practically applicable, addressing complex problems by breaking them into manageable components with iterative development and stakeholder involvement.

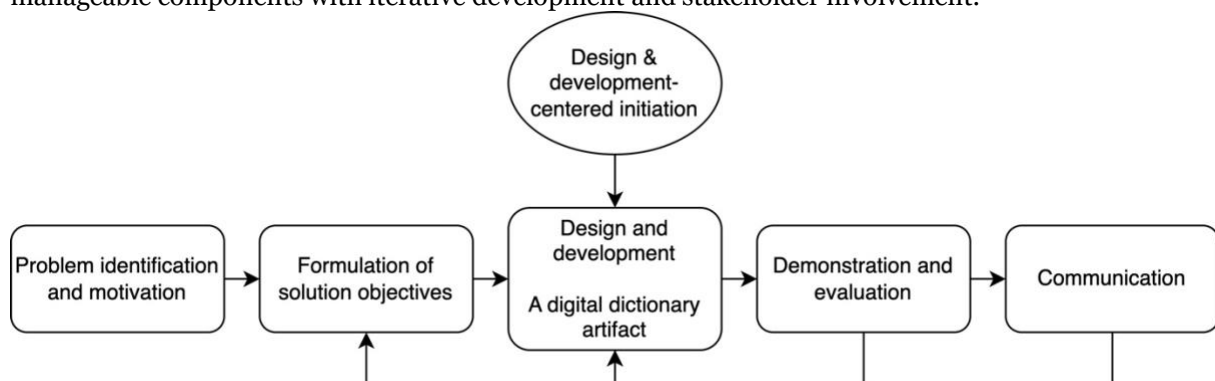


Figure 1. DSRM process for the Enggano digital dictionary development

Each stage of the DSRM will be closely monitored through specific metrics such as community engagement levels, contribution rates, and the frequency of dictionary usage. These metrics will serve

as assessment criteria for the project's success, ensuring that it meets its goals of both language preservation and community involvement.

3 Design and Development

The ongoing Enggano online digital dictionary project, accessible at <https://enggano.cirhss.org>, is built upon a robust and scalable system architecture designed to ensure high availability and performance (see Figure 2). At the foundation lies the cloud infrastructure, providing essential virtualized resources such as servers, storage, and networking, crucial for the deployment and scalability of the application. The operating system used is CentOS, a stable and secure platform that supports the execution of all software components. Docker, a containerization platform, is employed to create isolated environments for the application components, ensuring consistency across development, testing, and production environments. The application layer comprises backend services developed in Golang and a dynamic frontend built with React, facilitating efficient server-side processing and a rich user interface. The data management layer utilizes MySQL, a relational database system that stores and organizes dictionary entries, user data, and other pertinent information.

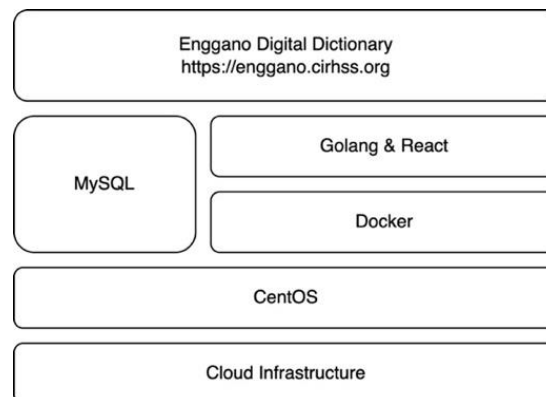


Figure 2. Enggano Dictionary Architecture

In addition to this architecture, the Enggano Digital Dictionary incorporates a Continuous Integration and Continuous Deployment (CI/CD) pipeline. CI/CD is a series of automated processes and tools that allow developers to merge their code changes frequently, ensuring that new code integrates smoothly with the existing codebase. The pipeline includes stages for building, testing, and deploying the application, which helps to catch and fix bugs early, improve software quality, and deliver updates to users more quickly. Running Golang and React applications in containers (like Docker) facilitates this process by providing a consistent and isolated environment for each stage of the pipeline. This comprehensive architectural setup, integrating containerization with Docker and leveraging cloud infrastructure alongside a robust CI/CD pipeline, ensures the Enggano Digital Dictionary's robustness, scalability, and maintainability, thereby supporting its mission to provide a reliable and user-friendly digital resource.

The dictionary's iterative development process allows for gradual improvements, with new functionalities introduced in each cycle based on community feedback. Early demonstrations to the Enggano people will help prioritize features such as search optimization, additional language pairs, and ease of contribution.

The Enggano digital dictionary focuses on providing comprehensive information about the contemporary Enggano language while also including data on older Enggano words. It allows users to search the existing database for linguistic information and contribute new data to enhance and expand the dictionary's database. This collaborative approach ensures the dictionary remains up-to-date and relevant, preserving both modern and old aspects of the Enggano language.

Community feedback is integral to the dictionary's development. Workshops and surveys with Enggano speakers will evaluate the usability of the dictionary and provide linguistic contributions. A structured system will manage the feedback, ensuring suggestions are reviewed and incorporated into future iterations.

The landing page of the Enggano digital dictionary (Figure 3) serves as the entry point for users accessing the database, offering a user-friendly and informative interface. Key features include a prominent search bar that allows users to quickly find specific lexical entries by typing in queries. Users can select the

source language from options such as Enggano, Indonesian, English, and German. The German language is included as a source or target language to our digital dictionary due to many of the old Enggano language resources sourcing the Kähler dictionary (Kähler 1987). Additionally, the page includes a search history feature, enabling users to easily return to previous search results. This design ensures efficient navigation and enhances the overall user experience.

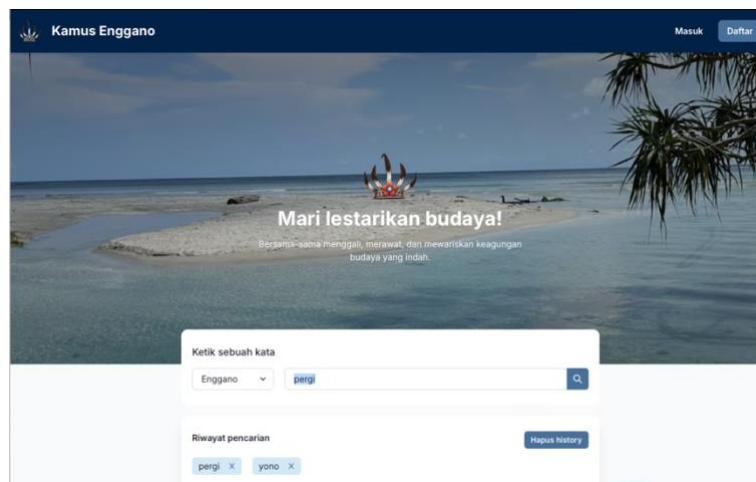


Figure 3. Landing page of the Enggano digital dictionary

Figure 4 shows the output results after entering the word *pergi* (go) in the search feature of the Enggano digital dictionary. It provides detailed entries for related Enggano words, including meanings, example sentences, translations in Indonesian and English, and derived word forms. The layout also includes a list of word classes on the side for easy reference, enhancing the user's ability to navigate and understand the lexical information provided.

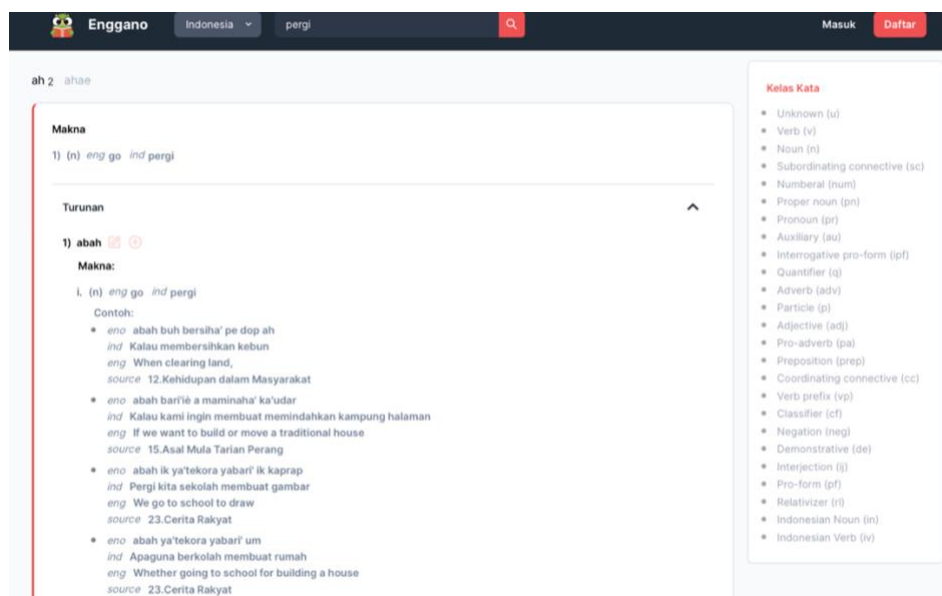


Figure 4. The output for the searching feature

Statistics for the Enggano digital dictionary (Figure 5) showcase the number of headwords (the main word or term being defined or explained in a dictionary or glossary) and derived words (terms that originate from a headword, often formed by adding prefixes, suffixes, or other modifications) available in the dictionary. This figure highlights the dictionary's extensive lexical database, reflecting the efforts to document and preserve the Enggano language comprehensively.

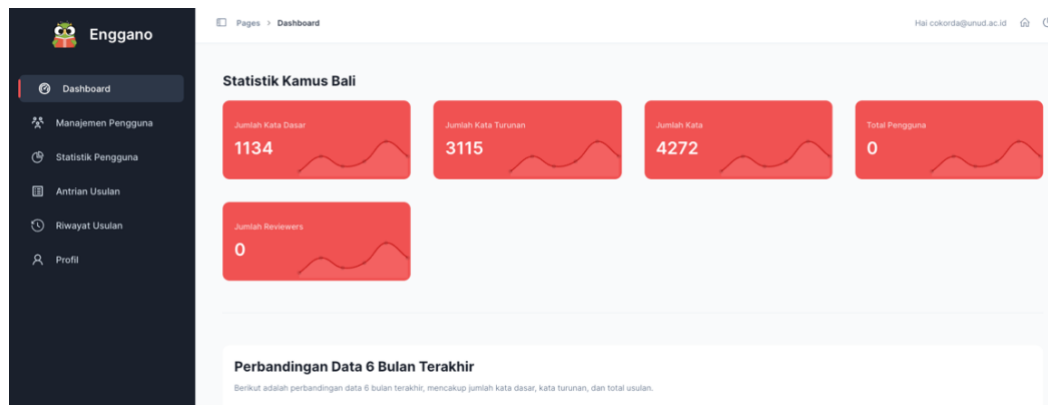


Figure 5. Stats for the number of headwords and derived words available on the Enggano digital dictionary

The list of word suggestions in the Enggano digital dictionary (Figure 6) aids users in finding related terms and enhancing their vocabulary. This feature supports language learning and exploration, contributing to the dictionary's educational value.

The 'Daftar Usulan' page shows a list of word suggestions with the following columns: ID, KATA, PENGUSUL, STATUS, BAHASA, TANGGAL DIUSULKAN, and AKSI. The data is as follows:

ID	KATA	PENGUSUL	STATUS	BAHASA	TANGGAL DIUSULKAN	AKSI
1	kata dasar	mahardikagede0@gmail.com	Kata Baru		2024-07-17	Review
2	dasar	mahardikagede0@gmail.com	Kata Baru		2024-07-14	Review
3	dasar	mahardikagede0@gmail.com	Kata Baru		2024-07-14	Review
4	dasar	mahardikagede0@gmail.com	Kata Baru		2024-07-14	Review
5	dasar	mahardikagede0@gmail.com	Kata Baru		2024-07-14	Review

Figure 6. List of word suggestions

4 Conclusion and Future Work

The development of the Enggano digital dictionary marks an important step in preserving the endangered Enggano language, spoken by around 1500 people. By applying the Design Science Research Methodology (DSRM), the project offers a scalable and accessible digital resource for both language learning and cultural preservation. Through community collaboration and iterative development, the dictionary addresses the shift from Enggano to Indonesian and promotes active language use, ensuring it remains relevant for future generations.

This project contributes to the field of Information Systems (IS) by showcasing how DSRM can be effectively applied to real-world challenges, particularly in community-based projects. It also advances IS theory by demonstrating how modern technologies like cloud infrastructure and continuous integration can support digital preservation efforts. Furthermore, it engages in the broader discussion on decolonizing the digital landscape by empowering indigenous communities to preserve their languages through digital tools, challenging the dominance of major languages online.

Future work will focus on further enriching the dictionary with additional content and features, such as integrating natural language processing (NLP) capabilities. Ongoing community feedback will shape the next iterations, ensuring that the tool continues to meet the linguistic and cultural needs of the Enggano people. By fostering a participatory approach, the project provides a blueprint for similar digital initiatives aimed at preserving endangered languages and promoting cultural sustainability.

5 References

- Beatus, T., Izak Habel, W., Suwarjono, S., Martha, L., Aenal Fuad, A., and Hariyanto, H. 2021. "Preservation of the Muyu Indigenous Language with an Android-Based Dictionary," (4), pp. 85-101.
- Boyd, M., Martine, M., Alexis, M., Séverine, G., Alexandre, F., and Evangelia, A. 2014. "Documenting and Researching Endangered Languages: The Pangloss Collection," *Language Documentation & Conservation* (8), pp. 119-135.
- Buys, P. W. 2024. "A Design Science Methodology," in *Crafting Efficiency in Managerial Costing System Design: An Integrated Design Science Approach*. Springer, pp. 17-32.
- Delport, P. M., Von Solms, R., and Gerber, M. 2024. "Methodological Guidelines for Design Science Research," *Procedia Computer Science* (237), pp. 195-203.
- Ewa, C.-H., Martin, H., and Sarah, M. K. 2014. "Using Tei for an Endangered Language Lexical Resource: The Nxaʔamxáin Database-Dictionary Project," *Language Documentation & Conservation* (8), pp. 1-37.
- Hemmings, C., and Arka, I. W. 2023. "Evidence of Contact with Malay/Indonesian in the Enggano Language," *NUSA* (75), pp. 19-51.
- Jamilah, D. R. N. J. 2022. "English Language Imperative Level in Indonesia," *Intensive Journal* (5:1), pp. 36-36.
- Khalid, A., Mika, H., Jack, R., and Niko, P. 2020. "Ve'rdd. Narrowing the Gap between Paper Dictionaries, Low-Resource Nlp and Community Involvement," pp. 1-6.
- Kähler, H. 1987. "Enggano-Deutsches Wörterbuch," in: *Enggano-Deutsches Wörterbuch*. Dietrich Reimer Verlag.
- Mardhiya, H., Sumarni, A., and Anggit Dwi, H. 2020. "Lexicon-Based Indonesian Local Language Abusive Words Dictionary to Detect Hate Speech in Social Media," *Journal of Information Systems Engineering and Business Intelligence* (6:1), pp. 9-17.
- Muhammad Dian Saputra, T. 2022. "Enggano Language Viability: Diglossia Leakage Phenomenon and Endangered Language Defense Strategy in Enggano," *Linguistik, Terjemahan, Sastra (LINGTERSA)* (3:1), pp. 50-65.
- Pramartha, C., Davis, J. G., and Kuan, K. K. Y. 2017. "Digital Preservation of Cultural Heritage: An Ontology-Based Approach," *The 28th Australasian Conference on Information Systems*, Hobart, Australia, pp. 1-12.
- Pramartha, C., Mahendra, I. M. Y., Rajeg, G. P. W., and Arka, I. W. 2023. "The Development of Semantic Dictionary Prototype for the Balinese Language," *International Journal of Cyber and IT Service Management* (3:2), pp. 96-106.
- Racquel-Maria, S., and Tracy, H.-E. 2019. "Using Existing Documentation for Teaching and Learning Endangered Languages," *Language and Education* (33:6), pp. 560-576.
- Sarmila, N., Andi Muhammad, Y., and Herman, R. 2023. "The Effectiveness of Electronic Dictionary in Learning Vocabulary at Indonesian Junior High School Context," *International Journal of Research on English Teaching and Applied Linguistics* (3:2), pp. 40-44.
- Selly Rizki, Y. 2023. "Digitalizing a Local Language Dictionary," *Lexicography* (10:1), pp. 59-85.
- Tuunanen, T., Winter, R., and Vom Brocke, J. 2024. "Dealing with Complexity in Design Science Research: A Methodology Using Design Echelons," *MIS Quarterly* (48:2).
- UNESCO. (n.d). "Digital Preservation of Indigenous Languages: At the Intersection of Technology and Culture." Retrieved 13 September, 2024, from <https://www.unesco.org/en/articles/digital-preservation-indigenous-languages-intersection-technology-and-culture>
- Wenqi, L. 2023. "Measures to Sustain Endangered Languages: A Bilingual Competition Model with Sliding Mode Control," *PLOS ONE* (18:6), pp. e0287850-e0287850.
- Winata, G. I., Aji, A. F., Cahyawijaya, S., Mahendra, R., Koto, F., Romadhony, A., Kurniawan, K., Moeljadi, D., Prasajo, R. E., and Fung, P. 2022. "Nusax: Multilingual Parallel Sentiment Dataset for 10 Indonesian Local Languages," *arXiv preprint arXiv:2205.15960*.
- Wongso Adi, S. 2022. "Multilingualism in Indonesia: An Introduction to Makassarese Language," *EDULEC EDUCATION LANGUAGE AND CULTURE JOURNAL* (2:2), pp. 161-169.
- Yvonne, S. 2022. "Linguistic Pragmatism, Lingua Francae, and Language Death in Indonesia," *Journal of language teaching* (2:11), pp. 15-19.

Acknowledgements

This project received funding from the Arts & Humanities Research Council (AHRC; Grant ID: AH/W007290/1), United Kingdom, titled "Lexical resources for Enggano, a threatened language of Indonesia"

Copyright

Copyright © 2024 Prammartha, Rajeg, Dalrymple, and Arka. This is an open-access article licensed under a [Creative Commons Attribution-Non-Commercial 4.0 Australia License](https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and ACIS are credited.