Fuzzy Systems and Data Mining VII
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA210190

Open Access Digital Thesaurus on Ethnic Groups in the Mekong River Basin

Wirapong Chansanam[a,1], Kanyarat Kwiecien[a], Marut Buranarach[b] and Kulthida Tuamsuk[a]

[a] *Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Khon Kaen 40002, Thailand*

[b] *National Electronics and Computer Technology Center, Pathumthani 12120, Thailand*

Abstract. The ethnic group domain, in particular, is characterized by rich and diverse data sets in the Mekong River Basin (MRB). Ethnic groups' vocabulary and relevant data come from various sources that cross history, language, and geography. As a result, distinct language is used by specialized groups to characterize their artifacts. Data interoperability among multiple catalogs is highly challenging as a result of this. The usage of controlled vocabularies and thesauri is generally considered a major practice for making preparations for standardization, which is essential for data reuse and sharing. In contrast, when used together, thesauri eliminate ambiguity in natural language, making it easier to identify and integrate data from different sources and allow scholars and computer programs to understand data more efficiently. This paper describes the modeling process of the EGMRB Thesaurus, its integration and role in the infrastructure, its publication as Linked Open Data, and the results of this work after six months of development. This paper presents the rationale behind the realization of this thesaurus. Thesaurus EGMRB (http://thesaurus.asiana.net/vocab/) provides a semantic resource on ethnic groups in the Mekong river basin. EGMRB is the outcome of interdisciplinary cooperation of specialists from the domains of ethnic groups and information science, who collaborated in the context of collaborative research. The thesaurus was developed in Simple Knowledge Organization System (SKOS), a standard data format based on the Resource Description Framework (RDF), using semantic web standard technologies. EGMRB is freely available online, with a SPARQL endpoint (http://thesaurus.asiana.net/vocab/sparql.php) for querying and an API (http://thesaurus.asiana.net/vocab/services.php) for system integration. Digital collections, digital exhibits, and a virtual study environment are being built as part of a digital platform that will give scholars and the general users search and content curation services. EGMRB, which provides unified ideas with related unique and resolvable URIs, can profoundly reduce the barriers to data discovery, integration, and sharing if adopted as a standard and carefully implemented and expanded by the academic community.

Keywords. Thesaurus, Mekong River Basin, Digital Platform, Linked Data, Web service

[1] Corresponding Author, Wirapong Chansanam, Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Khon Kaen 40002; E-mail: wirach@kku.ac.th.

1. Introduction

An online platform providing search and content curation services will be available for free to anybody who wishes to use it via the EGMRB Thesaurus. This will feature digital displays, digital collections, and access to a virtual study environment for scholars and the general users. The EGMRB Thesaurus infrastructure will be used to help project researchers in the development of data open access management strategies and to disseminate any data by the KOS principles. According to [1] DCAT 2 (Data Catalogue Vocabulary) specifications like the Getty Thesaurus will be easier to locate and utilize because of the DCAT 2 specifications.

Nowadays, information can be transferred rapidly and simply according to such new technologies as the Internet and electronic data records. To conduct effective research and analysis, users must have access to the appropriate datasets and efficient methods for properly explaining and exploring basic concepts inside and across datasets [2, 3]

Metadata, controlled vocabularies, thesauri, or even ontologies are used in this method since they complement one another [4]. A thesaurus is a controlled and structured vocabulary that contains concepts by terms. It is delivered so that concepts are interconnected so that links between them may be made clear. Preferred terms are accompanied with comes entries for synonyms or "quasi-synonyms" [5]. By reducing ambiguity in plain language, thesauri are essential method for discovering and organizing information, especially for discovering material in non-literary, uncategorized formats. They help academics and the public to discover data and other materials consistently.

The process of building a thesaurus is rather complicated. It involves identifying terms from a specific area, analyzing and linking those phrases into a categorization model that can be used to classify resources from the same domain. The form and characteristics of monolingual and multilingual thesauri have changed throughout the years to normalize these structures [5-8]. ISO 25964, the most current international standard for thesauri, describes five fundamental kinds of thesaurus ideas, thesaurus terms, thesaurus arrays, and notes and sets forth a data format and derived XML schema. There is an explicit distinction between “terms” and “concepts” in the data model. The data model includes information on additional elements that might be considered a part of a thesaurus, including ConceptGroup, a unique grouping for terms related to a particular domain, theme, or other categories [5-8]. While it does not have a specific building technique proposed, the ISO 25964 does not offer specific advice on developing and maintaining thesauri but does provide guidelines for both aspects.

Initiatives in development for digital platform assistance for content discovery and management will be using the EGMRB Thesaurus. A helpful way to share and link to controlled vocabularies in the Semantic Web is by using the W3C recommendation, the SKOS (Simple Knowledge Organization System), as a model [9]. The thesaurus of the EGMRB is detailed in this article, including how it functions inside the infrastructure and as Linked Open Data (LOD). The second section of this paper will consider similar work that has modeled thesauri and other controlled vocabularies and other research infrastructures and LOD initiatives in the framework of other projects and initiatives addressing specific ethnic group. Section 3 will discuss the methods and tools used to manage and publish vocabularies in the infrastructure. This section details our preliminary findings. Our findings and recommendations conclude with this section.

2. Related work

In GLAM (Galleries, Libraries, Archives, and Museums) applications, subject indexing, keyword-based search in local databases, and federated search across multiple databases [10, 11] are extensively used through the use of thesauri and other knowledge organization systems, such as taxonomies, classification schemes, and classifications. Within LOD efforts, it is estimated that SKOS is one of the most often used vocabularies and is widely utilized by other Semantic Web vocabularies. Using controlled vocabularies for Semantic Web applications such as query expansion, search term recommendations, and semantic search engines is possible through the ‘skosification' of those vocabularies [12, 13, 14]. Here, the Getty Vocabularies Project is a classic case, as it contains reference vocabularies that describe the work of artists and architects, such as the Art and Architecture Thesaurus (AAT).

The two publication models identified by Pohorec, Zorman, and Kokol include both inline and parallel Web publications (2013). Another popular approach of making massive datasets available as Linked Data is by publishing them in parallel, as machine-processable collections of RDF data. However, the findings of this study include an observation that RDF material published using this method is not often indexed by standard search engines and has limited exposure [15]. Instead of just utilizing an established standard like Microdata or Resource Description Framework (RDF), authors can use an extension like the Microdata or RDF markup to add semantic markup to HTML text and create structured data. As a result, the Linked Data concepts are syntax agnostic, making the technology compatible with many markup languages and serialization methods.

Linked Data is defined as a method of publishing and interconnecting structured information while also encouraging the growth and contribution of other participants in the Linked Data ecosystem [16], The machines can “understand” the meaning of the relationships by making inferences about the content and then interlinking that data to new information automatically, discovering connections that were previously hidden [17, 18].

3. Methodology

Development of the EGMRB Thesaurus

EGMRB Thesaurus development follows the procedure in Fig. 1, based on standard for building multilingual thesauri [19-21].

In the beginning, we are just getting started when we are synthesizing several language resources provided by our partnering academic institutes and various KOS from the domains of humanities and social sciences. Mapping or alignment can be performed with regard to the words and ideas included in the following canonical sources: these sources serve not only as reference material for terms and concepts in the EGMRB Thesaurus but also as subjects of that mapping or alignment. A comprehensive mapping of the ideas included in LOD to reference and top-level concept schemes, such as the AAT and the BBT, should be achievable. In our process, we include interactions with subject matter experts and information professionals. It will be essential to provide the local topic heading lists, thesauri and validate third-party subject heading lists. It will also help test EGMRB Thesaurus macro and microstructure.

This thesaurus contains approximately 4,609 ethnic terms organized hierarchically, although other essential ideas are also included. A collaborative working team method was employed throughout the development of the EGMRB Thesaurus. Different duties were divided up amongst editors and knowledge engineers, who worked together. As a result, editors have knowledge of many ethnic groups' subjects, and they are responsible for planning, design, implementation, and maintenance of the thesaurus.

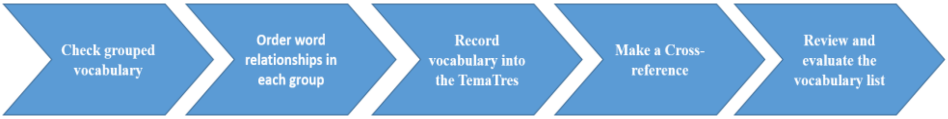
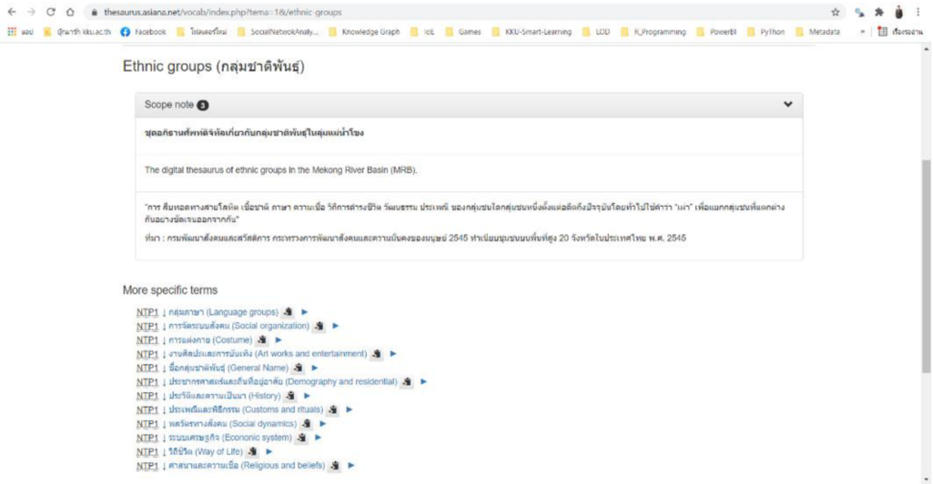

Check grouped vocabulary
Order word relationships in each group
Record vocabulary into the TemaTres
Make a Cross-reference
Review and evaluate the vocabulary list

Figure 1. The EGMRB digital thesaurus development



Ethnic groups (กลุ่มชาติพันธุ์)

Scope note 3

ชุดอภิธานศัพท์ดิจิทัลเกี่ยวกับกลุ่มชาติพันธุ์ในลุ่มแม่น้ำโขง

The digital thesaurus of ethnic groups in the Mekong River Basin (MRB).

"การ สืบทอดทางสายโลหิต เชื้อชาติ ภาษา ความเชื่อ วิถีการดำรงชีวิต วัฒนธรรม ประเพณี ของกลุ่มชนใดกลุ่มชนหนึ่งตั้งแต่อดีตถึงปัจจุบันโดยทั่วไปใช้คำว่า "เผ่า" เพื่อแยกกลุ่มชนที่แตกต่างกันอย่างชัดเจนออกจากกัน"

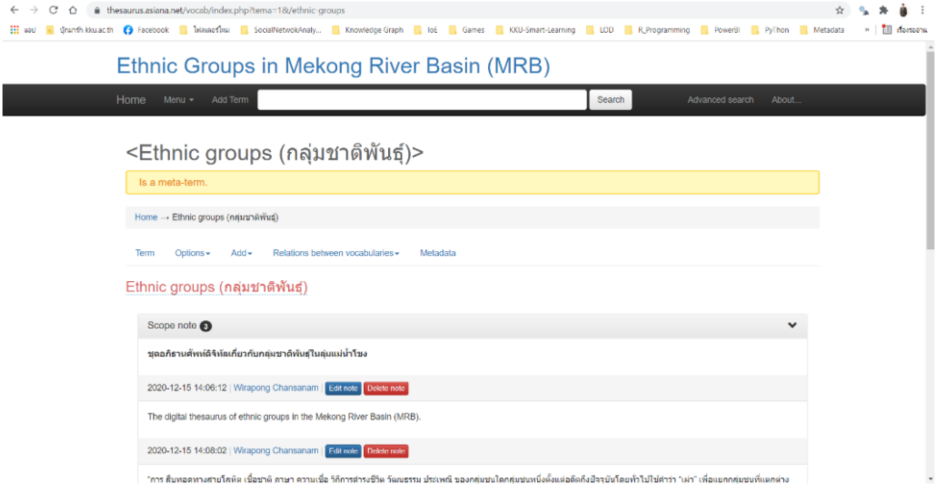
ที่มา : กรมพัฒนาสังคมและสวัสดิการ กระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ 2545 ทำเนียบชุมชนบนพื้นที่สูง 20 จังหวัดในประเทศไทย พ.ศ. 2545

More specific terms

NTP1 ↓ กลุ่มภาษา (Language groups) ▶
NTP1 ↓ การจัดระบบสังคม (Social organization) ▶
NTP1 ↓ การแต่งกาย (Costume) ▶
NTP1 ↓ งานศิลปะและการบันเทิง (Art works and entertainment) ▶
NTP1 ↓ ชื่อกลุ่มชาติพันธุ์ (General Name) ▶
NTP1 ↓ ประชากรศาสตร์และถิ่นที่อยู่อาศัย (Demography and residential) ▶
NTP1 ↓ ประวัติและความเป็นมา (History) ▶
NTP1 ↓ ประเพณีและพิธีกรรม (Customs and rituals) ▶
NTP1 ↓ พลวัตรทางสังคม (Social dynamics) ▶
NTP1 ↓ ระบบเศรษฐกิจ (Econonic system) ▶
NTP1 ↓ วิถีชีวิต (Way of Life) ▶
NTP1 ↓ ศาสนาและความเชื่อ (Religious and beliefs) ▶

Figure 2. The screen shot of hierarchical structure with narrower terms and meta terms for "Ethnic groups" respectively.

TemaTres (http://www.vocabularyserver.com/) was chosen to make SKOS thesaurus' development less difficult [22]. TemaTres' relative simplicity, and the fact that it was possible to interact with it through the web, made it a good option for developing EGMRB Thesaurus. Fig. 3 presents the TemaTres user interface, which contains a basic but helpful feature for modifying ideas. As shown, it has advanced search capabilities and the possibility to link to external concepts (e.g., SKOS-Core, JSON, JSON-LD). In addition, the TemaTres allows a little more basic taxonomy input for tab-indented files or RDF-XML encoded SKOS documents [23].


thesaurus.asiana.net/vocab/index.php?tema=1&/ethnic-groups
Home
Menu
Add Term
Search
Advanced search
About...
Term
Options
Add
Relations between vocabularies
Metadata
Edit note
Delete note

Ethnic Groups in Mekong River Basin (MRB)

<Ethnic groups (กลุ่มชาติพันธุ์)>

Is a meta-term.

Home → Ethnic groups (กลุ่มชาติพันธุ์)

Ethnic groups (กลุ่มชาติพันธุ์)

Scope note 3

ชุดอภิธานศัพท์ดิจิทัลเกี่ยวกับกลุ่มชาติพันธุ์ในลุ่มแม่น้ำโขง

2020-12-15 14:06:12 | Wirapong Chansanam

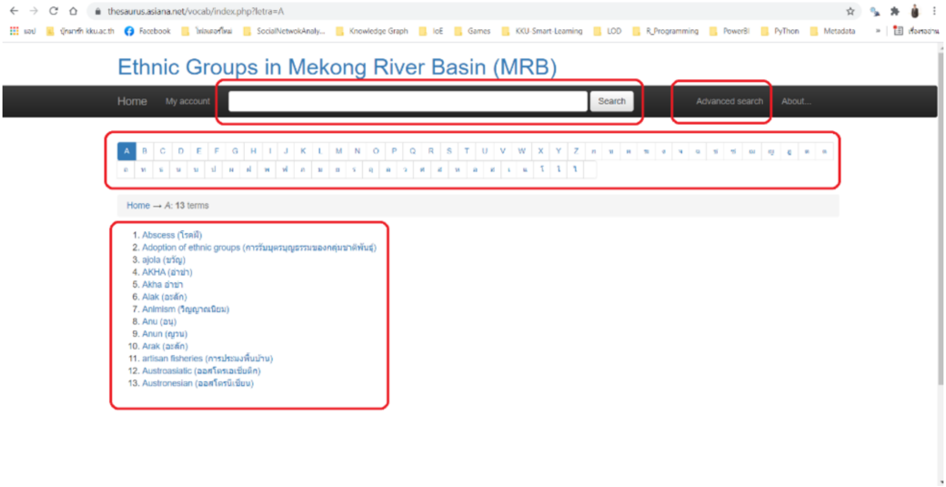
The digital thesaurus of ethnic groups in the Mekong River Basin (MRB).

2020-12-15 14:08:02 | Wirapong Chansanam

"การ สืบทอดทางสายโลหิต เชื้อชาติ ภาษา ความเชื่อ วิถีการดำรงชีวิต วัฒนธรรม ประเพณี ของกลุ่มชนใดกลุ่มชนหนึ่งตั้งแต่อดีตถึงปัจจุบันโดยทั่วไปใช้คำว่า "เผ่า" เพื่อแยกกลุ่มชนที่แตกต่าง

Figure 3. In the admin interface of the EGMRB thesaurus main page, there are several tabs on the screen.

SPARQL endpoints are also accessible, enabling anybody (human or otherwise) to utilize the SPARQL language to search the thesaurus [24]. Results are often delivered in machine-processable forms. In addition to the SPARQL interface, TemaTres also provides an online form that knowledge engineers may use to query the thesaurus (Fig. 5). While inexperienced users will generally handle both the formulation of questions and the human-readable display of results, developing ad-hoc user interfaces and software are required for more proficient users.

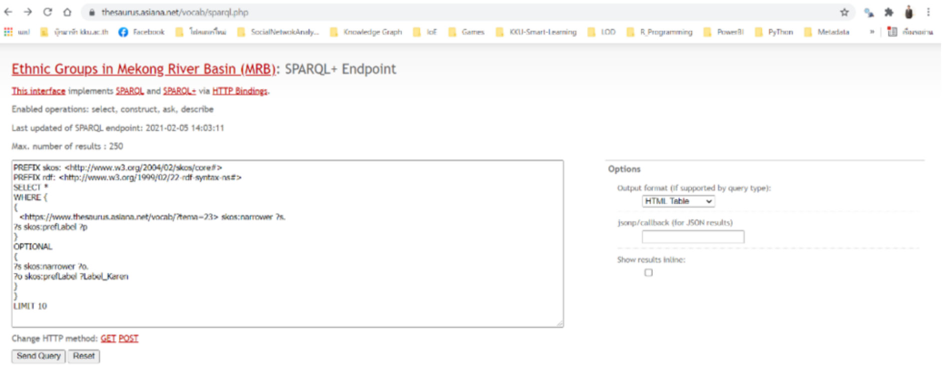


Ethnic Groups in Mekong River Basin (MRB)

Home → A: 13 terms

1. Abscess (โรคฝี)
2. Adoption of ethnic groups (การรับบุตรบุญธรรมของกลุ่มชาติพันธุ์)
3. ajola (ขวัญ)
4. AKHA (อ่าข่า)
5. Akha อ่าข่า
6. Alak (อะลัก)
7. Animism (วิญญาณนิยม)
8. Anu (อนุ)
9. Anun (ญวน)
10. Arak (อะลัก)
11. artisan fisheries (การประมงพื้นบ้าน)
12. Austroasiatic (ออสโตรเอเชียติก)
13. Austronesian (ออสโตรนีเซียน)

Figure 4. EGMRB Thesaurus's variety of user-friendly search engines included: simple search engine; auto-completed text box search, alphabetic index, and advanced search.



Ethnic Groups in Mekong River Basin (MRB): SPARQL+ Endpoint

This interface implements SPARQL and SPARQL+ via HTTP Bindings.

Enabled operations: select, construct, ask, describe

Last updated of SPARQL endpoint: 2021-02-05 14:03:11

Max. number of results : 250

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT *
WHERE {
{
  <https://www.thesaurus.asiana.net/vocab/?tema=23> skos:narrower ?s.
?s skos:prefLabel ?p
}
OPTIONAL
{
?s skos:narrower ?o.
?o skos:prefLabel ?Label_Karen
}
}
LIMIT 10
```

Change HTTP method: GET POST

Options

Output format (if supported by query type): HTML Table

jsonp/callback (for JSON results)

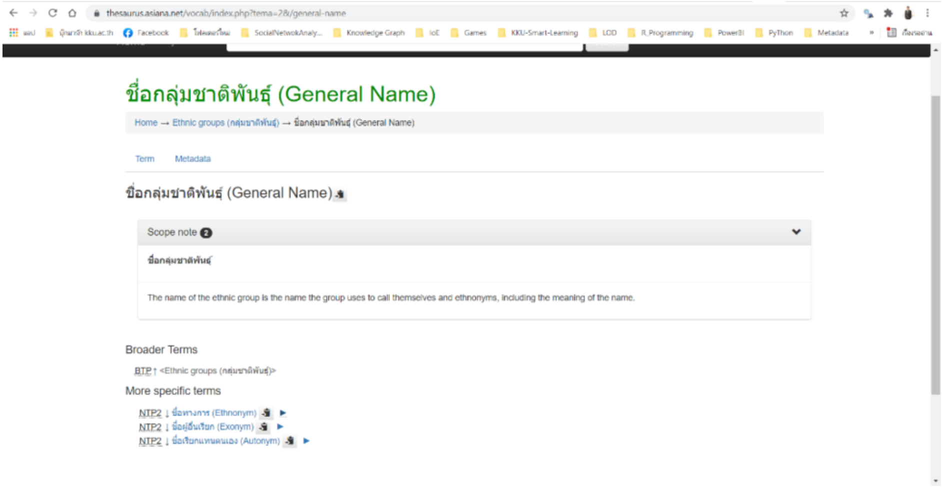
Show results inline: □

Figure 5. SPARQL endpoint interface for the concept "Karen" which offers an example of a query in the form of a query string, with a response in HTML table format.

We assigned attributes and relations to the terms, then changed them based on how they were relevant to the webpage's structure. Editorial touch-ups: Users could modify the structural elements of each term (preferred term, non-preferred term, definition, associated bibliographic note, etc.) by using the TemaTres user interface, while simultaneously restructuring the hierarchical structure if applicable. Each of the chosen terms was accompanied by a definition that frequently linked to a bibliographic reference and was presented within a hierarchical relationship. In addition, preferred terminology such as related terms and a synonym were given where it is usual to use those phrases.

When the thesaurus was built, editors sent it to the knowledge engineer, who went through and assessed various terms and their structural elements, then confirmed or excluded them and submitted revisions to the editors in the form of private notes from TemaTres. The knowledge engineers submitted their comments, and the editors evaluated the information, after which the stable version of the thesaurus was issued. Codes and Uniform Resource Identifiers (URIs) are published, and each code and URI changes are monitored.

EGMRB Thesaurus is accessible over the internet at the URL https://thesaurus.asiana.net/vocab/ This feature is flexible and might be applied to many uses. Additionally, we utilize the EGMRB Thesaurus as well as other existing controlled vocabularies to help manage data sources, including databases, research documents, and to support semantic interoperability at the Princess Maha Chakri Sirindhorn Anthropology Centre (Public Organization) (https://www.sac.or.th/databases/ethnic-groups/ethnicGroups)



ชื่อกลุ่มชาติพันธุ์ (General Name)

Home → Ethnic groups (กลุ่มชาติพันธุ์) → ชื่อกลุ่มชาติพันธุ์ (General Name)

ชื่อกลุ่มชาติพันธุ์ (General Name)

Scope note 2

ชื่อกลุ่มชาติพันธุ์

The name of the ethnic group is the name the group uses to call themselves and ethnonyms, including the meaning of the name.

Broader Terms

BTP ↑ <Ethnic groups (กลุ่มชาติพันธุ์)>

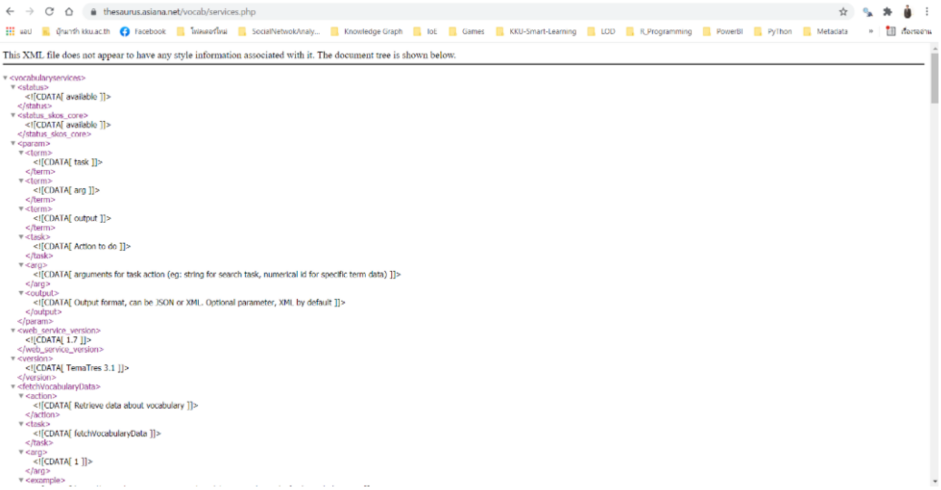
More specific terms

NTP2 ↓ ชื่อทางการ (Ethnonym) ▶
NTP2 ↓ ชื่อผู้อื่นเรียก (Exonym) ▶
NTP2 ↓ ชื่อเรียกแทนตนเอง (Autonym) ▶

Figure 6. A screenshots of vocabulary provided for the concept "General Name".

INSPIRE, the national standard of metadata, includes guidelines for SAC metadata which meet these guidelines (Drafting Team Metadata and European Commission Joint Research Centre, 2013). The "Dataset Metadata" user interface incorporates an approved user input form, as well as a synonym entry form from the EGMRB Thesaurus, within its "Dataset Metadata" section to allow authorised users to describe datasets in line with SAC Metadata Schema. While the metadata standard does not specify where to put metadata terms like "keyword" for the explanation of datasets on ethnic groups variety or correlated domains, EGMRB Thesaurus can be used for metadata annotations, and the process of identifying terms to use for metadata annotations begins by looking for relevant metadata terms in the data. If the user uses clear terms from shared thesauri for their metadata schema's population, searchers' efforts will be helped tremendously in data discovery and retrieval.

Additional anthropological thesauri and controlled vocabularies are necessary to the academic community. Furthermore, the EGMRB Thesaurus is utilized by the Linked Data system for data annotation [25]. When contributors uploading data resources through the portal utilize the data ingestion web interface, they must name their dataset's fields after the taxonomies, which are located inside the thesauri. These ideas, when put together, form the "EGMRB Data Schema". The examples shown in figure 7 illustrate the annotation of data accessible via the data portal by leveraging APIs (JSON format).



This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼<vocabularyservices>
  ▼<status>
     <![CDATA[ available ]]>
   </status>
  ▼<status_skos_core>
     <![CDATA[ available ]]>
   </status_skos_core>
  ▼<param>
    ▼<term>
       <![CDATA[ task ]]>
     </term>
    ▼<term>
       <![CDATA[ arg ]]>
     </term>
    ▼<term>
       <![CDATA[ output ]]>
     </term>
    ▼<task>
       <![CDATA[ Action to do ]]>
     </task>
    ▼<arg>
       <![CDATA[ arguments for task action (eg: string for search task, numerical id for specific term data) ]]>
     </arg>
    ▼<output>
       <![CDATA[ Output format, can be JSON or XML. Optional parameter, XML by default ]]>
     </output>
   </param>
  ▼<web_service_version>
     <![CDATA[ 1.7 ]]>
   </web_service_version>
  ▼<version>
     <![CDATA[ TemaTres 3.1 ]]>
   </version>
  ▼<fetchVocabularyData>
    ▼<action>
       <![CDATA[ Retrieve data about vocabulary ]]>
     </action>
    ▼<task>
       <![CDATA[ fetchVocabularyData ]]>
     </task>
    ▼<arg>
       <![CDATA[ 1 ]]>
     </arg>
    ▼<example>
```

Figure 7. A portion of the database's schema description including information on the semantic annotations of fields with thesauri concepts.

In addition, while searching for the ethnic group concept, all these terms will be returned. The terminology and URIs in a standardized data markup are connected explicitly with one another, allowing records to be understood by the machine. The potential for computer-aided change, distribution, and long-term reuse opens up with the opening of datasets to a larger universe of possibilities.

The EGMRB Thesaurus has a stated objective of developing a web portal that will increase the amount of ethnic historical MRB resources in the public domain and help in the advanced search and display of information linked to this inheritance. To enhance the quality of the study, it seeks to utilize as many scientific references as feasible. This thesaurus is specifically designed for academics, students, and anthropologists. For example, the researcher may connect words they may have found in manuscript documentation with thesaurus entries, which are more modern and conventional in their nomenclature. Standardizing vocabulary in a monument requires the use of a thesaurus. Students in the humanities and social sciences will learn about many ethnic groups. In addition, this vocabulary's ultimate goal is to promote the knowledge, conservation, and distribution of ethnic groups information.

The EGMRB Thesaurus has an increasingly international terminology, implying that all terms have to be translated. Initially, simple words, synonyms, and related ideas were translated directly. In other situations, specialized resources provide translations in between Thai and English. Because terms only existed in the original ethnic language or needed more than one translation, difficulties occurred while translating them.

4. Discussion

This project aims to ensure its long-term sustainability. The project's initial goal was to create the EGMRB Thesaurus to address ambiguity concerns with natural language by standardizing how words, meanings, and relationships are constructed.

As a result, building upon the various semantic resources, such as the EGMRB Thesaurus, is essential for facilitating the discovery and interchange of existing and new knowledge and accomplishing a more thorough considerate within the particular domain as well as a greater understanding of the broader context. The first step towards semantic interoperability is dataset annotation using taxonomies [17].

This perspective argues that the EGMRB Thesaurus is created according to LOD principles to ensure that computers may run in a way that assists data users in finding and processing relevant material [26]. In addition, EGMRB Thesaurus, maintaining standards about Web development, boosts interoperation and use of automatic data interchange across various sources. EGMRB Thesaurus ideas for the semantic annotation of their data enable simultaneous searches on various data.

The thematic thesaurus is expected to be an essential initial step in creating an ethnic group knowledge base for the Mekong River Basin. While [27] talk about creating ontologies, they also suggest using thesauri as a starting point. In this view, EGMRB Thesaurus employed knowledge structure [28] based on ethnicity in six countries, moreover utilized semantic data categorization using a LOD framework [25], expanding researchers' boundaries.

5. Conclusion and future work

The purpose of this article is to construct the EGMRB Thesaurus. Instead, they aim to enable the discovery and administration of different content types in the EGMRB Thesaurus database, providing resources for topics such as politics, liberal arts, linguistics, and culture. We hope that our resource will become a valuable part of the Linked Open Data cloud and function as a reference tool in subjects of anthropology domain. While important work remains to be done in creating and publishing the EGMRB Thesaurus, significant progress has been made. The first complete edition of the thesaurus is expected to be released in October 2021. The EGMRB Thesaurus digital platform will have another assignment: integration of the thesaurus into the platform. This means that academics will be able to comment on virtual exhibits, virtual collections, and other content developed on the platform. In order to increase the reach of our platform, we will employ the thesaurus to expand search throughout our existing collection and the semantic domain by accessing a smaller dataset given by a researcher of the Digital Humanities Research Group (DHRG) who specializes in folktales.

References

[1] R. Albertoni et al., “Data Catalog Vocabulary (DCAT)-Version 2,” W3C Candidate Recommendation, vol. 3, 2019.

[2] M. B. Jones, M. P. Schildhauer, O. Reichman, and S. Bowers, “The new bioinformatics: integrating ecological data from the gene to the biosphere,” Annu. Rev. Ecol. Evol. Syst., vol. 37, pp. 519–544, 2006.

[3] J. Madin, S. Bowers, M. Schildhauer, S. Krivov, D. Pennington, and F. Villa, “An ontology for describing and synthesizing ecological observation data,” Ecological informatics, vol. 2, no. 3, pp. 279–296, 2007.

[4] B. Haslhofer and W. Klas, “A survey of techniques for achieving metadata interoperability,” ACM Computing Surveys (CSUR), vol. 42, no. 2, pp. 1–37, 2010.

[5] ISO 25964-1: Information and documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval. ISO, Geneva (2011)

[6] BS 8723-1, 2005. Structured vocabularies for information retrieval - Guide - Part 1: Definitions, symbols and abbreviations. British Standards Institution, London.

[7] E. G. Fayen, “Guidelines for the construction, format, and management of monolingual controlled vocabularies: A revision of ANSI/NISO Z39. 19 for the 21st century,” Information Wissenschaft und Praxis, vol. 58, no. 8, p. 445, 2007.

[8] I. O. for Standardization, ISO 25964-2:2013: Information and Documentation : Thesauri and Interoperability with Other Vocabularies. Interoperability with other vocabularies, no. pt. 2. ISO, 2013. [Online]. Available: https://books.google.co.th/books?id=703ooAEACAAJ

[9] A. Miles and S. Bechhofer, “SKOS simple knowledge organization system reference,” W3C recommendation, 2009.

[10] Coudyzer, E.: First release GLAM sector reference terminologies (Sep 2013), https://www.athenaplus.eu/getFile.php?id=187

[11] P. Harpring, Introduction to controlled vocabularies: terminology for art, architecture, and other cultural works. Getty Publications, 2010.

[12] D. Koutsomitropoulos, G. Solomou, and K. Kalou, “Federated semantic search using terminological thesauri for learning object discovery,” Journal of Enterprise Information Management, 2017.

[13] H. Nagy, T. Pellegrini, and C. Mader, “Exploring structural differences in thesauri for SKOS-based applications,” in Proceedings of the 7th International Conference on Semantic Systems, 2011, pp. 187–190.

[14] Y. Yang, J. Xiong, and S. Wang, “A semantic search engine based on SKOS model ontology in agriculture,” in International Conference on Computer and Computing Technologies in Agriculture, 2010, pp. 110–118.

[15] S. Pohorec, M. Zorman, and P. Kokol, “Analysis of approaches to structured data on the web,” Computer Standards & Interfaces, vol. 36, no. 1, pp. 256–262, 2013.

[16] C. Bizer, T. Heath, and T. Berners-Lee, “Linked data: The story so far,” in Semantic services, interoperability and web applications: emerging concepts, IGI global, 2011, pp. 205–227.

[17] T. Berners-Lee, J. Hendler, and O. Lassila, “The semantic web,” Scientific american, vol. 284, no. 5, pp. 34–43, 2001.

[18] W3.org. 2021. Ontologies - W3C. [online] Available at: <http://www.w3.org/standards/semanticweb/ontology> [Accessed 19 June 2021].

[19] E. G. Fayen, “Guidelines for the construction, format, and management of monolingual controlled vocabularies: A revision of ANSI/NISO Z39. 19 for the 21st century,” Information Wissenschaft und Praxis, vol. 58, no. 8, p. 445, 2007.

[20] J. Aitchison, D. Bawden, and A. Gilchrist, Thesaurus construction and use: a practical manual. Routledge, 2003.

[21] V. Broughton, Essential thesaurus construction. Facet Publishing, 2006.

[22] A. Gonzales-Aguilar, M. Ramírez-Posada, and D. Ferreyra, “TemaTres: software para gestionar tesauros,” Profesional de la Información, vol. 21, no. 3, pp. 319–325, 2012.

[23] D. Beckett and B. McBride, “RDF/XML syntax specification (revised),” W3C recommendation, vol. 10, no. 2.3, 2004.

[24] E. Prud’hommeaux, “SPARQL query language for RDF, W3C recommendation,” http://www. w3. org/TR/rdf-sparql-query/, 2008.

[25] W. Chansanam, K. Tuamsuk, and J. Chaikhambung, “Linked Open Data Framework for Ethnic Groups in Thailand Learning,” International Journal of Emerging Technologies in Learning (iJET), vol. 15, no. 10, pp. 140–156, 2020.

[26] E. Garnier et al., “Towards a thesaurus of plant characteristics: an ecological contribution,” Journal of Ecology, vol. 105, no. 2, pp. 298–309, 2017.

[27] W. Chansanam, K. Tuamsuk, K. Kwiecien, T. Ruangrajitpakorn, and T. Supnithi, “Development of the Belief Culture Ontology and Its Application: Case Study of the GreaterMekong Subregion,” in Joint International Semantic Technology Conference, 2014, pp. 297–310.

[28] J. Chaikhambung and K. Tuamsuk, “Knowledge classification on ethnic groups in Thailand,” Cataloging & Classification Quarterly, vol. 55, no. 2, pp. 89–104, 2017.