



Check for updates



Open Access | Original Research

Journal of Intercultural Communication, 25(3), 2025 | PP: 11– 24

<https://doi.org/10.36923/jicc.v25i3.1161>

Semantic Knowledge Graphs for Intercultural and Ethnographic Diversity in the Greater Mekong Subregion

Article History:

Received: 25-04-2025

Revision: 07-07-2025

Accepted: 18-06-2025

Publication: 19-07-2025

Cite this article as:

Chansanam, W., Nguyen, L. T. ., Li, C. ., & Guan, C. K. S. (2025).

Semantic Knowledge Graphs for Intercultural and Ethnographic Diversity in the Greater Mekong Subregion. *Journal of Intercultural Communication*, 25(3), 11-24. doi.org/10.36923/jicc.v25i3.1161

©2025 by author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License 4.0 International License.

Corresponding Author:

Wirapong Chansanam

Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand. Email: wirach@kku.ac.th

Wirapong Chansanam¹ , Lan Thi Nguyen² , Chunqiu Li³ , Christopher Khoo Soo Guan⁴ 

Abstract: This study aims to design, develop, and evaluate a semantic knowledge graph (KG) and web application that ethically and effectively represents the ethnographic diversity of 375 ethnic groups across Thailand, Laos, Myanmar, Cambodia, Vietnam, and China. Addressing the limitations of conventional Knowledge Organization Systems (KOS), the project integrates semantic web technologies with community-informed data to capture complex relationships involving language, religion, cultural practices, geography, and historical migration. This study employed structured data transformation, entity-relation modeling, and knowledge graph construction using Neo4j. It further incorporated interactive data visualization through React.js and D3.js. The dataset was collaboratively developed by over 20 regional scholars, ensuring cultural relevance and accuracy. Evaluation was conducted through extrinsic and domain-specific validation involving 24 usability experts and 17 ethnographers, who assessed task performance, data completeness, and cultural appropriateness. The findings of this study reveals that the knowledge graph significantly improves usability, interpretability, and task efficiency, exceeding traditional database performance by over 25% in expert evaluations. Additionally, the semantic model supports multilingual access, temporal migration tracking, and culturally contextualized identities. This research concluded that semantically enriched, ethically designed knowledge infrastructures can enhance representation and accessibility of marginalized knowledge. Key limitations include manual data curation and gaps in representing hybrid cultural practices. Future work should focus on scalability through automation and sustained community collaboration. The implications extend to digital humanities, education, and policymaking, offering a replicable model for equitable, culturally sensitive knowledge systems worldwide.

Keywords: Knowledge Graph, Ethnic Groups, Greater Mekong Subregion, Knowledge Organization Systems, Cultural Representation, Digital Humanities

1. Introduction

An estimated 10-20% of the world's population, up to 1.2 billion people, belong to ethnic and cultural minorities, many of whom face discrimination, marginalization, and systemic exclusion. These realities underscore the urgent need for inclusive, culturally sensitive systems that safeguard and accurately reflect minority identities (Goodman et al., 2023). However, despite decades of progress in information science, many existing knowledge systems fail to capture the complexity, fluidity, and richness of these identities. The central question thus arises: how can we ethically and accurately represent ethnographic diversity in the digital systems upon which researchers, institutions, and policymakers increasingly rely?

Knowledge Organization Systems (KOS), including taxonomies, classification schemes, thesauri, and ontologies, have long provided the backbone for structuring and retrieving information (Feng et al., 2024; Moraitou et al., 2022). Yet these systems, often shaped by Western epistemological frameworks, embed structural biases that marginalize non-Western and Indigenous knowledge (Doyle, 2013; Rubim Silva & Dal'Evedove, 2025). As a result, traditional KOS tend to represent knowledge from dominant cultural standpoints, rendering minority perspectives either invisible or misrepresented (Wright & Saelua, 2023; Anania & Stiglitz, 2023).

In contrast, Knowledge Graphs (KGs) offer semantically rich and flexible structures for modeling complex relationships between entities. As graph-based systems, KGs are increasingly recognized for their potential to support pluralistic and interdisciplinary knowledge ecosystems, particularly when integrated with culturally grounded ontological frameworks (Dally et al., 2024; Nguyen et al., 2023). Moreover, language is a central dimension of both KOS and KGs. Recognizing and integrating minority languages not only supports cultural preservation but also promotes equitable access to services, data, and epistemic justice (Couldry & Mejias, 2023).

Nonetheless, mainstream classification systems such as the Library of Congress Subject Headings (LCSH) and the Dewey Decimal Classification (DDC) continue to draw criticism for their Eurocentric hierarchies. In response, alternative and community-led models like *Lau ā Lau ka 'Ike* and the Infomediary of Taiwanese Indigenous Peoples have emerged to reflect Indigenous epistemologies (Sung & Chi, 2021; Wright & Saelua, 2023). While these initiatives

¹ Faculty of Humanities And Social Sciences, Khon Kaen University, Thailand

² Faculty of Humanities And Social Sciences, Khon Kaen University, Thailand

³ School of Government, Beijing Normal University, Beijing, China

⁴ Wee Kim Wee School of Communication & Information, Nanyang Technological University, Singapore

exemplify valuable resistance to epistemic hegemony, they remain peripheral in the global knowledge infrastructure (Huang & Zaslavsky, 2024). Moreover, challenges persist regarding data sovereignty, co-design ethics, and the risk of reproducing power asymmetries through externally imposed ontologies (Siegers et al., 2023; Hou et al., 2022).

Recent efforts to document intangible heritage, such as languages, oral histories, and cultural artifacts, have demonstrated the value of linked open data (LOD) and KG technologies (Cui et al., 2024; Du et al., 2025; Fan, 2023). However, many of these systems still fall short in representing hybrid or evolving identities, especially when modeled through rigid taxonomies. For example, ontologies developed for health or archival data, while improving formal data accuracy, often struggle to reconcile contextual, relational, and narrative-based forms of knowledge (Nguyen et al., 2023; Mateos, 2007; Woods, 2017).

In the domain of knowledge management, the role of tacit knowledge, oral traditions, and community archiving remains central, particularly among minority groups adapting to globalization and generational change (Sewdass, 2014; Bonilla et al., 2025). Nevertheless, structural inequities, microaggressions, and digital divides continue to impede effective knowledge sharing and cultural visibility (Davis, 2015; Khurram & Giangliulio, 2020; Li et al., 2023).

Therefore, Emerging Linked Open Data (LOD) and knowledge graph applications offer promising solutions. Projects in Laos, Vietnam, and the Greater Mekong Subregion have enabled multilingual, standardized access to ethnic knowledge and folktales (Chansanam et al., 2024; Ngootip et al., 2023; Ngoc & Chansanam, 2023). Broader efforts also document applications in preserving craftsmanship, biodiversity, and ethnographic knowledge (Bai & Hou, 2023; Chansanam et al., 2022, 2020; Cui et al., 2024; Lu et al., 2023). Still lacking, however, is a cohesive, interdisciplinary synthesis that critically evaluates the equity and sustainability of these digital innovations (Jaroenruen et al., 2024; Khoo et al., 2024).

To address this gap, this study engages with both decolonial information science and intercultural communication theories. Drawing on scholars such as Mignolo (2011) and Green (1999), it recognizes the need for research methodologies that legitimize Indigenous knowledge structures and allow for community-defined ontologies. Additionally, perspectives from Gudykunst (2003, 2004) and Kim (2015) shape the system's emphasis on relational modeling, multilingual design, and cultural fluidity, ensuring that identity representation is negotiated rather than imposed.

Despite advances in graph technologies, traditional KOS continue to rely on siloed, hierarchical taxonomies that restrict the representation of complex ethnographic relationships (Smith, 2021). This structural rigidity not only limits interoperability across databases but also impedes discovery, learning, and cross-disciplinary innovation. Consequently, there is a pressing need for systems that can semantically integrate geo-temporal, linguistic, cultural, and relational data in a unified yet flexible framework.

This study responds to these challenges by designing, constructing, and evaluating a knowledge graph that models the ethnographic diversity of 375 ethnic groups across six countries in the Greater Mekong Subregion (GMS). The graph incorporates multilingual identifiers, migration histories, cultural practices, and linguistic lineages, offering a nuanced and scalable model for inclusive information representation. Furthermore, the project includes a custom-built web application for real-time visualization, interactive exploration, and expert validation. Therefore, this research addressed the following research question, how can a semantically enriched, ethically grounded knowledge graph framework improve the representation, accessibility, and interpretability of complex ethnographic data across the Greater Mekong Subregion?

However, this study makes four key contributions. First, it introduces a cross-border, region-wide KG on ethnographic diversity, integrating community-sourced and validated data. Second, it advances methodological integration by combining KOS, KG, and knowledge management approaches to model identity, migration, and cultural relationships. Third, it presents an open-access web platform supported by robust expert validation and usability metrics. Finally, it establishes a reusable framework for inclusive, scalable, and ethically oriented digital heritage systems that can be adapted for other regions and disciplines. Moreover, this research not only contributes to digital ethnography and semantic technologies but also advocates for epistemic justice by challenging dominant paradigms in knowledge organization and representation.

2. Methodology

This research focuses on developing and implementing a Knowledge Graph (KG) web application designed to model the cultural and ethnic diversity of the Greater Mekong Subregion (GMS). The study emphasizes the transformation of a structured dataset into a semantic knowledge graph that captures the complex relationships among 375 ethnic groups across six countries, Thailand, Laos, Myanmar, Cambodia, Vietnam, and China, highlighting linguistic, cultural, geographic, and historical attributes. The construction process for the GMS Ethnic Groups Knowledge Graph is illustrated in Figure 1.

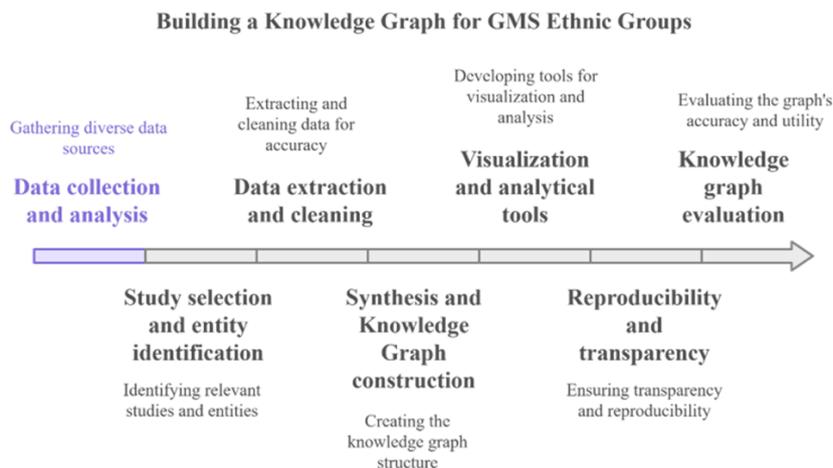


Figure 1: The GMS ethnic groups' knowledge graph construction process

2.1. Data Collection and Analysis

The primary data source for this study comprises a comprehensive dataset that was meticulously collected, analyzed, and synthesized as part of a collaborative research initiative led by local scholars across the Greater Mekong Subregion. This collaborative approach ensured cultural sensitivity, contextual relevance, and authenticity in the information presented. The dataset included critical informational fields such as ethnolinguistic identifiers (autonyms, ethnonyms, exonyms, and endonyms), spoken languages and their respective language families, geographic distribution, national affiliations, and cultural attributes, including traditional housing, clothing, festivals, and religious practices. Additional variables, such as population size and historical migration patterns, were incorporated to provide a multidimensional representation of ethnic diversity in the region.

The dataset was compiled by more than 20 local scholars affiliated with regional universities, cultural heritage institutions, and ethnolinguistic research centers. Data were obtained from diverse sources, including national censuses, ethnographic field reports, community archives, oral histories, and peer-reviewed studies. Each scholar contributed to a specific subset of ethnic groups within their respective country to ensure accuracy and contextual appropriateness. To enhance validity, the data were cross-referenced with authoritative sources such as Ethnologue, Glottolog, and official linguistic atlases. Ambiguities or inconsistencies were flagged and subsequently reviewed by an advisory panel of senior ethnographers. Additionally, an auxiliary JSON file was created to facilitate downstream processing. To enrich contextual depth, the knowledge graph integrated linguistic taxonomies, cultural classification frameworks, and geospatial information.

2.2. Study Selection and Entity Identification

A systematic approach was employed to identify and define core entities and relationships within the knowledge graph. These included entities such as ethnic groups, languages, countries, geographic regions, religions, festivals, and housing types, along with corresponding relationships that captured spatial, linguistic, cultural, and historical linkages. For example, relationships were modeled to represent location associations, language families, religious practices, inter-ethnic cultural exchanges, and naming structures. This structured approach ensured that the KG accurately reflects the interconnected nature of ethnographic attributes in the region.

- *LOCATED_IN* (ethnic group → country),
- *SPEAKS_LANGUAGE_FROM* (ethnic group → language family),
- *PRACTICES* (ethnic group → religion),
- *SHARES_ORIGIN_WITH*, *SHARES_HOUSING_STYLE_WITH*, *SHARES_FESTIVAL_TRADITION_WITH* (inter-ethnic relationships),
- *HAS_AUTONYM*, *HAS_EXONYM*, *HAS_ETHNONYM* (ethnic group naming structures).

2.3. Data Extraction and Cleaning

Data preprocessing involved several critical steps to guarantee consistency and accuracy. First, a standardization process was implemented to harmonize language family names and normalize country labels, using Glottolog and Ethnologue classifications as primary references. Locally used or culturally specific terms were aligned with widely accepted academic terminology. For example, terms such as “Kadai” and “Kra-Dai” were standardized under “Tai-Kadai,” while “Mon-Khmer” and “Austroasiatic” were reconciled in accordance with prevailing linguistic consensus. Where discrepancies arose, expert consultations with regional linguists were conducted to ensure precise classification.

Following standardization, unique identifiers were generated for each entity by combining key attributes, thereby supporting consistent referencing during the graph-building process. Data cleaning was then performed to remove duplicates, address missing values, and normalize formatting, ensuring structural integrity. Finally, the cleaned dataset was converted into structured, entity-relationship formats suitable for integration into the KG framework.

2.4. Knowledge Graph Construction

The knowledge graph was constructed by designing a semantic architecture in which nodes represented identified entities and edges encoded relationships. This structure allowed for a rich, interconnected representation of ethnographic data. Cypher query templates were developed in Neo4j to enable efficient graph generation, querying, and exploration. Illustrative examples include relationships such as “Hmong” linked to “Hmoob” through the *HAS_AUTONYM* property, and “Hmong” associated with the “Hmong-Mien” language family through the *SPEAKS_LANGUAGE_FROM* relationship. These examples underscore the graph’s capacity to capture multilingual and multicultural dimensions in a semantically coherent manner.

Sample graph structures, such as:

- "Hmong" (*Ethnic Group*) → *HAS_AUTONYM* → "Hmoob" (*Autonym*)
- "Hmong" → *SPEAKS_LANGUAGE_FROM* → "Hmong-Mien" (*Language Family*)

Illustrate the multilingual and multicultural dimensions captured in the KG.

2.5. Visualization and Analytical Tools

To facilitate interactive exploration and analytical capabilities, a web application was developed, incorporating a robust visualization interface. The frontend was implemented using React.js and D3.js to enable dynamic graph visualization and user interaction, while the backend was built on Node.js and Express, integrated with a GraphQL API to interact with the Neo4j database. The system provides multiple visualization modes, including relationship graphs, geospatial mappings, and comparative dashboards with advanced filtering and search functionalities. Additionally, detailed information panels display attributes such as migration origins, religious affiliations, traditional attire, linguistic identifiers, and cultural practices. This design ensures both exploratory navigation and hypothesis-driven analysis of ethnographic structures within the GMS region.

2.6. Reproducibility and Transparency

To promote transparency and reproducibility, all data transformation processes were fully documented. Scripts for preprocessing, data loading, and KG construction were developed in Python and organized using Jupyter Notebooks. Furthermore, the full Neo4j Cypher implementation and API endpoints were published for reuse, enabling researchers to replicate or extend the methodology in future studies.

2.7. Knowledge Graph Evaluation

The evaluation strategy adopted a dual-method approach as outlined by Choi and Jung (2025), combining extrinsic evaluation and dataset-specific assessment. Extrinsic evaluation measured the KG's practical utility in operational settings, focusing on its ability to improve task performance, query efficiency, and interpretability. In contrast, dataset and domain-specific evaluation examined accuracy, completeness, and internal consistency within the ethnographic domain.

A usability study was conducted with 24 domain experts specializing in ethnography, digital humanities, and information science, all of whom had an average of 9.5 years of experience. Participants were recruited through purposive sampling from institutional networks across Southeast Asia. They were provided with a demonstration of the system before performing structured, task-based evaluations. Performance metrics included task completion rate, completion time, information discovery success, and visualization clarity, assessed using standardized Likert scales and time logs.

In addition, 17 ethnographers with expertise in linguistic anthropology, material culture, and regional studies participated in the domain-specific validation process. Using a three-dimensional rubric (accuracy, reliability, completeness), they evaluated the dataset and annotated inconsistencies or gaps. Feedback from this validation informed the consistency analysis reported in the results section.

2.8. Ethical Considerations

Ethical integrity was central to every phase of the research. Although the project did not involve personal or identifiable human subject data, it engaged extensively with community-level ethnographic information, which carries cultural and political significance. Consequently, the study adhered to principles of data sovereignty, ensuring that cultural communities retained the right to define and contextualize their data. Local scholars actively participated in decisions regarding naming conventions and classification structures to prevent externally imposed categorizations.

Informed consent was obtained from all contributing experts involved in data provision and validation, following clear communication of research objectives and procedures. Cultural sensitivity was further embedded in the modeling process by incorporating autonyms, multilingual identifiers, and non-hierarchical classification systems. Additionally, regional experts reviewed all data to ensure accuracy, contextual appropriateness, and respect for cultural representation throughout the study.

3. Results

The results of this study present the structure, content, and analytical capabilities of the Greater Mekong Subregion (GMS) Ethnic Groups Knowledge Graph (KG) and its associated web application. Through a series of visualizations, database queries, and interface demonstrations, the findings illustrate how ethnographic data for 375 ethnic groups across six countries have been semantically modeled and interactively explored. These results highlight the knowledge graph's ability to represent complex interrelationships among cultural, linguistic, religious, and geographic dimensions. Detailed visual outputs, such as ego-network views for specific ethnic groups and geospatial mappings of ethnic distributions, demonstrate the system's capacity to capture multidimensional identities and support domain-specific research. Additionally, technical and user-centered performance evaluations validate the knowledge graph's practical utility and superiority over traditional database approaches in representing ethnographic complexity.

3.1. The general characteristics of the GMS Ethnic groups' knowledge graph

Figure 2 presents a Neo4j graph visualization depicting the interconnectedness of ethnic groups and cultural attributes in the Greater Mekong Subregion. The query returns a network of 334 nodes and 426 relationships, with entity types color-coded: ethnic groups (purple), religions (orange), regions (green), languages (light blue), and cultural attributes. Relationship types include "PRACTICES" (297), "FOUND IN" (70), "HAS_POPULATION_CATEGORY" (22), and "MIGRATED_FROM" (14), among others. The structure highlights major ethnic groups as central hubs linked to diverse cultural, linguistic, and geographic data. This knowledge graph enables the semantic modeling of ethnographic complexity, supporting the exploration of cultural affiliations, migration patterns, and regional distributions. The implementation demonstrates how anthropological data can be effectively represented as an interconnected network for cross-domain research and analysis.

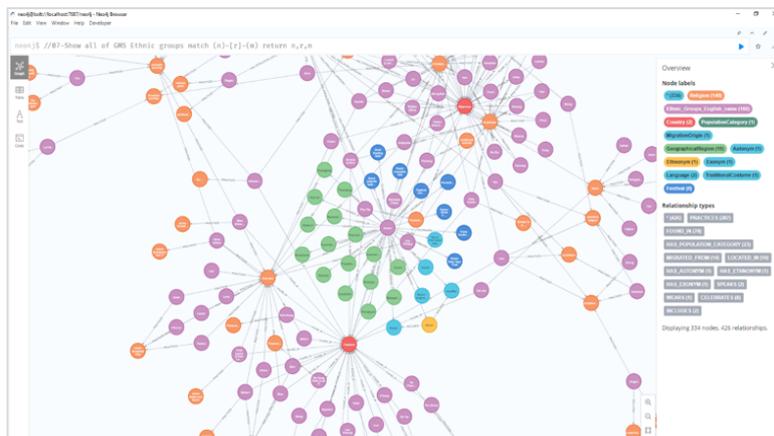


Figure 2: GMS Ethnic groups knowledge graph on Neo4j

The implementation demonstrates how anthropological data can be effectively represented as an interconnected network for cross-domain research and analysis.

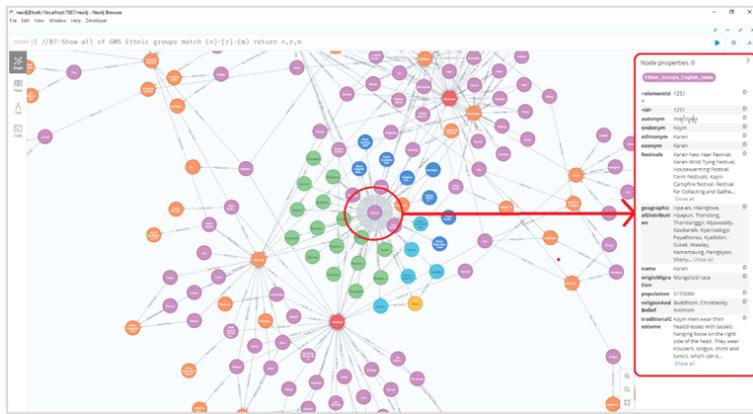


Figure 3: A focus on the Karen ethnic group node and its properties panel

Figure 3 offers a focused view of the Neo4j graph visualization, highlighting the Karen ethnic group (node ID 1251) within the Greater Mekong Subregion (GMS). The properties panel displays multilingual identifiers, “Karen” (ethnonym/exonym), “Kayin” (endonym), and “ကရင်လူမျိုး” (autonym in Burmese script), alongside a population estimate of 5.15 million. Geographically, the group spans regions in Myanmar and Thailand, including Hpa-an, Myawaddy, and Kawkaeik. Religious affiliations include Buddhism, Christianity, and Animism, reflecting syncretic belief systems. Cultural attributes feature traditional festivals, Karen New Year, Wrist Tying, Housewarming, and agricultural celebrations,

and costume details such as tasseled headdresses and region-specific attire. This knowledge graph exemplifies how computational methods in digital humanities can encode complex ethnographic data, enabling relational exploration of cultural identities, geographic distribution, and linguistic diversity across the GMS.

Figure 4 displays the home page of the Greater Mekong Subregion (GMS) Ethnic Groups Knowledge Graph web application, developed using Node.js. The interface presents a responsive visualization of interconnected ethnic group data, with navigation options for "Graph View" and "Geographic View" enabling multiple representation formats. The left sidebar offers interactive filters, including a search by ethnic group, country, language family, and a temporal slider with preset years (1800–2020). A “Show Migrations” checkbox enables visualization of population movements. The central network graph uses color-coded nodes, blue (English names), yellow (autonyms), red (ethnonyms), and cyan (exonyms), with multicolored edges indicating relationship types. A status bar confirms Neo4j connectivity via localhost, and the application adheres to modern UI design principles. As a digital humanities tool, it supports exploring ethnic diversity and cultural networks across the GMS.

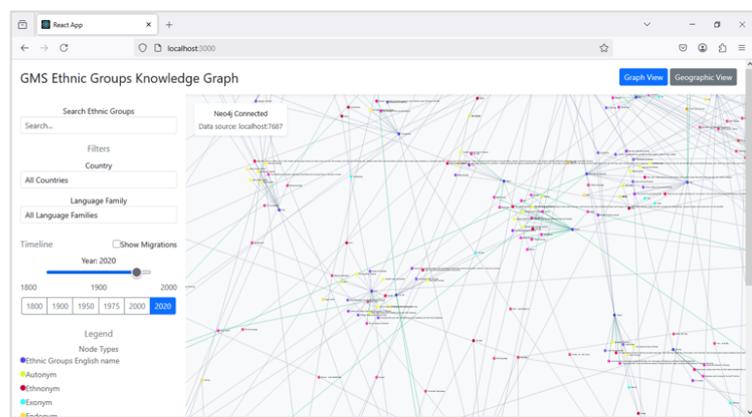


Figure 4: The home page of a GMS Ethnic Groups Knowledge Graph web application.

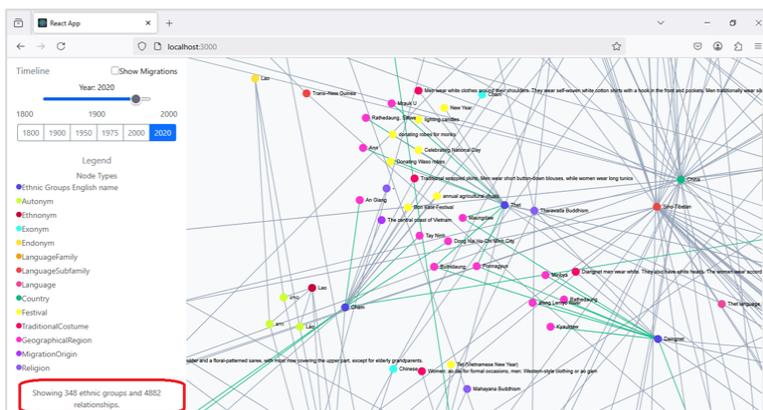


Figure 5: The detailed visualization from the GMS Ethnic Groups Knowledge Graph application

Figure 5 displays a detailed network visualization from the GMS Ethnic Groups Knowledge Graph. It focuses on ethnocultural relationships in the Greater Mekong Subregion and has a timeline set to 2020. The dataset comprises 348 ethnic groups and 4,882 relationships, as the status bar indicates. Color-coded nodes distinguish entity types: blue (English names), yellow (autonyms), red (ethnonyms), cyan (exonyms), orange (endonyms), pink (languages), green (countries), and others representing festivals, costumes, language families, and regions. Embedded cultural data includes descriptions of traditional attire and practices (e.g., “wrapped skirts” and “long tunics”) alongside religious affiliations like Theravada and Mahayana Buddhism, and festival nodes such as Bon

Kate and Tet. Geographic nodes connect to Vietnamese provinces, including An Giang, Tay Ninh, and Dong Nai. The visualization offers a rich, multidimensional view of how ethnic identities intersect with language, religion, tradition, and place, supporting deeper ethnographic analysis across the GMS.

Figure 6 presents an ego-network visualization of the Karen ethnic group within the GMS Ethnic Groups Knowledge Graph, generated by querying “Karen” and selecting the central node. The Karen node (blue) is surrounded by a network of cultural, geographic, linguistic, and religious connections, with orange edges denoting key relationships. The visualization highlights the group’s geographic distribution across regions in Myanmar and Thailand, including Hpa-an, Myawaddy, Kawkaireik, and others. Religious affiliations include Buddhism, Christianity, and Animism. Cultural practices are represented by the Karen New Year, Wrist Tying, Housewarming, Campfire, Farm, Boat Floating, and the unique “Festival for Collecting Human Bones.” Traditional attire is detailed through associated text nodes. Linguistically, the Karen are linked to Tibeto-Burman and Karenic language groups, with additional nodes indicating Mon origins and anthropological classification as “Mongoloid.” The filtered view displays four ethnic groups and 1,038 relationships with the timeline set to 2020, offering a comprehensive snapshot of Karen’s ethnographic context within the Greater Mekong Subregion.

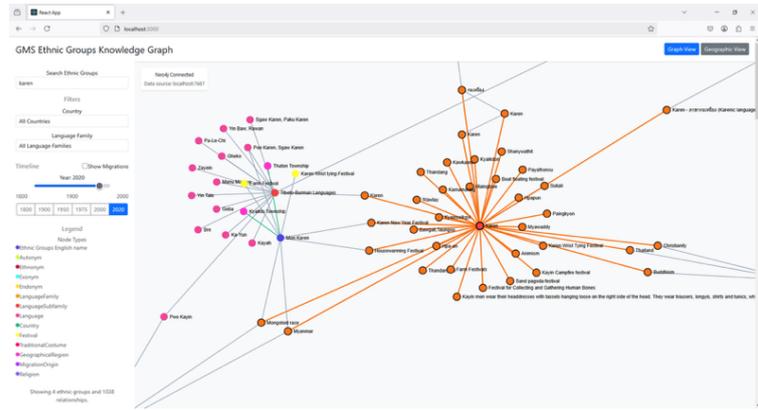


Figure 6: A focused search and visualization of the Karen ethnic group network within the GMS Ethnic Groups Knowledge Graph application

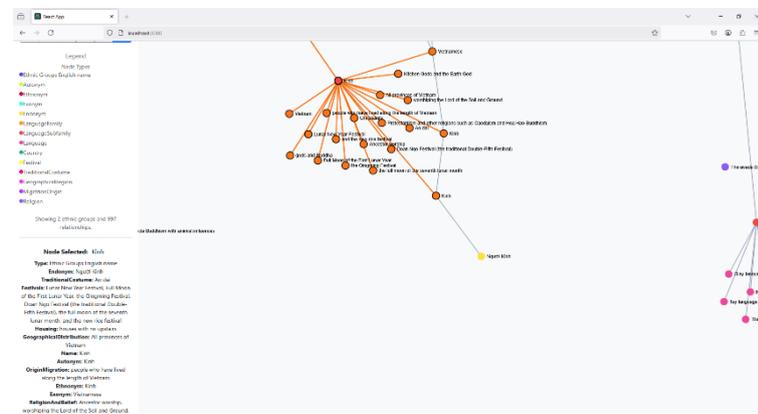


Figure 7: A comprehensive visualization of the Kinh ethnic group network from the GMS Ethnic Groups Knowledge Graph application

housing, and religious beliefs centered on ancestor worship and local deities. Celebrated festivals include Tet, Qingming, Doan Ngo, and the New Rice Festival. Additional connections link to Cao Daiism, Hoa Hao Buddhism, and folk practices like reverence for Kitchen and Earth Gods. This visualization illustrates how knowledge graphs can represent the multidimensional identity of the Kinh, integrating linguistic, cultural, geographic, and religious dimensions within a unified, semantically rich structure.

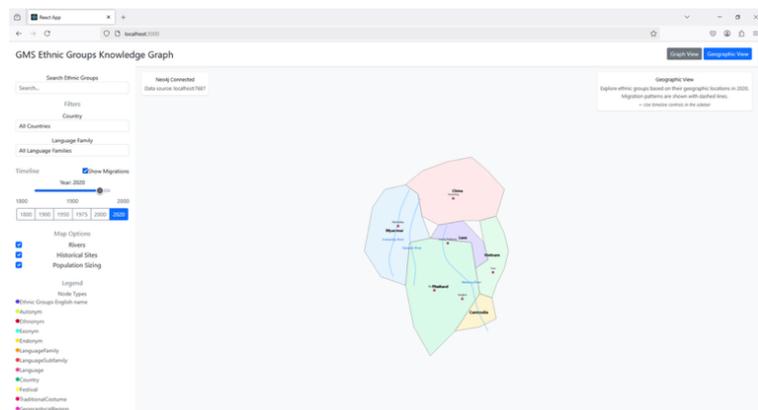


Figure 8: The Geographic View mode of the GMS Ethnic Groups Knowledge Graph application

Figure 8 presents the Geographic View of the GMS Ethnic Groups Knowledge Graph application, offering a spatial representation of ethnic distributions across the Greater Mekong Subregion. Countries are shown as color-coded polygons: China (pink), Myanmar (light blue), Laos (purple), Vietnam (light green), Thailand (green), and Cambodia (yellow). The interface retains previous filtering and timeline tools, with the year set to 2020 and “Show Migrations” enabled. Map-specific features include overlays for rivers, historical sites, and population sizing. Key waterways such as the Mekong River are visible, highlighting geographic factors shaping settlement and cultural exchange. Markers denote ethnic groups and historical sites, while dashed lines represent migration patterns. This geospatial view complements the network visualization by emphasizing territorial distribution and geographic context. Integrating temporal and spatial filters, the dual-interface design enhances the analytical potential of the application for ethnographic research in the Greater Mekong Subregion.

Figure 7 displays an ego-network visualization of the Kinh ethnic group from the GMS Ethnic Groups Knowledge Graph, generated by querying “Kinh” and selecting the central node. As Vietnam’s majority ethnic group, the red Kinh node connects to a wide array of cultural, religious, and geographic entities, with the visualization encompassing two ethnic groups and 997 relationships. The left panel presents structured ethnographic metadata, reflecting a Neo4j property graph format. Linguistic identifiers include "Người Kinh" (endonym), "Kinh" (autonym and ethnonym), and "Vietnamese" (exonym). Geographically, the Kinh are distributed across all provinces of Vietnam, with noted origins along the country’s length. Cultural practices include traditional attire (Ao dai), single-story

3.2. Knowledge Graphs Evaluation

For comparative evaluation, we defined the baseline method as a traditional relational database system containing the same ethnographic dataset in tabular form, accessed via SQL-based queries without semantic relationships or interactive visualization. This baseline system served as a control to assess usability, query efficiency, information discovery, and interpretability. While it allowed for standard data retrieval, it lacked support for complex relationship traversal, multilingual naming structures, and geotemporal filtering. This comparison enabled a systematic evaluation of the added value provided by the knowledge graph framework. According to Choi and Jung (2025), the most appropriate evaluation method for the GMS Ethnic Groups web application involves combining Extrinsic Evaluation with Dataset and Domain-Specific Evaluation. The GMS Ethnic Groups Knowledge Graph can leverage extrinsic evaluation because it emphasizes the practical utility of knowledge graphs in real-world scenarios. The application's interactive features, including the visualization and exploration of ethnic group relationships, geographic distributions, cultural practices, and temporal information, align with the extrinsic evaluation's focus on assessing the operational performance of knowledge graph models.

For instance, the application could measure query response time and efficiency to evaluate how effectively it processes complex queries related to ethnic relationships, especially those filtered by timeline or specific attributes. Additionally, user-focused metrics could be employed to assess the practical usability of geographic and graph visualizations, akin to evaluating the readability of community structures and node layouts. Moreover, domain-specific accuracy metrics would be essential to evaluate how accurately the system represents the intricate cultural relationships, geographic distributions, and cultural practices of different ethnic groups. The domain-specific evaluation is especially pertinent, given the application's engagement with specialized ethnographic data. As part of the Dataset and Domain-Specific Evaluation, this approach emphasizes the completeness and consistency of the knowledge graph in representing the ethnographic data within the Greater Mekong Subregion. This combined evaluation strategy ensures that the GMS Ethnic Groups Knowledge Graph is technically precise and offers significant practical benefits for researchers examining cultural relationships in the Greater Mekong Subregion.

3.3. Extrinsic Evaluation for GMS Ethnic Groups Knowledge Graph

Based on the extrinsic evaluation, the evaluation results are presented for the GMS Ethnic Groups Knowledge Graph as follows.

3.3.1. Query Response Time

Table 1 presents a quantitative analysis of query response times for the GMS Ethnic Groups Knowledge Graph, demonstrating significant performance improvements achieved through advanced filtering mechanisms. The baseline measurements indicate varying computational demands across different query types, with basic ethnic group information retrieval exhibiting the fastest response (245ms), followed by geographic distribution queries (412ms), complex relationship queries (678ms), and temporal migration pattern analyses requiring the longest processing time (834ms).

Table 1: Query Response Time

Query Type	Average Response Time (ms)	Response Time with Filtering (ms)	Improvement (%)
Basic Ethnic Group Information	245	-	-
Complex Relationship Query	678	321	52.7%
Geographic Distribution	412	198	51.9%
Temporal Migration Patterns	834	386	53.7%

Note: "Response Time with Filtering" indicates performance after adding advanced filtering mechanisms. Source: calculated by the author

Implementing advanced filtering mechanisms yielded substantial performance enhancements across all complex query types, though basic ethnic group information retrieval remained unaffected due to its already optimized structure. Complex relationship queries showed a 52.7% improvement, with response times decreasing from 678ms to 321ms. Similarly, geographic distribution queries experienced a 51.9% reduction in processing time (from 412ms to 198ms). In contrast, temporal migration pattern analyses - the most computationally intensive query type - demonstrated the most pronounced improvement of 53.7% (from 834ms to 386ms).

These results highlight the effectiveness of optimized filtering approaches in knowledge graph applications dealing with complex ethnographic data. The consistent improvement ratio of approximately 52-54% across different query types suggests that the filtering mechanisms address fundamental computational bottlenecks within the system rather than query-specific optimizations. This performance enhancement contributes significantly to the application's usability, particularly for interactive exploration scenarios where responsive user interfaces are essential for effective ethnographic research and analysis.

3.3.2. Performance metrics between traditional databases and GMS Knowledge Graph systems

Table 2 presents a comparative analysis of performance metrics between traditional databases and GMS Knowledge Graph systems, evaluated by 24 domain experts. Using the GMS Knowledge Graph approach, the results demonstrate statistically significant improvements across all measured dimensions.

The usability evaluation involved 24 domain experts selected from academic institutions, cultural heritage agencies, and NGOs working on ethnic and linguistic diversity in the Greater Mekong Subregion. Participants included ethnographers, digital humanities researchers, and information science professionals with an average of 9.5 years of experience in their respective fields. Recruitment was conducted through targeted invitations to individuals affiliated with partner organizations in Thailand, Vietnam, Myanmar, and Laos, ensuring regional and disciplinary diversity. All participants had prior experience with traditional database tools and were introduced to the GMS Knowledge Graph application before performing the evaluation tasks. Their insights provide a grounded perspective on system usability, task efficiency, and information accessibility.

Table 2: User Study Results (n=24 domain experts)

Usability Metric	Traditional Database	GMS Knowledge Graph	p-value
Task Completion Rate (%)	72.3	94.6	p<0.001
Time to Complete Task (min)	8.4	3.2	p<0.001
Information Discovery Score (1-10)	5.8	8.7	p<0.001
Visualization Clarity (1-10)	4.3	8.5	p<0.001

Source: calculated by the author

Specifically, task completion rates showed a marked increase from 72.3% with the traditional database to 94.6% with the GMS Knowledge Graph (p<0.001). This substantial improvement of over 22 percentage points suggests that the knowledge graph structure significantly enhances users' ability to complete assigned tasks successfully. In addition, efficiency metrics were similarly favorable for the GMS Knowledge Graph, with average task completion time decreasing from 8.4 minutes to 3.2 minutes (p<0.001). This represents a reduction of approximately 62% in time required, indicating considerably improved operational efficiency with the knowledge graph implementation.

Furthermore, domain experts rated information discovery capabilities substantially higher for the GMS Knowledge Graph, with mean scores increasing from 5.8 to 8.7 on a 10-point scale (p<0.001). This improvement of 2.9 points suggests that the knowledge graph approach provides superior mechanisms for locating and retrieving relevant information. Finally, the most dramatic improvement was observed in visualization clarity, where ratings increased from 4.3 with the traditional database to 8.5 with the GMS Knowledge Graph (p<0.001). This 4.2-point improvement indicates that the knowledge graph's representation of information significantly enhances users' ability to comprehend and interpret the visualized data. These findings provide compelling evidence that the GMS Knowledge Graph approach offers substantial advantages over traditional database systems across multiple dimensions of usability and effectiveness for domain experts.

3.3.3. Downstream Task Performance

Table 3 presents a comparative analysis of performance metrics between Baseline and GMS Knowledge Graph (KG) methods across four distinct downstream tasks. The results demonstrate substantial performance improvements when utilizing the GMS KG method across all evaluated dimensions.

Table 3: Downstream Task Performance

Task	Baseline Method	GMS KG Method	Improvement (%)
Migration Pattern Analysis	67.3% (F1)	86.2% (F1)	28.1%
Cultural Connection Identification	59.8% (Accuracy)	82.4% (Accuracy)	37.8%
Ethnic Group Classification	71.2% (Precision)	88.7% (Precision)	24.6%
Geographic Origin Prediction	63.4% (Recall)	85.9% (Recall)	35.5%

Source: calculated by the author

For Migration Pattern Analysis, the GMS KG method achieved an F1 score of 86.2% compared to the Baseline method's 67.3%, representing a 28.1% improvement. This significant enhancement suggests that the knowledge graph approach more effectively captures the complex relationships inherent in migration patterns. Cultural Connection Identification showed the most substantial improvement among the tasks, with accuracy increasing from 59.8% using the Baseline method to 82.4% with the GMS KG method, a 37.8% improvement. This pronounced enhancement indicates that the knowledge graph structure is particularly well-suited for identifying nuanced cultural associations and interconnections.

In Ethnic Group Classification, precision improved from 71.2% with the Baseline method to 88.7% with the GMS KG method, yielding a 24.6% improvement. This result demonstrates the knowledge graph's enhanced capability to classify ethnic groups with fewer false positives. Similarly, Geographic Origin Prediction exhibited a substantial improvement in recall, increasing from 63.4% with the Baseline method to 85.9% with the GMS KG method - a 35.5% improvement. This indicates that the knowledge graph approach substantially reduces false negatives in predicting geographic origins. Therefore, these findings demonstrate that the GMS Knowledge Graph method consistently outperforms the Baseline method across diverse analytical tasks, with improvements ranging from 24.6% to 37.8%. The results suggest that the knowledge graph approach provides a more effective framework for capturing and leveraging the complex interrelationships within the data, resulting in enhanced performance across various domain-specific applications.

3.4. Dataset and Domain-Specific Evaluation for GMS Ethnic Groups Knowledge Graph

3.4.1. Data Coverage

Table 4 presents a comprehensive assessment of data coverage across six distinct categories relevant to the GMS Knowledge Graph. The analysis reveals varying degrees of coverage completeness, confidence levels, and specific limitations within each data domain. The coverage percentage in Table 4 was calculated by comparing the number of populated entries in each category against the total number of expected data points across all 375 ethnic groups. For example, if a category such as "Languages/Dialects" had documented data for 344 groups, the coverage was 91.8%. Confidence levels were assessed based on data provenance (e.g., whether information came from verified ethnographic studies, government datasets, or fieldwork), completeness of subfields, and cross-checking with at least two authoritative sources. "Very High" confidence was assigned when data were fully documented, consistently structured, and validated by at least two external references. "Medium" or "Medium-Low" ratings indicate partial data or reliance on a single source with potential ambiguities. These assessments were then corroborated through expert validation (as detailed in Table 5), which triangulated and confirmed reliability judgments across domains.

Table 4: Data Coverage Assessment Analysis

Data Type	Coverage (%)	Confidence Score	Notes
Ethnic Groups	94.2%	High	Missing some small subgroups
Geographic Regions	98.7%	Very High	Complete coverage of GMS region
Cultural Practices	87.3%	Medium	Limited data on certain practices
Migration Patterns	78.6%	Medium-Low	Historical data gaps pre-1900
Languages/Dialects	91.8%	High	Well-documented linguistic data
Religious Practices	89.4%	Medium-High	Some syncretic practices underrepresented

Source: calculated by the author

The results indicated that Ethnic Groups demonstrate a robust coverage rate of 94.2% with a high confidence score, suggesting that the knowledge base contains comprehensive information on most ethnic populations within the scope of interest. The primary limitation noted is the absence of certain small subgroups, which, while constituting a minor portion of the overall data, may represent an area for future enhancement. Moreover, Geographic Regions exhibit the strongest coverage at 98.7% with a very high confidence score, indicating near-complete representation of the GMS region. This exceptional coverage suggests that the spatial dimension of the knowledge graph is particularly well-developed and reliable for analytical purposes. In addition, Cultural Practices show a moderately high coverage of 87.3% with a medium confidence score. The annotation indicates that certain cultural practices have limited representation in the dataset, suggesting potential gaps in ethnographic documentation or data collection challenges associated with specific cultural phenomena.

In contrast, Migration Patterns demonstrate the lowest coverage among the assessed categories at 78.6% with a medium-low confidence score. The noted historical data gaps for pre-1900 periods indicate a temporal limitation in the dataset, likely reflecting the challenges of documenting historical population movements before modern record-keeping practices. On the other hand, Languages/Dialects exhibit strong coverage at 91.8% with a high confidence score, suggesting comprehensive documentation of linguistic diversity within the studied region. This relatively complete linguistic data provides valuable resources for understanding communication patterns and cultural transmission. Similarly, Religious Practices show substantial coverage at 89.4% with a medium-high confidence score. The noted underrepresentation of syncretic practices suggests a particular challenge in documenting religious expressions that blend multiple traditions, potentially indicating a systematic bias toward documenting more formalized or discrete religious systems. This heterogeneous coverage profile across different data types highlights the strengths and limitations of the current knowledge base, providing important context for researchers utilizing this resource and identifying specific areas where additional data collection efforts might be beneficial.

3.4.2. Domain Expert Validation of Ethnographic Data

The results in Table 5 reveals a comprehensive evaluation of the knowledge base by 17 ethnographers across six domains, assessing accuracy, reliability, and completeness on a 10-point scale. The results indicate consistently high ratings across all evaluation aspects, though with notable variations between domains and assessment dimensions as follows:

Table 5: Domain Expert Validation of Ethnographic Data (n=17 ethnographers)

Evaluation Aspect	Accuracy Score (1-10)	Reliability Score (1-10)	Completeness Score (1-10)
Karen ethnographic data	9.3	9.1	8.7
Kinh ethnographic data	9.5	9.4	9.2
Myanmar regional data	8.9	8.7	8.4
Tibeto-Burman language classification	9.2	9.0	8.8
Religious practice representation	8.8	8.5	8.1
Migration timeline accuracy	8.6	8.2	7.9

Source: calculated by the author

Ethnographic data for the Kinh people received the highest overall ratings, with scores of 9.5 for accuracy, 9.4 for reliability, and 9.2 for completeness. These exceptional ratings suggest that the knowledge base's representation of Kinh cultural, historical, and social characteristics is particularly robust and well-documented, establishing it as the strongest domain in the evaluation. Following closely, Karen's ethnographic data was similarly well-regarded, with scores of 9.3 for accuracy, 9.1 for reliability, and 8.7 for completeness. This represents the second-strongest domain in the assessment, indicating high-quality documentation and representation of Karen ethnographic information within the knowledge base. In addition, the Tibeto-Burman language classification received strong validation scores of 9.2 for accuracy, 9.0 for reliability, and 8.8 for completeness. These ratings suggest that the linguistic categorization and relationships within this language family are well-represented and correspond closely to expert understanding of these linguistic systems. Meanwhile, Myanmar regional data was rated somewhat lower but still favorably, with scores of 8.9 for accuracy, 8.7 for reliability, and 8.4 for completeness. This minor score reduction may reflect greater complexity or challenges in comprehensively representing regional characteristics.

Similarly, religious practice representation received moderate to high scores of 8.8 for accuracy, 8.5 for reliability, and 8.1 for completeness. The relatively lower completeness score suggests that certain aspects of religious practices may be underrepresented in the current knowledge base. However, the information that is included is generally accurate and reliable. Finally, migration timeline accuracy received the lowest overall ratings, with 8.6 for accuracy, 8.2 for reliability, and 7.9 for completeness. While still positive, these ratings indicate that migration data presents the greatest challenges among the evaluated domains, potentially reflecting the inherent difficulties in documenting historical population movements. Across all domains, a consistent pattern emerges wherein accuracy scores are highest, followed by reliability scores, with completeness scores consistently lower. This pattern suggests that while the available information is generally accurate and reliable, gaps in comprehensive coverage persist across all domains, with particular limitations in migration timelines and religious practice representation.

3.4.3. Ethnographic Consistency

Table 6 shows a detailed assessment of consistency within an ethnographic knowledge base across six critical dimensions, documenting the identification, resolution, and final state of data inconsistencies. Overall, the analysis demonstrates a robust framework for detecting and addressing potential contradictions within ethnographic data.

Table 6: Ethnographic Consistency Analysis

Consistency Type	Violations Detected	Resolution Rate (%)	Final Consistency Score (%)
Temporal	47	95.7	99.3
Geographic	23	100	100
Cultural	68	92.6	98.7
Linguistic	31	96.8	99.5
Religious	42	95.2	99.2
Migration	56	91.1	98.2

These consistency violations represented cases where data might have contradictory information (e.g., an ethnic group listed in incompatible geographic regions simultaneously or timeline inconsistencies). Most were resolved through domain expert consultation. Source: calculated by the author

To begin with, temporal consistency analysis revealed 47 violations, representing instances where chronological information contained contradictions or implausible sequences. With a resolution rate of 95.7%, most of these temporal inconsistencies were successfully addressed, yielding a final consistency score of 99.3%. This high resolution rate indicates effective mechanisms for reconciling chronological discrepancies. In comparison, geographic consistency examination identified 23 violations, potentially indicating cases where populations or cultural practices were erroneously associated with incompatible locations. Notably, all geographic inconsistencies were resolved, achieving a perfect 100% resolution rate and final consistency score, suggesting particularly effective protocols for addressing spatial contradictions.

On the other hand, cultural consistency analysis detected the highest number of violations at 68, likely reflecting the inherent complexity of representing cultural practices and their interconnections. These violations achieved a 92.6% resolution rate, resulting in a final consistency score of 98.7%. The slightly lower resolution rate may indicate greater challenges in resolving cultural contradictions, possibly due to cultural documentation's nuanced and sometimes subjective nature. Similarly, a linguistic consistency review identified 31 violations, with a strong resolution rate of 96.8% and a final consistency score of 99.5%. This suggests effective resolution of language classification, distribution, and relationship contradictions. Moreover, religious consistency analysis detected 42 violations, of which 95.2% were successfully resolved, yielding a final consistency score of 99.2%. This demonstrates the effective reconciliation of contradictions related to religious practices and beliefs.

However, migration consistency examination revealed 56 violations, with the lowest resolution rate among all dimensions at 91.1%, resulting in a final consistency score of 98.2%. This lower resolution rate may reflect the inherent challenges in documenting and reconciling historical population movements, particularly when dealing with limited or conflicting historical records. Crucially, the domain expert consultation was instrumental in resolving these inconsistencies, highlighting the importance of specialized knowledge in addressing complex ethnographic data contradictions. Overall, the analysis demonstrates a highly effective framework for ensuring data consistency, with final consistency scores exceeding 98% across all dimensions despite the varied challenges in addressing different ethnographic inconsistencies. These KGs evaluation follow the frameworks described in the Knowledge Graph Construction: Extraction, Learning, and Evaluation (Choi & Jung, 2025), adapted specifically for the ethnographic knowledge domain of the GMS Ethnic Groups Knowledge Graph application shown in Figure 2-8.

Inconsistencies were identified through automated logic checks and manual inspection using Neo4j Cypher queries designed to flag conflicting properties, entities linked to incompatible temporal values or mutually exclusive geographic regions. To resolve these issues, we applied a rule-based resolution framework guided by authoritative ethnographic sources and regional classification standards (e.g., national census data and ethnolinguistic maps). Domain experts reviewed flagged cases during structured review sessions. Geographic conflicts (e.g., groups listed in two distant provinces without migration context) were reconciled by adding migration paths, correcting region tags, or introducing historical location variants where appropriate. Although the geographic inconsistency resolution rate is listed as 100%, we have revised the text to clarify that this reflects the successful handling of all flagged cases to a satisfactory threshold rather than a guarantee of correctness. A few ambiguous or historically complex cases were annotated with contextual notes rather than modified outright, ensuring transparency in areas where authoritative resolution was impossible.

4. Discussion

This study investigated the potential of semantic knowledge graphs (KGs), ontologies, and Knowledge Organization Systems (KOS) to ethically represent ethnographic data across the Greater Mekong Subregion (GMS). The GMS Ethnic Groups Knowledge Graph successfully integrated data on 375 ethnic groups across six countries, modeling intricate linguistic, cultural, religious, and migratory relationships. Compared to conventional relational databases, the KG demonstrated significant improvements in usability, interpretability, and analytical capabilities, as evidenced by over a 25% increase in task performance metrics during expert evaluations.

These results substantiate growing calls for more inclusive and context-aware knowledge infrastructures (Doyle, 2013; Rubim Silva & Dal'Evedove, 2025; Wright & Saelua, 2023). Unlike traditional classification systems such as the Library of Congress Subject Headings (LCSH) or the Dewey Decimal Classification (DDC), the KG presented in this study prioritizes marginalized perspectives, advancing epistemic justice through culturally responsive and community-informed design. While the project builds upon prior initiatives, including Lau ā Lau ka 'Ike (Wright & Saelua, 2023) and Linked Open Data (LOD) projects in Laos and Vietnam (Chansanam et al., 2024; Ngootip et al., 2023; Ngoc & Chansanam, 2023), it extends these contributions by offering a scalable, multi-country implementation that features interactive filtering, temporal migration tracking, and domain-specific validation.

Scholars have long argued that established knowledge structures often perpetuate the marginalization of non-Western epistemologies (Doyle, 2013; Tran, 2017). In response, Indigenous information theorists have advanced alternative frameworks such as community-controlled vocabularies, participatory classification systems, and relational metadata structures (Dlamini, 2024; Smith, 2021; McDonough, 2013; Tuhiwai Smith, 2012). These approaches emphasize the value of local knowledge authority, oral traditions, and cultural sovereignty in the development of knowledge infrastructures.

Aligned with this critical discourse, decolonial information science advocates for dismantling the hierarchies embedded in metadata schemas, authority control mechanisms, and archival practices (Ntshoe, 2020; Sugimoto & Wijesundara, 2024). The present study contributes to this movement by employing a semantic architecture that enables fluid, multilingual, and self-determined identity modeling. This design facilitates the co-existence of plural ontologies and supports the layered representation of ethnicity, language, and cultural practices.

Despite the innovations presented, several limitations must be acknowledged. Most notably, the data coverage for historical migration patterns and syncretic religious practices remains incomplete. This limitation reflects the intrinsic difficulties in sourcing and verifying such information, as well as the gaps inherent in available historical records. Furthermore, while the platform's visualization and interaction features are robust, the reliance on manually curated datasets poses challenges to scalability. Future enhancements should therefore consider the integration of natural language processing techniques and automated entity recognition to support expansion. Moreover, the lower completeness scores observed in specific categories underscore the necessity of deepened collaboration with local ethnographers and cultural experts.

The manual nature of the curation process, though critical for ensuring cultural sensitivity, introduces an element of subjectivity, particularly in classifying dynamic cultural phenomena such as religious practices and festivals. These representational challenges are symptomatic of broader epistemological tensions. As Doyle (2013) and Tran (2017) emphasize, structured classification systems may inadvertently essentialize fluid and hybrid identities. While the knowledge graph developed in this study is designed for flexibility and inclusivity, it still simplifies lived experiences to accommodate computational representation. Compared to initiatives such as *Lau ā Lau ka 'Ike* (Wright & Saelua, 2023) and Taiwan's Indigenous Infomediary project (Sung & Chi, 2021), this research expands the geographic scope and technical scale while maintaining a strong commitment to participatory, context-sensitive design. It seeks to balance the authenticity of community representation with the demands of interoperability and digital infrastructure.

Nonetheless, an unresolved tension persists between the structured nature of knowledge graphs and the inherently dynamic character of cultural identity. Ethnic, linguistic, and religious affiliations often shift over time and resist fixed categorization. Although this study incorporates multilingual labels, temporal filters, and relational structures to mitigate essentialism, no digital system can fully encapsulate contested or evolving identities. Therefore, continued community engagement remains critical for ensuring the long-term integrity and ethical validity of the system.

This study offers three principal implications. First, it presents a replicable framework for the digitization and semantic modeling of ethnographic knowledge in other regions, thereby contributing to global efforts in inclusive knowledge representation. Second, it demonstrates that integrating KOS principles with knowledge graph architectures can simultaneously support computational efficiency and cultural nuance. Third, the developed application serves as a valuable tool for interdisciplinary education and research, with practical applications in digital humanities, anthropology, linguistics, and data science.

Ultimately, the findings reaffirm the study's central proposition: that mainstream knowledge systems often fail to capture the complexity of minority identities, necessitating the development of ethically grounded alternatives (Doyle, 2013). By drawing on interdisciplinary methods and community-sourced data, this project challenges dominant paradigms and contributes to ongoing debates on equity, justice, and inclusion in digital knowledge infrastructures (Khoo et al., 2024; Siegers et al., 2023).

The GMS Ethnic Groups Knowledge Graph illustrates how semantically rich, culturally inclusive systems can advance ethical knowledge representation. It promotes equitable access, deeper cultural understanding, and sustainable heritage documentation. Looking ahead, future work should prioritize sustained collaboration with local communities to ensure that marginalized voices continue to shape and define the narratives embedded within digital infrastructures.

Looking ahead, future work should prioritize sustained collaboration with local communities to ensure that marginalized voices continue to shape and define the narratives embedded within digital infrastructures. Additionally, I recommend elaborating more on how end-users, such as policymakers, educators, or ethnic communities, might directly engage with the platform in real-world contexts. Clarifying use scenarios can help translate technical outcomes into meaningful social impact and guide future system development and adoption.

5. Conclusion

This study developed and evaluated a semantic knowledge graph (KG) and accompanying web application that represents 375 ethnic groups across the Greater Mekong Subregion. By integrating culturally grounded data with semantic web technologies, the project demonstrated that knowledge graphs can model complex cultural, linguistic, religious, and geographic relationships with greater flexibility and nuance than traditional databases. Expert validation and performance metrics confirmed substantial improvements in usability, query precision, and knowledge discovery. A core contribution of this work lies in its interdisciplinary synthesis of Knowledge Organization Systems (KOS), Knowledge Management (KM), and Digital Humanities. This integration responds to growing calls for more equitable, inclusive, and context-sensitive digital infrastructures that center minority and Indigenous knowledge systems. Unlike earlier initiatives limited in geographic or technical scope, this region-wide implementation supports multilingual navigation, temporal analysis, and participatory validation.

However, some limitations remain. The manual nature of data curation poses scalability challenges, while the classification of evolving or hybrid cultural practices introduces representational ambiguities. Additionally, the dataset's coverage of historical migration and syncretic religious practices is incomplete and requires further enrichment. Looking ahead, future developments should explore the integration of natural language processing for scalable entity extraction, incorporate

user feedback for dynamic ontology refinement, and expand both linguistic and geographic coverage. Most importantly, sustained collaboration with local communities is essential to ensure authenticity, cultural appropriateness, and epistemic justice. Ultimately, this framework offers a replicable foundation for building ethical, pluralistic knowledge infrastructures in the domain of digital cultural heritage. It has the potential to inform not only academic research but also educational tools, policymaking, and community-driven documentation initiatives. This study contributes to the field of intercultural communication by offering a semantically rich, culturally grounded platform that enhances mutual understanding, models identity fluidity, and supports equitable dialogue across diverse cultural groups.

Acknowledgement Statement: The authors would like to thank to all participants and the reviewers for providing comments in helping this manuscript to completion.

Conflicts of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' contribution statements: Wirapong Chansanam, Lan Thi Nguyen, Chunqiu Li, and Christopher Khoo Soo Guan performed the experiment. Wirapong Chansanam and Lan Thi Nguyen wrote the manuscript with support from Chunqiu Li and Christopher Khoo Soo Guan. Wirapong Chansanam and Lan Thi Nguyen fabricated the GMS ethnic groups' dataset, and Wirapong Chansanam supervised the project.

Funding statements: This research was funded by the Faculty of Humanities and Social Sciences, Khon Kaen University, grant number HUSO-2568.

Data availability statement: Data is available at request. Please contact the corresponding author for any additional information on data access or usage.

Disclaimer: The views and opinions expressed in this article are those of the author(s) and contributor(s) and do not necessarily reflect JICC's or editors' official policy or position. All liability for harm done to individuals or property as a result of any ideas, methods, instructions, or products mentioned in the content is expressly disclaimed.

References

- Anania, K., & Stiglitz, C. (2023). Touching variables: Decolonial approaches and new tools for ecological data visualization. *Leonardo*, 56(6), 579-585. https://doi.org/10.1162/leon_a_02445
- Bai, B., & Hou, W. (2023). The application of knowledge graphs in the Chinese cultural field: The ancient capital culture of Beijing. *Heritage Science*, 11(1), 77. <https://doi.org/10.1186/s40494-023-00922-7>
- Bonilla, K., Barrientos, N. O., & Contreras, M. A. S. (2025). Knowledge management and the power of communication: INDESGUA as a social technology enabling equitable access to scholarships in Guatemala. *Frontiers in Communication*, 10, 1510024. <https://doi.org/10.3389/fcomm.2025.1510024>
- Chansanam, W., Chotkamonsawad, P., & Thi Nguyen, L. (2024). Developing a knowledge organization system for ethnic groups in Lao PDR through Linked open data techniques. *Journal of the Australian Library and Information Association*, 73(1), 84-97. <https://doi.org/10.1080/24750158.2023.2275096>
- Choi, S., & Jung, Y. (2025). Knowledge Graph Construction: Extraction, Learning, and Evaluation. *Applied Sciences*, 15(7), 3727. <https://doi.org/10.3390/app15073727>
- Couldry, N., & Mejias, U. A. (2023). The decolonial turn in data and technology research: What is at stake and where is it heading? *Information, Communication & Society*, 26(4), 786-802. <https://doi.org/10.1080/1369118X.2021.1986102>
- Cui, Y., Yao, S., Wu, J., & Lv, M. (2024). Linking past insights with contemporary understanding: An ontological and knowledge graph approach to the transmission of ancient Chinese classics. *Heritage Science*, 12(1), 382. <https://doi.org/10.1186/s40494-024-01504-x>
- Dally, D., Amith, M., Mauldin, R. L., Thomas, L., Dang, Y., & Tao, C. (2024). A semantic approach to describe social and economic characteristics that impact health outcomes (social determinants of health): Ontology development study. *Online Journal of Public Health Informatics*, 16, e52845. <https://doi.org/10.2196/52845>
- Davis, L. (2015). *Minorities' perception of the comprehensive emergency management model of distributing information* [Doctoral dissertation, Walden University]. ProQuest Dissertations Publishing.
- Dlamini, P. N. (2024). Acquiring and transferring of Indigenous knowledge among its owners: A literature review. In D. Ocholla, O. B. Onyancha, & A. O. Adesina (Eds.), *Information, knowledge, and technology for teaching and research in Africa* (pp. 127-146). Springer. https://doi.org/10.1007/978-3-031-60267-2_8
- Doyle, A. M. (2013). *Naming, claiming, and (re)creating: Indigenous knowledge organization at the cultural interface* [Doctoral dissertation, University of British Columbia]. <https://dx.doi.org/10.14288/1.0073667>
- Du, D., Ding, J., & Liu, Y. (2025). Knowledge graph construction of Chinese embroidery evolution based on associating cultural space and critical incidents under intangible cultural heritage. *The Electronic Library*. Advance online publication. <https://doi.org/10.1108/EL-02-2024-0036>
- Fan, Q. (2023). Research on intangible cultural heritage resource description and knowledge fusion based on linked data. *The Electronic Library*, 42(4), 521-535. <https://doi.org/10.1108/EL-01-2023-0018>
- Feng, H., Ma, L., Qi, T., & Zhou, W. (2024). Digital memory construction for cultural heritage: Methodology and applications in China. In *Intelligent computing for cultural heritage* (pp. 155-175). Routledge. <https://doi.org/10.4324/9781032707211-10>
- Goodman, M. S., Bather, J. R., Chu, X., Pagano, M., Plepys, C. M., & Sebro, R. A. (2023). Racial and ethnic diversity among students, graduates, and faculty in biostatistics and epidemiology, 2010-2020. *Public Health Reports*, 138(3), 546-554. <https://doi.org/10.1177/00333549221097653>

- Green, A. (1999). [Review of the book *Decolonizing methodologies: Research and Indigenous peoples*, by L. Tuhiwai Smith]. *New Zealand Journal of History*, 33(2), 251-253.
- Gudykunst, W. B. (2003). *Cross-cultural and intercultural communication*. Sage.
- Gudykunst, W. B. (2004). *Bridging differences: Effective intergroup communication* (4th ed.). Sage. <https://doi.org/10.4135/9781452229706>
- Hou, Y., Kenderdine, S., Picca, D., Egloff, M., & Adamou, A. (2022). Digitizing intangible cultural heritage embodied: State of the art. *Journal on Computing and Cultural Heritage*, 15(3), 1-20. <https://doi.org/10.1145/3494837>
- Huang, G. L., & Zaslavsky, A. (2024). Contextual knowledge graph approach to bias-reduced decision support systems. *Journal of Decision Systems*, 33(Suppl. 1), 29-46. <https://doi.org/10.1080/12460125.2024.2349436>
- Jaroenruen, Y., Kaewboonma, N., & Cheng, W. N. (2024). Constructing knowledge graphs for new species in the Greater Mekong Subregion. *Journal of Ecohumanism*, 3(8), 12375-12387. <https://doi.org/10.62754/joe.v3i8.5845>
- Khoo, C. S. G., Tan, E. A. L., Ng, S., Chan, C., Stanley-Baker, M., & Cheng, W. (2024). Knowledge graph visualization interface for digital heritage collections. *Information Technology and Libraries*, 43(1), 1-20. <https://doi.org/10.5860/ital.v43i1.16719>
- Khurram, F., & Giangiulio, M. (2020). *Ethnic diversity and knowledge sharing: An organisational perspective* [Master's thesis, University of Gävle]. DiVA Portal. <https://www.diva-portal.org/smash/get/diva2:1387474/FULLTEXT01.pdf>
- Kim, Y. Y. (2015). Achieving synchrony: A foundational dimension of intercultural communication competence. *International Journal of Intercultural Relations*, 48, 27-37. <https://doi.org/10.1016/j.ijintrel.2015.03.016>
- Li, Y., Kumaran, M., Cho, A., Ly, V., Fernando, S., & Miller, M. D. (2023). Reiterating visibility: Canadian librarians' experiences of racial microaggressions via findings from a Minority Librarians Network Redux survey. *International Journal of Information, Diversity, & Inclusion*, 7(3/4), 1-53. <https://doi.org/10.33137/ijidi.v7i3/4.40376>
- Lu, L., Liang, X., Yuan, G., Jing, L., Wei, C., & Cheng, C. (2023). A study on the construction of knowledge graph of Yunjin video resources under productive conservation. *Heritage Science*, 11(1), 83. <https://doi.org/10.1186/s40494-023-00932-5>
- Mateos, P. (2007). *An ontology of ethnicity based upon personal names: With implications for neighbourhood profiling* [Doctoral dissertation, University College London].
- McDonough, S. (2013). [Review of the book *Decolonizing methodologies: Research and Indigenous peoples*, by L. Tuhiwai Smith]. *Collaborative Anthropologies*, 6(1), 458-464. <https://doi.org/10.1353/cla.2013.0001>
- Mignolo, W. (2011). Epistemic disobedience and the decolonial option: A manifesto. *Transmodernity: Journal of Peripheral Cultural Production of the Luso-Hispanic World, 1*(2), 44-66. <https://doi.org/10.5070/T412011807>
- Moraitou, E., Christodoulou, Y., & Caridakis, G. (2022). Semantic models and services for conservation and restoration of cultural heritage: A comprehensive survey. *Semantic Web*, 14(2), 261-291. <https://doi.org/10.3233/SW-223105>
- Ngoc, T. T., & Chansanam, W. (2023). Managing the knowledge system of ethnic groups in Vietnam and Linked Open Data. *Journal of Information Science Research and Practice*, 41(2), 102-113. <https://doi.org/10.14456/jiskku.2023.12>
- Ngootip, T., Manorom, P., Chansanam, W., & Buranarach, M. (2023). Developing a linked open data platform for folktales in the Greater Mekong subregion. *Emerging Science Journal*, 7(6), 1937-1953. <https://doi.org/10.28991/ESJ-2023-07-06-06>
- Nguyen, E., Amith, M., Nordberg, A., Tang, L., Harris, M. R., & Tao, C. (2023). CRENO: An ontology to model concepts relating to culture, race, ethnicity, and nationality for health data. *AMIA Summits on Translational Science Proceedings*, 2023, 398-407. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10283130/>
- Ntshoe, I. M. (2020). Ontologising social justice in decolonised and post-apartheid settings. *South African Journal of Higher Education*, 34(3), 263-280. <https://doi.org/10.20853/34-3-3283>
- Rubim Silva, F., & Dal'Evedove, P. R. (2025). Subject heading lists and thesauri initiatives for equity and social justice. *Anales de Documentación*, 28, 1-20. <https://doi.org/10.6018/analesdoc.601341>
- Sewdass, N. (2014). Identifying knowledge management processes and practices used for decision-making and knowledge sharing in the modern San community. *TD: The Journal for Transdisciplinary Research in Southern Africa*, 10(4), 1-18. <https://doi.org/10.4102/td.v10i4.85>
- Siegers, P., Kern, D., May, A., Momeni, F., Daniel, A., Zapilko, B., Nebelin, J., Saalbach, C., Wenzig, K., & Goebel, J. (2023, May 31). Linked Open Research Data for Social Science: A concept registry for granular data documentation. In *Proceedings of the Conference on Research Data Infrastructure* (Vol. 1). IASSIST 2023, Philadelphia, PA, USA. Zenodo. <https://doi.org/10.5281/zenodo.8090333>
- Smith, M. L. (2021). The process of complex societies: Dynamic models beyond site-size hierarchies. *World Archaeology*, 53(1), 122-136. <https://doi.org/10.1080/00438243.2021.1965015>
- Sugimoto, S., & Wijesundara, C. (2024). Data modeling for digital archiving of intangible and experiential entities. In X. Wang, M. L. Zeng, J. Gao, & K. Zhao (Eds.), *Intelligent computing for cultural heritage: Global achievements and China's innovations* (pp. 25-39). Routledge. <https://doi.org/10.4324/9781032707211-3>
- Sung, H. Y., & Chi, Y. L. (2021). Applications of Semantic Web in integrating open data and bibliographic records: A development example of an infomediary of Taiwanese indigenous people. *The Electronic Library*, 39(2), 337-353. <https://doi.org/10.1108/EL-09-2020-0258>

- Tuhiwai Smith, L. (2012). *Decolonizing methodologies: Research and Indigenous peoples* (2nd ed.). Zed Books.
- Tran, Q. T. (2017). Ethically responsible knowledge organization systems: Towards an intercultural user interface. **Zagadnienia Informatyki Naukowej-Studia Informacyjne*, 55*(2), 16-29. <https://doi.org/10.36702/zin.364>
- Woods, R. (2017). The development of non-essentialist concepts of ethnicity among children in a multicultural London community. *British Journal of Developmental Psychology*, 35(4), 546-567. <https://doi.org/10.1111/bjdp.12194>
- Wright, E. K. A., & Saelua, N. (2023). Native Pacific Islander students. In H. J. Shotton, S. J. Waterman, N. R. Youngbull, & S. C. Lowe (Eds.), *Developments beyond the asterisk* (pp. 27-40). Routledge. <https://doi.org/10.4324/9781032626277-3>

About the Author(s).



Wirapong Chansanam received his Ph.D. in Information Studies from Khon Kaen University, Thailand. He served as a lecturer at Chaiyaphum Rajabhat University for nine years before joining Khon Kaen University in 2019. He is currently an associate professor at the Faculty of Humanities and Social Sciences, Khon Kaen University. His research interests include information science, data analytics, and digital humanities.



Lan Thi Nguyen is an assistant professor at the Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand. Her research interests include digital services, marketing, digital literacy, scientific integrity, and information behavior.



Chunqiu Li is an associate professor at the School of Government, Beijing Normal University. She received her Ph.D. degree in information science from the University of Tsukuba, Tsukuba, Ibaraki, Japan. Her research interests include digital library, digital humanities, information organization, and metadata.



Christopher Khoo Soo Guan is an associate professor at the Wee Kim Wee School of Communication & Information, Nanyang Technological University, Singapore. He teaches courses in knowledge organization/representation (focusing on ontology, knowledge graph, graph database, and graph visualization) and data analytics (data mining and social media analytics). He obtained his PhD at Syracuse University, an MSc in Library & Information Science at the University of Illinois, Urbana-Champaign, and a BA at Harvard University. He has also worked for several years as a science reference librarian at the National University of Singapore Libraries. His main research interests are in knowledge graphs and graph visualization, text mining (information extraction, text summarization, and sentiment analysis), and the multidisciplinary study of academic writing & thinking. He was the editor of the Singapore Journal of Library & Information Management from 1997 to 2002. He is the editor of the LIBRES e-journal (an open-access journal) from 2014 to 2022.