



Representation Gaps in Digital Humanities: A Case Study of Marginalized Languages and Cultures in Indonesia

Alim Jaizul^{1*}, Mugi Lestari²

¹*Research Collaboration Community, Bandung, Indonesia*

²*Communication in Research and Publications, Bandung, Indonesia*

**Corresponding author email: alimjaizul47@gmail.com*

Abstract

This study examines the gap in representation of marginal languages and cultures in digital humanities in Indonesia. As a country with more than 700 regional languages, Indonesia faces serious challenges in preserving linguistic diversity in the digital era. Urbanization, globalization, and the dominance of the Indonesian language have accelerated the extinction of regional languages, with only 13 regional languages having more than one million speakers. In the digital context, local languages and cultures tend to be marginalized by the dominance of the majority and global languages. Through a qualitative approach and literature review, this study identifies several key issues: the inequality of local language representation in digital humanities projects, limited data and resources, and minimal participation of local communities. Several initiatives such as BASAbali and NusaWrites demonstrate good practices in preserving languages through community-based approaches and the development of quality corpuses. However, structural challenges such as uneven digital infrastructure, low digital literacy, and the dominance of Western epistemology remain significant obstacles. This study recommends a collaborative approach between academics, local communities, government, and the private sector to build inclusive policies and initiatives. Digital technology needs to be positioned as a means of empowering local culture, not an instrument of global cultural domination, so that digital humanities can become a fair and representative space for the richness of Indonesian languages and cultures.

Keywords: Cultural preservation, community-based technologies, digital humanities, digital representation, marginalized languages

1. Introduction

Digital humanities have revolutionized the way we understand, document, and disseminate cultural heritage. By combining digital technology and humanities studies, the field opens up new opportunities for cultural preservation and exploration. However, behind this progress lies significant challenges related to the representation of marginalized languages and cultures, especially in Indonesia, which has extraordinary linguistic and cultural diversity.

Indonesia is a country with more than 700 regional languages, making it one of the most linguistically diverse countries in the world. Unfortunately, many of these languages are on the verge of extinction. According to data from the Language Agency, only 13 regional languages have more than one million speakers, while the rest are experiencing a drastic decline in the number of speakers. Factors such as urbanization, globalization, and the dominance of Indonesian as an official language have accelerated this process (Septiyana & Margiansyah, 2018; Halim & Sukanto, 2023).

In the digital context, marginalized languages and cultures often do not get adequate space. Digital platforms tend to be dominated by content in the majority language or global languages such as English. As a result, local languages and cultures are underrepresented, potentially leading to the erosion of cultural identity (Ullah & Akram, 2023).

Several initiatives have been undertaken to address this issue. For example, BASAbali is a wiki platform that aims to preserve the Balinese language through community participation. The platform provides a digital dictionary, learning modules, and discussion space for Balinese language users (Pramartha et al., 2020). However, such initiatives are still limited and do not cover all marginalized languages and cultures in Indonesia.

This gap in representation is also evident in existing digital humanities projects. Many projects focus on majority cultures and languages, while minority cultures and languages receive less attention. This creates an imbalance in cultural preservation and documentation, and hinders efforts to preserve culture as a whole.

In addition, the lack of data and resources for marginalized languages and cultures is a barrier to the development of technologies that support the preservation of these cultures. Projects such as NusaWrites are trying to address this issue by building high-quality corpora for underrepresented languages in Indonesia (Cahyawijaya et al., 2023). However, challenges such as limited data and lack of community participation remain obstacles.

The importance of equitable representation in the digital humanities is not only related to cultural preservation, but also to social justice and inclusivity. When marginalized languages and cultures are not given space in the digital world, the communities that use them are also marginalized. Therefore, efforts to improve representation must involve the active participation of these communities.

In the ever-evolving digital era, it is important to ensure that technology is used as a tool to strengthen, not erode, cultural diversity. This requires collaboration between government, academia, communities, and the private sector to develop policies and initiatives that support the preservation of marginalized languages and cultures (Saaida, 2023).

This study aims to identify and analyze the representation gaps in the digital humanities related to marginalized languages and cultures in Indonesia. By understanding the factors that cause these gaps, it is hoped that strategies can be developed to improve inclusivity and representation in digital humanities projects.

Through a qualitative approach and case studies, this study will explore how current digital humanities projects represent marginalized languages and cultures, and how local communities can play a role in this process. The results of this study are expected to contribute to the development of a more inclusive and representative digital humanities. Thus, the preservation of marginalized languages and cultures in Indonesia can be strengthened through the use of digital technology, ensuring that the nation's cultural diversity is maintained amidst the current of globalization.

2. Literature Review

Zahra (2024), explores the application of digital humanities methods, such as text mining and natural language processing (NLP), in analyzing Indonesian literary works. This study shows that these techniques are effective in identifying thematic patterns and semantic structures in large text corpora, enriching the understanding of cultural and social influences in literature. However, the focus of this research is still limited to the majority language, so that regional languages and marginalized cultures have not received adequate attention. This highlights the need to develop more inclusive methodologies to cover Indonesia's linguistic diversity.

The NusaWrites project by Cahyawijaya et al. (2023) aims to build a high-quality corpus for underrepresented languages in Indonesia. They compare three data collection methods: online scraping, human translation, and paragraph writing by native speakers. The results show that data produced by native speakers has higher lexical diversity and cultural relevance. These findings emphasize the importance of local community involvement in language preservation through digital technology, as well as the need for an approach that focuses on quality and cultural context in the development of linguistic resources.

Ariyani et al. (2022) in their study evaluated the effectiveness of digital tools such as mobile applications, online dictionaries, and social media platforms in preserving endangered languages in Indonesia. They highlighted that although these tools have great potential, challenges such as lack of infrastructure, limited access to technology, and low community participation remain significant barriers. Therefore, language preservation strategies must include increasing access to technology, community training, and developing policies that support the sustainability of digital initiatives.

Wiryomartono (2023) criticized the dominance of Western languages and cultures in global digital humanities projects, which often ignore non-Western languages and cultures such as those in Indonesia. He highlighted that institutional structures and global technology policies tend to be inclusive of linguistic diversity, creating knowledge gaps and reinforcing epistemic injustice in the digital world. This highlights the need for decolonization of digital humanities and development of policies that support more equitable representation of marginalized languages and cultures.

BASAbali is a wiki platform that aims to preserve the Balinese language through community participation. The platform provides a digital dictionary, learning modules, and discussion space for Balinese language users. The success of BASAbali shows that a community-based approach can be effective in preserving local languages and cultures through digital technology. However, challenges such as sustainability, funding, and expanding coverage to other languages remain concerns. The BASAbali model can be used as an example for similar initiatives in other regions with adaptations according to local contexts (Pramartha et al., 2020).

3. Methods

This study uses a descriptive qualitative approach with a literature review method (library research) as the main foundation. This method was chosen to explore and analyze the gap in the representation of marginalized languages and cultures in the development of digital humanities in Indonesia. Literature review allows researchers to critically examine various relevant scientific sources, including journal articles, research reports, academic books, policy documents, and related digital projects that have been carried out previously both at the national and international levels.

The data used in this study were obtained through a review of the latest literature that discusses the application of digital technology in the humanities, especially in the context of preserving local languages and cultures. Several key studies that were analyzed in depth include Zahra (2024) research on the application of NLP in Indonesian text analysis, the NusaWrites project by Cahyawijaya et al. (2023) which built a corpus for regional languages, and Ariyani et al. (2022) study which evaluated the effectiveness of digital tools in revitalizing endangered languages. In addition, attention is also paid to the critique of global representation in digital humanities projects by Wiryomartono (2023), as well as the case study of BASAbali as a model of community-based preservation (Pramartha et al., 2020).

The analysis was conducted by identifying key themes in the literature, then examining how these studies highlight issues of representation, community participation, digital resource development, and challenges in policy and infrastructure. The results of this analysis will be used to formulate a mapping of gaps in digital humanities research and practice in Indonesia, as well as to develop conceptual recommendations for creating a more inclusive and equitable approach to marginalized languages and cultures in the digital realm.

4. Discussion

4.1. Inequality in representation of local languages and cultures in digital humanities projects

The results of the literature review show significant inequalities in the representation of marginalized local languages and cultures in digital humanities projects, both at the national and global levels. As stated by Wiryomartono (2023), digital humanities projects tend to prioritize dominant languages and cultures, especially from the West, which have better infrastructure and access to resources. This leads to further marginalization of local communities in countries such as Indonesia that have high linguistic and cultural diversity. This inequality is evident in the lack of digital documentation in local languages, the lack of user interfaces in local languages, and the absence of local culture-based educational materials in public digital spaces.

This phenomenon has broad implications for the sustainability of minority languages and cultures. When digital representation is limited to dominant languages such as Indonesian or English, local languages increasingly lose their existential space. In fact, digitalization should be an inclusive means to expand participation in local culture and knowledge. The lack of attention to digital inclusion of local languages will exacerbate the inequality of access to information, as well as create new forms of cultural and knowledge injustice in the digital era.

4.2. Preservation strategy through local community involvement

Cahyawijaya et al. (2023), in the NusaWrites project, underlines the importance of active involvement of local communities in the collection and development of regional language corpora. Through a comparative study between online scraping, translation, and writing by native speakers, it was found that participatory methods gave the best results in terms of completeness, accuracy, and cultural relevance. This shows that the success of digital humanities projects focused on language preservation is highly dependent on the direct involvement of the community as producers and guardians of their own culture.

Community involvement not only improves data quality but also strengthens the sense of ownership of their language and culture. This is especially important in the context of revitalizing endangered languages, where the survival of a language is highly dependent on its active users. Therefore, a participatory approach must be a primary principle in the development of digital projects oriented towards preserving local culture.

In addition, community involvement can be expanded through technology-based training and education models, such as digital storytelling training, local-based data labeling workshops, and digital content development training using mother tongue. Initiatives like these will strengthen the capacity of communities to become active subjects in the digital ecosystem, not just objects of documentation.

4.3. Potential and Limitations of Digital Technology in Language Preservation

Ariyani et al. (2022) emphasized that digital technologies such as mobile applications, online dictionaries, and social media platforms offer great potential for preserving regional languages, especially in increasing access to and use of languages in everyday life. However, structural challenges such as unequal access to ICT infrastructure, limited human resources, and low digital literacy in remote areas are still major obstacles.

Not only that, technological approaches that are not sensitive to local cultural contexts often create resistance from the community. For example, many applications do not consider local norms or customs in delivering their content. Therefore, the development of language preservation technology must adopt a culture-based design approach, which places local values and preferences at the heart of digital application development.

Strategies to overcome these limitations include cross-sector collaboration between the government, NGOs, academics, and technology developers. The government needs to provide regulatory and infrastructure support; NGOs and academics can act as liaisons with local communities; while technology developers must open up space for participation in the digital product design process. Thus, language preservation does not only depend on the availability of technology, but also on a collaborative ecosystem that ensures the sustainability of the project.

4.4. Good practices (BASAbali Case Study)

One real example of good practices in language preservation through community-based digital media is BASAbali. This platform not only provides an online dictionary in Balinese, but also offers language learning features, discussion forums, and open contribution channels for the community. The success of BASAbali, as explained by Pramatha et al. (2020), lies in its ability to create a dynamic and participatory online community. In addition, BASAbali has also succeeded in establishing partnerships with local governments, educational institutions, and community leaders in disseminating the spirit of cultural preservation.

However, BASAbali's success is not without challenges. One of the main challenges is the financial sustainability and institutionalization of this model to other regions. In many other regions in Indonesia, there has not been a similar model that has been consistently implemented. Therefore, the BASAbali model can be used as a strategic reference for the development of similar platforms in other regions, of course with contextual adjustments to the characteristics of different local languages and cultures.

In the long term, BASAbali proves that language preservation is not only possible in the digital space, but can also be a means of empowering local identities, if carried out with a collaborative and community-based model.

4.5. The need for decolonization of digital humanities in Indonesia

The literature also highlights the importance of a decolonial approach in Indonesian digital humanities. Decolonization here does not only target hegemonic Western narratives, but also patterns of knowledge production and consumption that do not reflect Indonesia's own cultural diversity. As stated by Wiryomartono (2023), digital humanities institutions and projects often reproduce exclusive structures that discriminate against local languages and cultures. For example, digital projects that archive ancient manuscripts or literary works are often only in Indonesian or English formats, without translation or adaptation in local languages.

Decolonization also means making space for non-Latin, non-textual, and oral-based forms of cultural expression – which are characteristic of many indigenous communities in Indonesia. The digitization of local cultures must take into account the diversity of local epistemologies, and not impose Western documentation standards or narratives alone. This is important to maintain the authenticity of cultural representations and to encourage alternative, community-based documentation models.

4.6. The role of public policy and educational institutions

The success of preserving language and culture in the digital realm also depends on regulations and policy interventions that support cultural diversity. Saaida (2023) underlines the importance of the state's role in facilitating and supporting a digital ecosystem that supports multilingualism. Unfortunately, national policies related to digital transformation still do not explicitly encourage the digitalization of regional culture, especially with a community-based approach.

Educational institutions, from elementary schools to universities, have a great responsibility to include cultural and technological literacy in their curriculum. Training activities on the use of technology to preserve local languages can be part of community service or the independent campus program. That way, language preservation is not only the responsibility of researchers or language activists, but becomes an integral part of the education and national development process.

These findings emphasize that preserving local languages and cultures in the context of digital humanities requires an interdisciplinary, participatory, and equitable approach. Fair and equal representation in the digital space is a primary prerequisite for maintaining the sustainability of Indonesian cultural identity in the era of globalization.

5. Conclusion

Based on the results of the literature study conducted, it can be concluded that digital humanities has made a significant contribution to the preservation and exploration of cultural heritage, but there is still a striking gap in representation of marginalized languages and cultures in Indonesia. Most digital projects tend to focus on majority or global languages, while regional languages, which number more than 700 with endangered status, have not received adequate space in the digital realm.

Previous studies have shown that community-based approaches, such as those applied to the BASAbali platform, as well as the development of quality corpuses such as in the NusaWrites project, are good practices that can be replicated by considering the local context. However, there are still structural challenges such as limited resources, uneven digital infrastructure, minimal participation of local communities, and the dominance of Western epistemology in global digital humanities projects.

Therefore, a collaborative and interdisciplinary approach is needed that involves academics, local communities, government, and the private sector to build inclusive policies, platforms, and initiatives. Digital technology must be positioned as a means of empowering local culture, not as an instrument of global cultural domination. Thus, digital

humanities can function as a fair and representative space for the entire richness of Indonesian languages and cultures, as well as strengthen national identity amidst the current of globalization.

References

- Ariyani, F., Putrawan, G. E., Riyanda, A. R., Idris, A. R., Misliani, L., & Perdana, R. (2022). Technology and minority language: an Android-based dictionary development for the Lampung language maintenance in Indonesia. *Tapuya: Latin American Science, Technology and Society*, 5(1), 2015088.
- Cahyawijaya, S., Lovenia, H., Koto, F., Adhista, D., Dave, E., Oktavianti, S., ... & Fung, P. (2023). Nusawrites: Constructing high-quality corpora for underrepresented and extremely low-resource languages. *arXiv preprint arXiv:2309.10661*.
- Halim, S. W., & Sukamto, K. E. (2023). The (in) visibility of Torajan language: A study on linguistic landscape in South Sulawesi, Indonesia. *Studies in English Language and Education*, 10(3), 1585-1607.
- Pramartha, C., Iswara, I. B. A. I., & Mogi, I. K. A. (2020). Digital Humanities: Community Participation in the Balinese Language Digital Dictionary. *Jurnal Sistem Informasi (Journal of Information System)*, 16(2), 18-30.
- Saaida, M. (2023). Global catalysts: Empowering universities for social change in developing countries. *Edukasiana: Jurnal Inovasi Pendidikan*, 2(4), 318-341.
- Septiyana, I., & Margiansyah, D. (2018, December). Glocalization of intangible cultural heritage: strengthening preservation of Indonesia's endangered languages in globalized world. In *International Conference on Contemporary Social and Political Affairs (IcoCSPA 2017)* (pp. 85-88). Atlantis Press.
- Ullah, F., & Akram, A. (2023). Linguistic imperialism in the globalized world: Examining English dominance and its sociolinguistic consequences. *Harf-o-Sukhan*, 7(3), 279-288.
- Wiryomartono, B. (2023). *Globalization, Urbanization, and Civil Society: A Non-western Critique*. Routledge.
- Zahra, R. M. (2024). Digital humanities approaches to analyzing indonesian language texts as non-western languages. *JPPI (Jurnal Penelitian Pendidikan Indonesia)*, 10(4), 787-795.