

# Developing a Linked Open Data Platform for Folktales in the Greater Mekong Subregion

Treepidok Ngootip <sup>1</sup>, Paiboon Manorom <sup>1</sup>, Wirapong Chansanam <sup>1\*</sup>, Marut Buranarach <sup>2</sup>

<sup>1</sup> Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Khon Kaen 40002, Thailand.

<sup>2</sup> Thailand National Electronics and Computer Technology Center, Pathum Thani 12120, Thailand.

## Abstract

This research paper presents the development of a linked open data (LOD) platform that aims to organize and facilitate access to valuable knowledge about folktales and ethnic groups in the Greater Mekong Subregion countries. The study's methodology involved the creation of a linked open data platform, structuring folktales' knowledge, and evaluating its performance through expert assessment. The LOD platform was constructed through Google OpenRefine to establish connections with external data sources, and the RDF files (N-Triples) were deployed on Fuseki Server (Apache Jena) to serve as the SPARQL endpoint for querying the linked open data. The Pubby web app was chosen for further development to provide a user-friendly interface, which customized with the Bootstrap framework, featuring an intuitive homepage and a search box function for simplified data retrieval. For the expert evaluation, the study confirmed that the platform performs a high suitability in terms of congruence, reliability, integrity, understandability, collaboration, accessibility, and connectedness. The developed LOD platform exhibits significant potential for expanding its application to various content domains, offering a valuable resource for accessing and exploring the rich cultural heritage of folktales in the Greater Mekong Subregion countries.

## Keywords:

Linked Open Data Platform;  
Folktales;  
Greater Mekong Subregion;  
Knowledge Organization.

## Article History:

<b>Received:</b>	11	August	2023
<b>Revised:</b>	08	October	2023
<b>Accepted:</b>	03	November	2023
<b>Published:</b>	01	December	2023

## 1- Introduction

The Linked Open Data (LOD) platforms have emerged to facilitate the discovery, integration, and analysis of an open data set, which provide a variety of features, such as SPARQL endpoints, RDF storage, and graphical user interfaces (GUIs). To support LOD performance, the Pubby Web was adopted in this study as an open-source LOD platform that extends RDF libraries to support SPARQL endpoints. The Pubby web app can also be rendered as a simple HTML website, making it accessible to a wider audience [1]. The SparqPlug platform is a service that enables the extraction of LOD from web-based HTML documents without RDF data. The SparqPlug platform extracts data from the HTML DOM and converts it to RDF data, which can then be queried using SPARQL [2]. The Talis platform is a software-as-a-service (SaaS) LOD platform that provides RDF storage and access to SPARQL endpoints. The Talis platform also provides a REST API, which allows users to interact with the platform programmatically [3].

Replicating the previous studies, the LOD was employed in a number of study areas. For example, Reda and Carbonaro [4] developed a web platform for collecting and publishing Internet of Things (IoT) health and fitness datasets with Linked Data principles. By leveraging the IFO ontology and Semantic Web technologies, the platform enables the sharing of semantically meaningful and easily discoverable data, overcoming the challenges of manually integrating heterogeneous IoT datasets, which demonstrate that Semantic Web technologies offer a comprehensive solution to

\* **CONTACT:** [wirach@kku.ac.th](mailto:wirach@kku.ac.th)

**DOI:** <http://dx.doi.org/10.28991/ESJ-2023-07-06-06>

© 2023 by the authors. Licensee ESJ, Italy. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<https://creativecommons.org/licenses/by/4.0/>).

address data silos in the IoT landscape, facilitating valuable insights for better healthcare. Additionally, Tilahun et al. [5] explore the potential of Linked Open Data (LOD) technologies for representing, visualizing, and retrieving health information. Using RDF and Silk—a LOD-based system developed and evaluated by the WHO health observatory dataset, these systems allowed flexible queries and visualizations, receiving positive usability feedback, with some users finding SPARQL queries challenging. Ali et al. [6] also present the Linked Open COVID-19 Data System, utilizing LOD principles to structure and transform COVID-19 data into a semantic format to evaluate Semantic Web tools' efficacy in building LOD-based COVID-19 information systems, with promising results for providing valuable insights and information to stakeholders. Similar to D'Agostino et al. [7], they discuss the intersection of open data and public health, emphasizing the need for open access to public health data during disease outbreaks, which highlights the lack of articulated policy guidelines for data collection and management in many countries, particularly in Latin American countries.

The study aimed to develop a regional framework for open data in public health, adaptable to each country through appropriate national or subnational policies, as well as Dhayne et al. [8] examines the use of Linked Open Data (LOD) and Semantic Web technologies to address the challenge of data discovery, linking, and integration in big data variety to evaluate the efficacy of two frameworks, SILK and LIMES, for interlinking large-scale biomedical datasets, providing insights for future research and enhancement in this domain. The evaluation of SILK and LIMES indicated that the two frameworks were potential options for interlinking large-scale biomedical datasets within Linked Open Data (LOD) frameworks, comparing the two tools at various levels, including general features, comparison measures, resulting files, performance, and link effectiveness. The study contributes valuable insight view for choosing the most suitable tool in a biomedical context and offers opportunities for future research and enhancements in this domain.

Furthermore, Boilson et al. [9] launched a MIDAS project to develop a big data platform for integrating health and social care data to offer privacy-preserving analytics, forecasting tools, and visualizations. In this study, a realist evaluation framework, such as case studies and stakeholder feedback, was employed to ensure user requirements were aligned with the IT system development streamline. To enhance reliability in the study, longitudinal interviews and the Q-methodology were used to measure stakeholder attitudes. In addition, Wu et al. [10] propose a Linked Data-based knowledge navigation system (LDKNS) to enhance teaching effectiveness in art design education. This study employed a linked data visualization technology and learning resource ontology with undergraduate students' experiments, which illustrates increased motivation, improved achievement, and reduced cognitive load when using the LDKNS, indicating its potential for significant learning advantages. Besides, Monika Rani et al. [11] discuss Linked Open Data (LOD) practices and the importance of data quality in interlinking data on the Semantic Web to analyze the quality of linked data, existing frameworks for assessment, and the challenges in achieving high-quality LOD. On the one hand, Peña et al. [12] explore Information Visualization adaptability to the Semantic Web community, particularly Linked Open Data (LOD) principles, which emphasize the importance of creating smart visualizations with semantic descriptions to enable machines to understand and connect to external data sources. This study would advocate for the widespread adoption of semantic technologies, encouraging LOD-driven markets beyond the Semantic Web community. Not only for semantic technology, but the study also highlights the need for collaboration and process descriptions in LOD visual exploration tools to enhance data discoverability and reusability. This technology tends to improve LOD visualization, which is a greater implication of the Semantic Web on the Internet. Likewise, Dimitrakis et al. [13] survey recent work on stateless Question Answering (QA) systems, focusing on methods applied in RDF and Linked Data, documents, and their combinations, which identify challenges, categorize approaches, and review 21 systems and 23 evaluation datasets based on domain, knowledge source, tasks, and evaluation metrics.

In the study, Dimitrakis et al. [14] introduced LODQA, an approach for open domain Question Answering over Linked Open Data to cover factoid, confirmation, and definition questions, exploiting 400 Linked datasets and LODsyndesis services without using training data. The process involves question cleaning, entity recognition, and disambiguation using Linked Data and NLP methods, WordNet-based question expansion, and triple scoring for final answers. The results revealed that the effectiveness and efficiency of the approach played a significant role in the linked open data, which is consistent with the study of Cuijuan et al. [15] presents the implementation of a Genealogy Knowledge Service Platform using linked data technology, enabling knowledge reorganization and discovery. It describes the platform's design and development and showcases four examples of how it assists users in research, demonstrating Linked Data application in Digital Humanities.

Turning to the implications studies, Hawkins [16] discusses the potential of Linked Data to provide structured, integrated, and interoperable archival data for Digital Humanities research, which highlights the under-utilization of digital archives and proposes increased collaboration between archival and digital humanities scholars, with an emphasis on using AI and low-barrier tools to scale up the production of Archival Linked Data and enhance access to digitized and born-digital archives. Moreover, Zeng [17] explores the use of 'Semantic enrichment' to enhance the quality and discoverability of historical and cultural heritage data provided by libraries, archives, and museums (LAMs), which introduces various semantic enrichment methods and efforts that support deeper exploration and use of LAM data in digital humanities research. Real cases, research projects, experiments, and pilot studies demonstrate the potential for

LAM data regardless of their structure or types of artifacts, encouraging more effective initiatives to maximize their value in the Digital Humanities and Semantic Web. Meanwhile, Hyvönen et al. [18] proposed WarSampo, a system for publishing heterogeneous, distributed data about World War II on the Semantic Web, which comprises a Linked Open Data service for Digital Humanities research and a semantic portal for studying war history and family destinies. The portal had significant public interest upon its release in 2015, indicating a strong demand for such historical applications.

Nevertheless, despite the various research projects that have been focused on different domains and data sources, there seems to be a commonality in their approach—the utilization of Linked Open Data (LOD) for data application, which tends to be worth noting. On the contrary, these projects aim to develop and implement LOD platforms since a deficiency seems to be occurring between these platforms. This integration deficiency would pay a significant challenge as it hinders the seamless exchange and utilization of data across different LOD platforms. Without a cohesive framework for integrating information, the potential insights and benefits that could be derived from cross-referencing data from diverse domains remain untapped. To gain an extensive explanation, the study of Jazz Cats [19] would be a great example of the accessibility and usability issues of LOD platforms. His study explores a project that adheres to a 5-Star standard of Linked Open Data (LOD) but presents barriers for human users. Currently, the project's RDF data is stored in an instance of the Virtuoso triplestore, which can only be queried through the default SPARQL endpoint on the Jazz Cats website. While example queries demonstrate simultaneous results from multiple datasets, engaging with the data beyond these examples requires users to have the skills to construct new SPARQL queries. The authors acknowledge that the current site UI presents notable barriers to access, considering the dichotomy of skills between music scholars and query formulation. Similarly, Wathanti et al. [20] present a methodology aimed at creating and publishing Linked Data regarding Thai tradition through ontology construction. The approach incorporates a pilot model example and offers comprehensive explanations of each task, catering to the needs of private and public agencies handling cultural and lifestyle data in Thailand. The methodology discusses various tools for data linking, including OpenRefine, Jena, Fuseki, Jetty, Pubby, and the RDF extension. However, their pilot project lacks a search function on the home page, and the user interface is not user-friendly. The LOD platforms mentioned above have both advantages and limitations. Advantages: a) The platforms allow LOD to be searched from multiple sources simultaneously; b) Some platforms can also display LOD instantly. Limitations: a) Some platforms are packaged software, so customization is complicated; b) Some platforms have an outdated GUI, so users may need help finding information on the platforms. Table 1 summarizes the advantages and limitations of the LOD platforms mentioned above.

**Table 1. The comparing of the advantages and limitations of linked open data platforms**

Platforms	Advantages	Limitations	Data Type	LOD Star Rating
The Pubby web app	Open-source, can be rendered as a simple HTML website	Accessing and browsing the platform is challenging due to the absence of a dedicated home page and the inability of users to search for data effectively. Moreover, the process of customization is overly complex and lacks user-friendliness.	RDF	★★★★★
SparqPlug platform	Extracts LOD from HTML documents without RDF data	The data received through the SparqPlug platform may be incomplete as it relies on extracting information from the HTML Document Object Model (DOM) and subsequently converting it into RDF data.	RDF	★★★★★
Tails platform	Provides RDF storage and access to SPARQL endpoints	Platform customization is challenging due to its ready-made software nature, resulting in a complex and difficult process.	RDF	★★★★★

Addressing the limitation, this study emphasizes the development of a unified and interconnected LOD infrastructure, such as an integrated platform, which would enable researchers and practitioners to access and utilize data from multiple sources within a coherent ecosystem. The LOD platform might facilitate efficient data sharing and interlinking of information by establishing common data standards, ontologies, and linking mechanisms. In this study, integrating information across various LOD platforms would unlock new opportunities for interdisciplinary research, enriching the scope of data analysis and exploration. In terms of collaboration, the collaborative efforts among researchers, institutions, and data providers are essential to establish and maintain this integrated LOD infrastructure, fostering a harmonized ecosystem of linked data resources. Integrating information into LOD platforms would ultimately drive innovation and discovery across diverse fields, creating a powerful and dynamic knowledge network that transcends individual research projects to embrace the principles of open data and interoperability. Hence, the research community would unlock the full potential of LOD and advance the collective understanding of complex global challenges.

Since the existing literature on Linked Open Data (LOD) platforms reveals several gaps that require attention and further research to enhance the integration and accessibility of data, future studies would be concerned as follow. Firstly, there is a deficiency of integration between various LOD platforms developed for domains and data sources because the individual research projects focus on specific areas; a unified framework for linking and sharing data across these platforms limits the potential for cross-domain insights and knowledge exchange. The research should aim to develop interoperable LOD platforms that enable seamless data integration and interlinking between diverse domains. Secondly,

usability and accessibility issues are prevalent on some LOD platforms. For instance, the Jazz Cats project adheres to a high standard of LOD but presents challenges for human users due to complex query formulation requirements. In this regard, the researchers should explore user-centric design approaches to create intuitive interfaces that simplify data retrieval and exploration for non-experts, thereby widening the platform's user base and enhancing its overall accessibility.

Furthermore, there is a need to incorporate search functionalities on the home pages of LOD platforms. Wathanti et al. [20] pilot project for Thai tradition lacks a search function, which could hinder users' ability to find specific information efficiently, as well as a user-friendly search feature on the homepage that can significantly enhance data discoverability and user experience. Another gap in the literature is the exploration of semantic enrichment methods to enhance the quality and discoverability of historical and cultural heritage data provided by libraries, archives, and museums (LAMs). By leveraging semantic technologies and enriching LAM, the data would unlock valuable insights for digital humanities research, but further investigation into effective semantic enrichment approaches is still required. Lastly, while several research projects highlight the potential of LOD in specific domains, there is a need for more comprehensive efforts toward creating a linked data infrastructure encompassing diverse areas. Therefore, the collaborative initiatives, which promote data sharing, common standards, and ontology, might lead to the development of an integrated LOD ecosystem to foster interdisciplinary research and knowledge discovery. In this study, the gaps and contributions will be addressed to advance LOD platforms, enabling researchers, practitioners, and the public to harness the full potential of linked data for valuable insights and discoveries across various domains, respectively.

The limitations of LOD platforms can be addressed by developing more user-friendly platforms that provide better customization capabilities. In this case, the development of standards for LOD platforms would help to ensure that the platforms are interoperable and LOD can be easily shared and reused [21]. Then, the selection of the Pubby web application from Table 1 was generated for development purposes. However, it is important to note that despite its nearly comprehensive functionality, the Pubby web app has not undergone further development since 2014 [1]. Consequently, the application suffers from an outdated graphical user interface (GUI) and barely performs a search function within the platform [22]. Currently, data access is weighty since it solely relies on URI access, and to address these limitations, the researchers undertook the development of the Pubby web app, leveraging the content derived from the folktales of Mekong countries [23]. By customizing the Pubby web app's graphical user interface (GUI) using the Bootstrap framework, the researchers aimed to enhance its visual appeal and modernize the platform to encompass the creation of a home page to facilitate user navigation and the addition of a search function to simplify data retrieval.

The present research on Linked Open Data (LOD) platforms employed a multidisciplinary approach and incorporated various theoretical foundations to develop effective solutions, including Semantic Web and Ontology Design, Data Linking, User-Centric Design, Knowledge Organization, Open Data Principles, NLP, Information Retrieval, and Evaluation as follows: 1) Semantic Web technologies like RDF and OWL facilitate structured data representation and ontology design principles to enable domain-specific vocabularies for data integration; 2) Data linking techniques establish connections between related resources, enhancing the cohesiveness of the linked data network; 3) User-centric design ensures intuitive interfaces and human-computer interaction theories drive iterative improvements; 4) Knowledge organization and data management frameworks to optimize data structure for better discoverability; 5) Open data principles emphasize openness, accessibility, and reusability; 6) NLP processes to unstructured text data for semantic enrichment; and 7) Evaluation frameworks to assess data coherence and reliability. The integration of these approaches aims to create robust LOD platforms for fostering interdisciplinary research and knowledge exploration. Thus, the objectives of this study are to undertake the development and optimization of a LOD platform to create a platform to organize and present existing knowledge of folktales in the Greater Mekong Subregion countries. With the aim of accomplishing this goal effectively and ensuring its suitability for integration into a LOD environment, the existing knowledge relating to folktales in the Greater Mekong Subregion will be systematically organized. Since folktales are a valuable source of cultural heritage, which manifest an insight into the history, beliefs, and values of the people. In recent years, there has been a growing interest in using LOD to organize and query knowledge of folktales, and LOD is a way of representing data that is linked and reused. This would entail the possibility of discovering relationships between different pieces of data, which can be used to gain a deeper understanding of the data to structure the information in a way that aligns with the requirements and principles of LOD platforms, ensuring its compatibility and seamless integration.

## **2- Research Methodology**

The research methodology was designed were as follows:

### **2-1- Research Tools**

The research tools utilized in the development of the LOD platform were as follows:

#### **2-1-1- Ontology Structure Development**

To establish the ontology structure, the Hozo Ontology Editor Program [24] was employed to facilitate the creation and organization and ensuring the logical and coherent structure.

### ***2-1-2- Identification of Data Source Links***

The process of identifying links with other data sources was carried out using the Google OpenRefine tool [25]. This tool enabled the identification and establishment of connections with external data sources, enhancing the overall richness and context of the LOD platform.

### ***2-1-3- SPARQL Endpoint***

For querying and retrieving data, a SPARQL Endpoint was implemented through the Fuseki Server [26]. This server facilitated the execution of SPARQL queries and provided a seamless interface for interacting with the LOD platform.

## ***2-2- Research Process***

### ***2-2-1- Organizing Knowledge about Folktales of the Mekong Countries***

This stage involved organizing the knowledge system pertaining to folktales of the Mekong countries. To accomplish this, the study employed the knowledge management process outlined in the knowledge of folktales in the Greater Mekong Subregion [27, 28] and the concept of developing a metadata schema for folktale knowledge management in the Mekong countries [23]. These resources served as valuable references in augmenting the ontology structure. The study had divided into two main parts, as described by Tuamsuk et al. [28]: 1) ontology design and development, and 2) ontology documentation, the Hozo Ontology Editor was utilized as the ontology editor for this process.

While various techniques were available, they exhibited structural similarities and shared certain modeling elements included:

- a) Class or concept: Representing a group of entities that share common characteristics.
- b) Subclass or sub-concept: A subdivision of a class or concept.
- c) Relationship: Referring to the connections between concepts and their attributes, roles, parts, or links.
- d) Individuals: Referring to entities that belong to specific concepts, also known as instances.

#### ***Part 1: The Ontology Design and Development Phase, the Following Steps Were Undertaken:***

##### ***Step 1: Categorize Classes and Modifying Existing Knowledge***

This step has initiated with the arrangement of the folktale's characteristics of from the Mekong countries, including ensuring accuracy and completeness by identifying missing knowledge and refining the conceptual framework. The relationships between classes and instances, either within the same hierarchy or across different hierarchies, were carefully considered. This stage was crucial for the classes to possess sufficient accuracy and concreteness, with clearly defined characteristics. During this step, a main class and corresponding subclasses were defined to capture the initial conceptual framework. The main class included attributes such as the title of the folktales, folktale contents, folktale publishers, folktale contributors, tale synopsis, keywords, characters, points of thought, ethnic group, relationship, place, country, language, source, year published, and copyright. The subclasses were established to represent specific attributes like theme, main idea, main characters, subordinate characters, names used by characters to address themselves, academic names, names called by others, and language families

##### ***Step 2: Examine Accuracy and Completeness***

This step involved assessing the accuracy and completeness of each folktale hierarchy within the Mekong countries. This assessment was carried out by analyzing three key points: (a) validating the taxonomic relationship between classes, (b) determining the need for further specification or generalization of existing classes in the original hierarchy, and (c) ensuring the presence of all necessary classes or instances in the original hierarchy. Additionally, modifications were conducted to the original ontology, including adding or removing elements as deemed necessary based on the assessment.

#### ***Part 2: Ontology Documentation:***

Comprehensive documentation is important for ensuring the efficiency and shareability of valuable knowledge. In this stage, the ontology documentation was reused if deemed valid. These resulted in document creation necessitates conducting carefully and precisely to the details, and all generated documentation was carefully archived and preserved. As shown in Figure 1 illustrates that the primary classes and subclasses comprise the folktale ontology in the Greater Mekong Subregion. The visual depiction presented in Figure 1 serves as a succinct and comprehensive overview of the structure and organizational framework of the ontology.



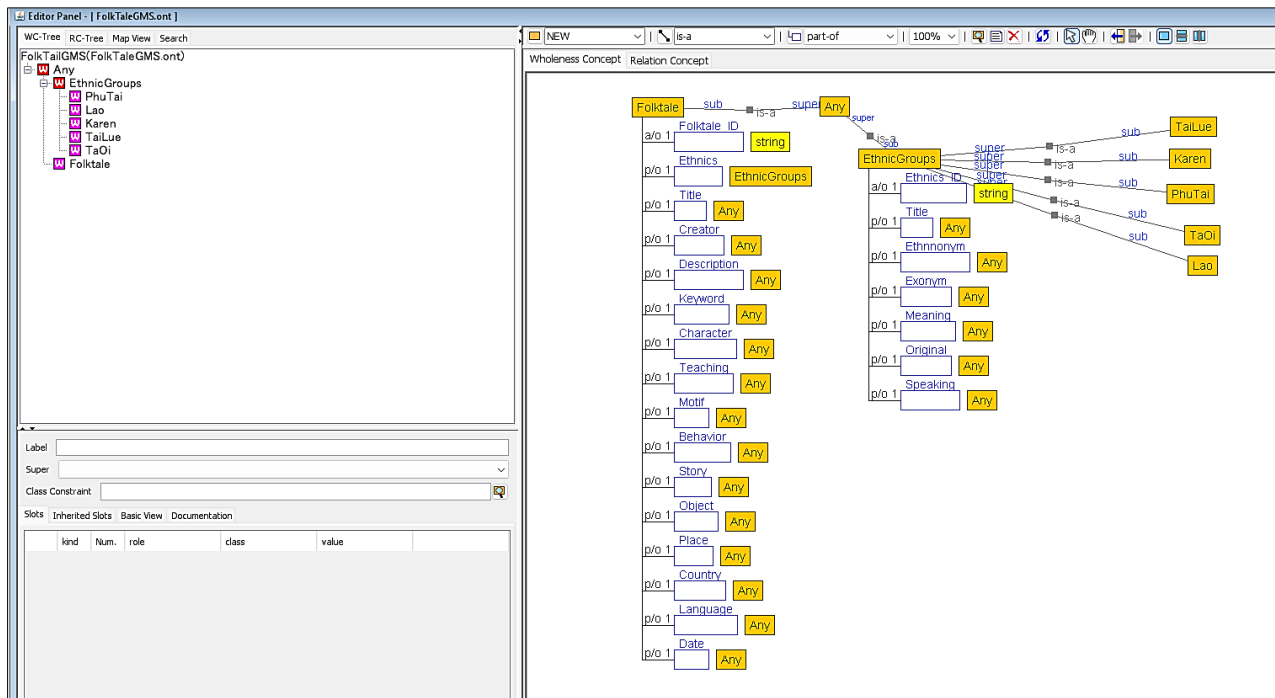


Figure 1. The main classes and subclasses of the folktale ontology in the Greater Mekong Subregion

### 2-2-2- Development for the LOD Platform for the Greater Mekong Subregion Folktales

To actualize the concept of a LOD platform [29], the development process was divided into several distinct sections e.g. the initial step began with the metadata and dataset of folktales preparation [23] in a spreadsheet format, then organized and structured related information about the folktales within a spreadsheet to ensure the data was in a proper format and ready for LOD platform integration, as shown in Figure 2.

FolkTale															
id	Title	Creator	Contributor	Description	Keyword	Character	Teaching	Motif	Behavior	Story	Object	Place	Country	Language	Date
TH001	สนทนาเรื่องพญาผาน	ทินกร	-	พญาผานเป็นนักประพันธ์ชาตินิโศพญา,ผาน พญาผาน	การเมือง	เกี่ยวกับเจลาด	การปกครองของป่า	xxx				ไทย	ภาษาไทย	1975-10	
TH002	กำพร้าหัวทอง	บุญยืน	-	จำวันเขามันเป็นทองลอม มั่นใจ	เจ้าเมือง	ความเชื่อ	การด่า	ความเชื่อ	การด่า	เจ้า	xxx	ไทย	ภาษาไทย	1975-10	
TH003	กำพร้าหมาดำ	กนกจิต	-	กาลครั้งหนึ่งมีหมาตัวหนึ่ง เก่ากว่า,หมาดำ	ความกล้า	การด่า	ความกล้า	การด่า	หมาดำ	xxx		ไทย	ภาษาไทย	1972-10	
TH004	ข้ายักษ์	รังสรรค์	-	-	ยักษ์	ยักษ์	-	-	-	-	xxx	ไทย	ภาษาไทย	1992-11	
TH005	ช้างเจ็ดหัวเจ็ดหาง	วันเพ็ญ	-	มีเด็กชายกำพร้าสองคนพอด ข้าง	ข้าง	-	-	-	-	-	xxx	ไทย	ภาษาไทย	1972-10	
TH006	จำปาสี่ต้น	ศักดิ์ศรี ศรีอักษร	-	มีแม่พญาอยู่เมืองหนึ่งชื่อว่า จำปา	จำปา	ความดี	ความดี	-	นอกเรื่อง	จำปา	YYY	ไทย	ภาษาไทย	1975-10	
LA007	พระทองพระนิม	ไกรสร	-	พระทองพระนิม เป็นพระลูกพี่เะ พญา,ผาน พญาผาน	การเมือง	เกี่ยวกับเจลาด	การปกครองของป่า	xxx				ลาว	ภาษาลาว	1975-10	
LA008	เจ้าอภัย	วงศ์สร	-	เจ้าอภัยเป็นบุตรของ... ยักษ์	ยักษ์	ยักษ์	-	-	-	-	xxx	ลาว	ภาษาลาว	1994-04	
LA009	จำ บักร 4 ขึ้น	ระชอนระเดช	-	ในวันนั้นมีพระขรรค์... จำ บักร จำ บักร	พระขรรค์	ความดี	-	-	-	จำ บักร	YYY	ลาว	ภาษาลาว	1994-04	
LA010	ท้าวโสวัตร	ก. กิ่งแก้ว ป. และ	-	วันหนึ่งท้าวโสวัตรไปเล่นน้ำ... ท้าวโสวัตร, ท้าวโสวัตร	ความรัก	ความรัก	-	-	-	ความพล่	-	ป่า,เมืองลาว	ภาษาไทย	-	
LA011	สุริยะคราส-จันทร์คราส	เดชาโร ธิกขุ (อิน	-	ครอบครัวหนึ่ง มีฐานะยากจน สุริยะคราส สุริยะคราส, ใจความดี	ทำดี	ทำดี	ทำได้	หมูป่าน	นกสีทอง	หมูป่าน	(เวียดนาม	ภาษาวีเียน	1975-10		
VN012	Những người bạn thiên chi	-	-	นิทานเรื่องนกสีทองถึงชีวิตของ... ดัน, อาไร, นู	เชื่อในพร	ความรัก	ทำได้	หมูป่าน	นกสีทอง	หมูป่าน	(เวียดนาม	ภาษาวีเียน	1975-10		

Figure 2. Metadata and dataset of the Mekong countries' folktales

To enhance functionality to LOD platform, the study has employed a Google OpenRefine as a powerful tool for identifying links with other data sources and facilitating the export of files in TTL (Turtle) format. This process has utilized OpenRefine's data manipulation capabilities and support for RDF (Resource Description Framework) data with following details. Firstly, Google OpenRefine serves as an invaluable tool for data cleaning, transformation, and reconciliation., the study has utilized OpenRefine's intuitive interface and powerful functions to preprocess the data and ensure the quality and consistency through data cleaning features, at this stage the study would be able to identify and resolve inconsistencies, duplicates, and errors in the dataset as well as optimizing the data reliability and accuracy. After refining and preparing data, the study had run the OpenRefine to identify associations and connections with other relevant data sources, as exemplified in Figure 3. Due to the OpenRefine supports various reconciliation services and APIs, which enable to match and link data with external datasets.

ID	Ethnic name	Ethnic group DBpedia	Ethnic group Wikidata	Ethnonym	Exonym	Meaning	Origin	Speaking	Speaking from DBpedia	Speaking from Wikidata	ID2	Title	Creator	Contributor	Description	Keyword
1	Lue ลื้อ	<a href="http://dbpedia.org/resource/Dai_people">http://dbpedia.org/resource/Dai_people</a>	<a href="https://www.wikidata.org/wiki/Q840530">https://www.wikidata.org/wiki/Q840530</a>	ลื้อ	ลื้อ		China	Tai-Lao	<a href="https://dbpedia.org/page/Tai_languages">https://dbpedia.org/page/Tai_languages</a>	<a href="https://www.wikidata.org/wiki/Q9211">https://www.wikidata.org/wiki/Q9211</a>	TH001	ลื้อ			ลื้อ เป็นกลุ่มชาติพันธุ์ที่อาศัยอยู่ในพื้นที่ภาคเหนือของประเทศไทยและประเทศจีน มีภาษาพูดของตนเองและใช้ภาษาไท-ลื้อในการสื่อสารกัน	ลื้อ
2	Karen Karen	<a href="https://dbpedia.org/page/Karen_people">https://dbpedia.org/page/Karen_people</a>	<a href="https://www.wikidata.org/wiki/Q271790">https://www.wikidata.org/wiki/Q271790</a>	Karen	Karen		Thai Myanmar	Tibeto-Burman	<a href="https://dbpedia.org/page/Sino-Tibetan_languages">https://dbpedia.org/page/Sino-Tibetan_languages</a>	<a href="https://www.wikidata.org/wiki/Q34064">https://www.wikidata.org/wiki/Q34064</a>	TH002	Karen			Karen เป็นกลุ่มชาติพันธุ์ที่อาศัยอยู่ในพื้นที่ภาคเหนือของประเทศไทยและประเทศจีน มีภาษาพูดของตนเองและใช้ภาษาไท-ลื้อในการสื่อสารกัน	Karen

**Figure 3. The Google OpenRefine implication to identify links with other data sources**

Through this process, meaningful connections have been established and the LOD platform has been enriched with additional information and context from external sources. Furthermore, Google OpenRefine also allows us to export our data files in TTL (Turtle) format as a human-readable serialization format for RDF data, specifically designed for representing RDF triples in a concise and readable manner. To export data in TTL format, compatibility with other linked data tools and systems was examined with the aim of facilitating interoperability and seamless integration with the wider linked data ecosystem. The Turtle, as an extension of N-Triples, provides a syntax that is both compact and expressive, making it suitable for representing complex linked data structures. With OpenRefine's export capabilities, the converted data were refined and reconciled into TTL files, adhering to the Turtle specification. To fulfill the study objectives, the exported data in TTL format not only enables the presentation of data in a standard and widely supported format but also ensures that the essential semantic information encoded in RDF triples is preserved. Since the TTL (Turtle) files generated through the utilization of OpenRefine possess compatibility with various linked data applications, SPARQL endpoints, and knowledge graph frameworks. This compatibility would help with data integration and widespread adoption in the larger linked data community, as shown in Figure 4. The LOD platform would be enhanced by utilizing data cleaning, reconciliation, and export functionalities to identify and establish links with other data sources, as well as enrich datasets with valuable context.

**RDF Schema alignment**

The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.

Base URI: <http://localhost:8080/> [Edit](#)

[RDF skeleton](#) [RDF Preview](#)

This is a sample Turtle representation of (up-to) the first 10 rows

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix ft: <http://www.hozo.jp/owl/FolkTale.owl#> .

<http://localhost:8080/ethnic/1> a ft:EthnicGroups;
foaf:name "Lue ลื้อ";
owl:sameAs <http://dbpedia.org/resource/Dai_people>, <https://www.wikidata.org/wiki/Q840530>;
ft:has_Ethnonym "ลื้อ";
ft:has_Exonym "ลื้อ ลื้อ ไท ไท ลื้อ ลื้อ";
ft:has-Origin "China ";
ft:has_Speaking "Tai-Lao", <https://dbpedia.org/page/Tai_languages>, <https://www.wikidata.org/wiki/Q9211> .
<http://localhost:8080/folktales/TH001> a ft:FolkTale;
foaf:name "ลื้อ";
owl:sameAs "ลื้อ";
ft:has_Creator "ลื้อ";
ft:has_Contributor "-";
ft:has_Character "ลื้อ";
ft:has_Teaching "ลื้อ";
ft:has_Motif "ลื้อ";

```

[OK](#) [Cancel](#)

**Figure 4. Using Google OpenRefine tool to convert data to RDF Schema**

To present LOD via a web browser, the study has combined the tools and technologies, i.e., the Fuseki Server as a SPARQL endpoint and the Jetty Web server as a server for the Pubby web app. These two applications would help to display data in HTML format while incorporating links to other URIs. Therefore, by enhancing the browsing experience for users, the Fuseki Server as our SPARQL endpoint was employed. SPARQL (SPARQL Protocol and RDF Query Language) is a query language specifically designed for querying RDF (Resource Description Framework) data. On the one hand, the Fuseki Server would help to store and manage RDF data efficiently and provide a convenient interface for executing SPARQL queries. With the Fuseki Server, the configuration to store the linked open data has examined whether the data was organized and accessible for retrieval, as well as defined the dataset and specified the permissions and security measures to protect the data for integrity and confidentiality.

Next, Jetty Web Server has been adopted as a server for the Pubby web app, which is a lightweight and highly scalable Java-based web server that allows us to serve web content efficiently. The server provides the necessary infrastructure for hosting the Pubby web app and making it accessible to users via a web browser. Within the Pubby web app, the Fuseki Server capabilities were leveraged to retrieve the desired LOD using SPARQL queries. The retrieved data was then processed and transformed into HTML format, suitable for rendering within the web browser, including appropriate formatting, styling, and presentation with clarity and readability. Crucially, when displaying the LOD in HTML format, the study has incorporated links to other URIs, which enable users to navigate and explore related resources, facilitating a seamless browsing experience. At this stage, the study would enable users to access additional information and traverse the interconnected web of LOD effortlessly to create a robust and efficient system for presenting LOD via a web browser. The integration of SPARQL queries, HTML formatting, and URI links would assist the users in exploring and interacting with the linked open data, enhancing their understanding. Therefore, the LOD platform for folktales in the Greater Mekong Subregion was developed using Google OpenRefine, Fuseki Server, and the Pubby web app, and Google OpenRefine was used to identify the links between different pieces of data about folktales.

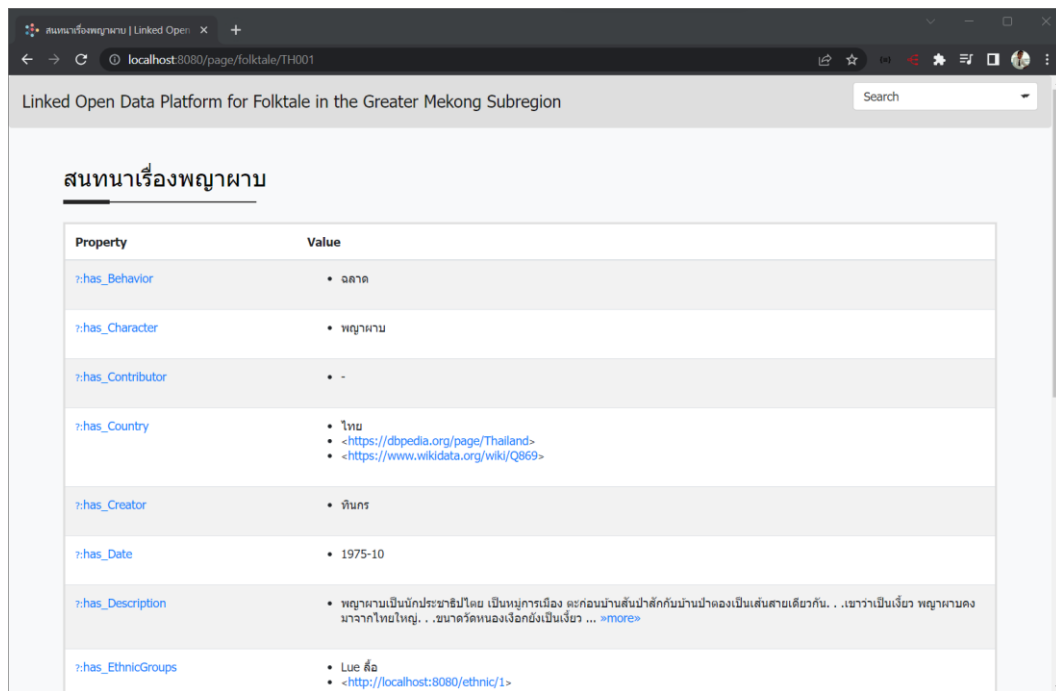
### 3- Results and Discussion

To investigate the specifics of the development of the LOD platform, the study illustrates further information regarding the available source code for reference. The source code for the project can be accessed at the following URL: <https://github.com/treepidokku/pubby-website> with the objective being as follows:

- a) To enhance the user experience of the Pubby web app, the Bootstrap framework was employed to meticulously tailor the graphical user interface (GUI). To leverage Bootstrap's extensive toolkit of pre-designed components and responsive grid system, we adeptly craft a visually appealing and user-friendly interface. Firstly, incorporate Bootstrap's CSS (Cascading Style Sheets) classes to seamlessly style the elements of the GUI. These classes provide consistent and professional-looking typography, buttons, forms, and other interface elements. To utilize Bootstrap's predefined styles, a cohesive and aesthetically pleasing design is needed throughout the application to confirm the process. Additionally, Bootstrap's responsive grid system allows us to create a layout that adapts fluidly to different screen sizes and devices. This ensures that users can seamlessly interact with the Pubby web app on various devices, including desktops, tablets, and smartphones. The responsive grid system also enables the organization of the app's content in a structured manner, optimizing readability and navigation. Moreover, we take advantage of Bootstrap's extensive collection of UI components. These components, such as navigation bars, drop-down menus, modals, and alerts, provide interactive and functional elements that enrich the user experience. By integrating these components into the Pubby web app's GUI, the study aims to provide users with intuitive and efficient ways to interact with the application's features and functionalities. Furthermore, Bootstrap's JavaScript plugins augment the functionality of the GUI. The study has leveraged these plugins to implement dynamic elements like carousels, tooltips, and scrollspy. These interactive features enhance engagement and provide users with an immersive experience while using the Pubby web app. We undertake the transformation of the graphical user interface (GUI) of the Pubby web application into a captivating and user-centric interface through the customizations. The study not only enhances the visual design to be visually appealing but also prioritizes usability and responsiveness, thus enabling a pleasant and seamless user experience throughout their interaction with the platform. The efficacy of these customizations is shown in Figure 5.
- b) To optimize the user experience and ensure accessibility across various devices, the study has designed and developed a responsive home page for the Pubby web app. For incorporating responsive design principles to provide users with a seamless and intuitive browsing experience, regardless of the device they use. Firstly, the study carefully considers the layout and structure of the home page and prioritizes the key information and features that users should have immediate access to upon landing on the app. This may include a brief introduction, a prominent search bar, and navigation elements that guide users to different sections or functionalities within the app. Then, the study has leveraged responsive web design techniques to ensure that the home page adapts fluidly to different screen sizes, which means that whether users access the app on a desktop, laptop, tablet, or smartphone, the content and layout of the home page will automatically adjust to fit the screen dimensions. This would guarantee that users can comfortably interact with the app without the need for excessive scrolling, zooming, or horizontal scrolling. To achieve responsiveness, the study utilized CSS (Cascading Style Sheets) media queries, which allow to define specific styles and layout rules based on the characteristics of the user's device by applying responsive styles such as flexible grid



systems, fluid typography, and media breakpoints. To ensure that the content of the home page rearranges and scales appropriately, maintaining readability and usability across different devices. Moreover, we optimize the home page for touch interactions on mobile devices. This study would confirm that interactive elements, such as buttons and links, have an adequate touch target size to accommodate users' fingertips accurately. Additionally, the study has employed touch-friendly gestures and animations to enhance the user experience and provide intuitive navigation on touch-enabled devices. In terms of performance, this optimized the home page's loading speed by minimizing file sizes, compressing images, and leveraging caching techniques to ensure that users can quickly access and navigate the home page without experiencing frustrating delays or sluggishness. Regarding aesthetics, the study manifested an aesthetically pleasing design for the home page of the Pubby web application, aligning it with the overarching branding and thematic elements. By incorporating suitable color palettes, typography selections, and graphical elements, we strive to create an engaging and visually harmonious experience for users. Our focus lies in prioritizing usability, accessibility, and visual appeal. Furthermore, integrating a responsive design ensures seamless user interaction with the application's features and content across diverse devices, thereby contributing to an elevated user experience, as shown in Figure 6.

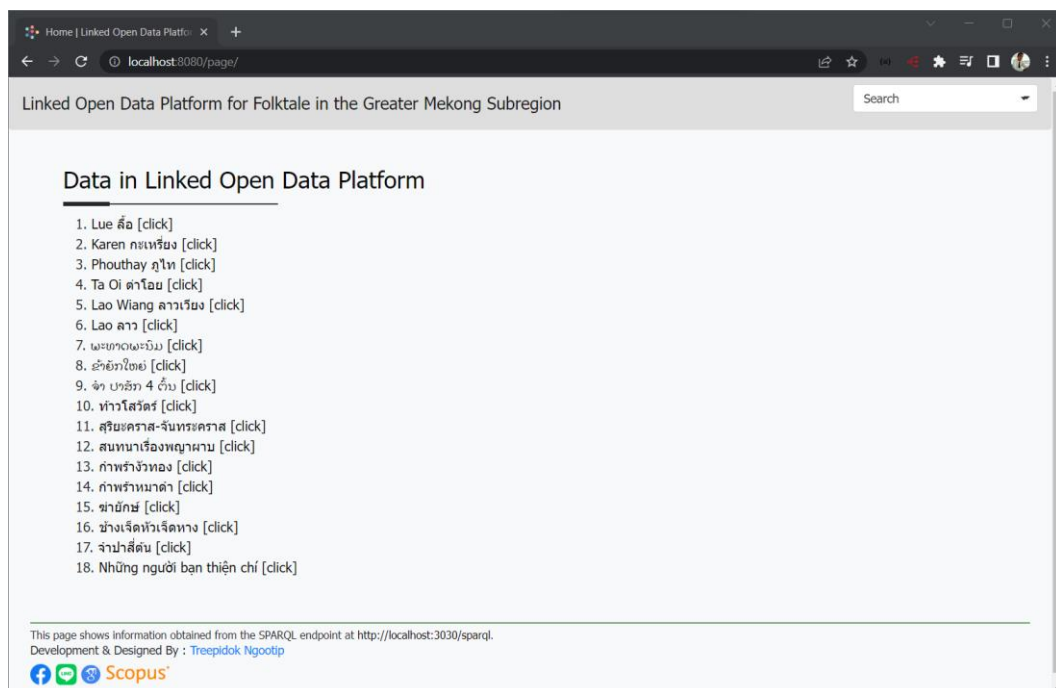


Linked Open Data Platform for Folktales in the Greater Mekong Subregion

สนทนาเรื่องพญาผาน

Property	Value
<a href="#">r:has_Behavior</a>	• ฉลาด
<a href="#">r:has_Character</a>	• พญาผาน
<a href="#">r:has_Contributor</a>	• -
<a href="#">r:has_Country</a>	<ul style="list-style-type: none"> <li>ไทย</li> <li><a href="https://dbpedia.org/page/Thailand">https://dbpedia.org/page/Thailand</a></li> <li><a href="https://www.wikidata.org/wiki/Q869">https://www.wikidata.org/wiki/Q869</a></li> </ul>
<a href="#">r:has_Creator</a>	• ทินกร
<a href="#">r:has_Date</a>	• 1975-10
<a href="#">r:has_Description</a>	<ul style="list-style-type: none"> <li>พญาผานเป็นนักประพันธ์ไทย เป็นเหตุการณ์เมืองละคอนบ้านสันป่าสักกับบ้านป่าดงเป็นเส้นสายเดียวกัน... เขาว่าเป็นเจ้าพญาผานคงมาจากไทยใหญ่... ขนาดวัดหนองเจ็ดยังเป็นเจ้า ... <a href="#">»more»</a></li> </ul>
<a href="#">r:has_EthnicGroups</a>	<ul style="list-style-type: none"> <li>ลue ลื้อ</li> <li><a href="http://localhost:8080/ethnic/1">http://localhost:8080/ethnic/1</a></li> </ul>

**Figure 5. Customization of the graphical user interface (GUI) with Bootstrap framework**



Linked Open Data Platform for Folktales in the Greater Mekong Subregion

Data in Linked Open Data Platform

1. Lue ลื้อ [click]
2. Karen กระเหรี่ยง [click]
3. Phouthay ภูไท [click]
4. Ta Oi ต่าโอย [click]
5. Lao Wiang ลาวเวียง [click]
6. Lao ลาว [click]
7. เขมรเขมร [click]
8. ลื้อ [click]
9. จีน 4 ส่วน [click]
10. ท้าวโสธร [click]
11. สุริยเคราะห์-จันทร์คราส [click]
12. สนทนาเรื่องพญาผาน [click]
13. ก้าวร้าวทอง [click]
14. ก้าวร้าวหมาดำ [click]
15. ชาลักษ์ [click]
16. ข้างเจ็ดหัวเจ็ดหาง [click]
17. จำปาสัก [click]
18. Những người bạn thiện chí [click]

This page shows information obtained from the SPARQL endpoint at <http://localhost:3030/sparql>.  
Development & Designed By : [Treepidok Ngootip](#)

[f](#) [t](#) [s](#) [Scopus](#)

**Figure 6. The Pubby web app responsive home page**

- c) To streamline data search within the Pubby web app platform, the study implemented a convenient search function in the top bar. This feature simplifies the process of finding specific information within the app by combining multiple operations, including querying data from RDF/TTL files, cleaning the data, and displaying the results in a select input search as follows: 1) Querying data from RDF/TTL files via the SPARQL API to initiate the search process; this also utilized the JavaScript language to interact with the SPARQL API running on the Fuseki Server. This API allowed us to query data from RDF/TTL files efficiently. Through the construction of suitable SPARQL queries using JavaScript, we initiated requests to the SPARQL API, facilitating the retrieval of pertinent data aligned with the specified search criteria, as shown in Figure 7; 2) Cleaning data to ensure accurate and relevant search results, the study implemented a data cleaning process. In this step, the study created code that removes excess words or irrelevant information from the retrieved data. This process may involve techniques such as string manipulation, pattern matching, or regular expressions by removing unnecessary words or characters to refine the data set to focus on the most meaningful and relevant information. Furthermore, to store the correct words for future use, the study created an array data structure as a container to hold the cleaned and refined words obtained from the data by organizing the words in an array to facilitate efficient data handling and retrieval during the search process; 3) To display words in the array at Select input search to enable users to interact with the search function, the study presented the cleaned and refined words from the array in a selected input search interface. This selected input allows users to choose from a list of available words as search criteria, providing a user-friendly and intuitive way to explore the data within the Pubby web app platform by displaying the words in a selected input search, then simplifying the search process for users, eliminating the need for manual entry or remembering specific keywords. This approach serves to augment the usability of the search function, rendering it more accessible to a diverse spectrum of users, as shown in Figure 8. In conclusion, the implementation of the search function in the top bar of the Pubby web app platform involves querying data from RDF/TTL files via the SPARQL API, cleaning and refining the retrieved data, and presenting the refined words in a selected input search interface. The combined execution of these operations substantively contributes to a streamlined and effective data search experience within the application, empowering users to locate the desired information swiftly and effortlessly, as shown in Figure 9 as an illustrative framework outlining the functionality of the search function within the Pubby web application.

```

<form class="form-inline mt-2 mt-md-0">
  <div class="row-fluid">
    <select id="search" class="selectpicker" data-show-subtext="true" data-live-search="true" onchange="window.location.href=this.value;">
      <option value="">Search</option>

    <script>
      1 var xmlhttp = new XMLHttpRequest();
      xmlhttp.onreadystatechange = function() {
        if (this.readyState == 4 && this.status == 200) {
          2 let str1 = this.responseText;
          let newStr1 = str1.replace("uri,title", "");
          let newStr11 = newStr1.replace(/http/g, "",http");
          var strArray1 = newStr11.split(",");
          let newStr12 = "";

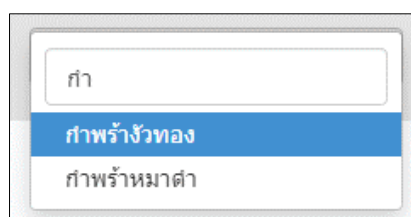
          // Display array values on page
          3 for(var i = 0; i < strArray1.length; i++){
            var strArray11 = strArray1[i].split(",");
            for(var x = 1; x < strArray11.length; x++){
              var option = "<option value="" + strArray11[0] + "">" + strArray11[1] + "</option>"
              document.getElementById('search').innerHTML += option;
            }
          }
        }
      };
      xmlhttp.open("GET", "http://localhost:3030/sparql?query=%23%23+prefix+standard%00%0APREFIX+folktale%3A+%3Chttp%3A%2F%2Fwww.hozo.jp%2Fow
      xmlhttp.send();
    </script>

    </select>
  </div>

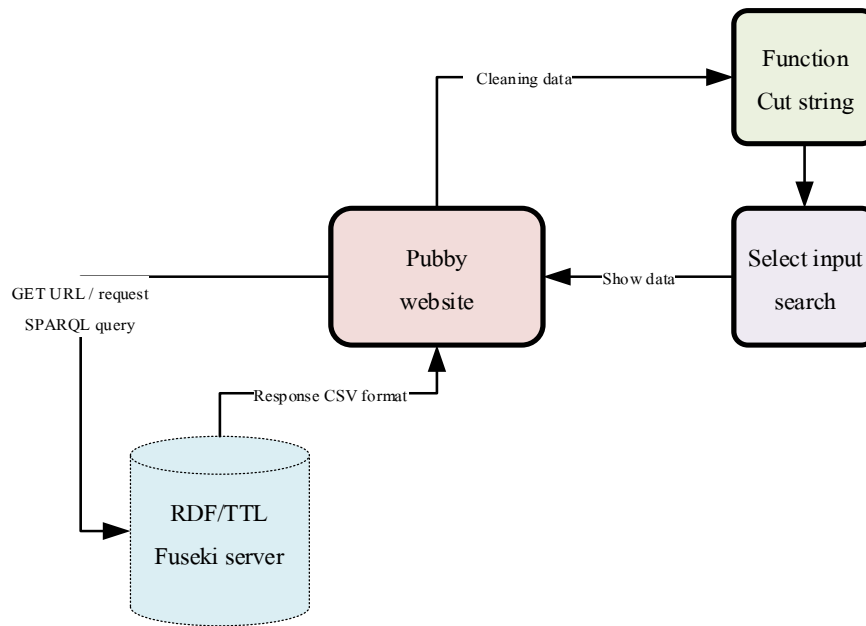
  <script src="https://ajax.googleapis.com/ajax/libs/jquery/1.11.2/jquery.min.js"></script>
  <script src="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.2/js/bootstrap.min.js"></script>
  <script src="//cdnjs.cloudflare.com/ajax/libs/bootstrap-select/1.6.3/js/bootstrap-select.min.js"></script>
</form>

```

**Figure 7.** An example source codes of search function in the Pubby web app



**Figure 8.** The search box function on the top bar of the Pubby web app

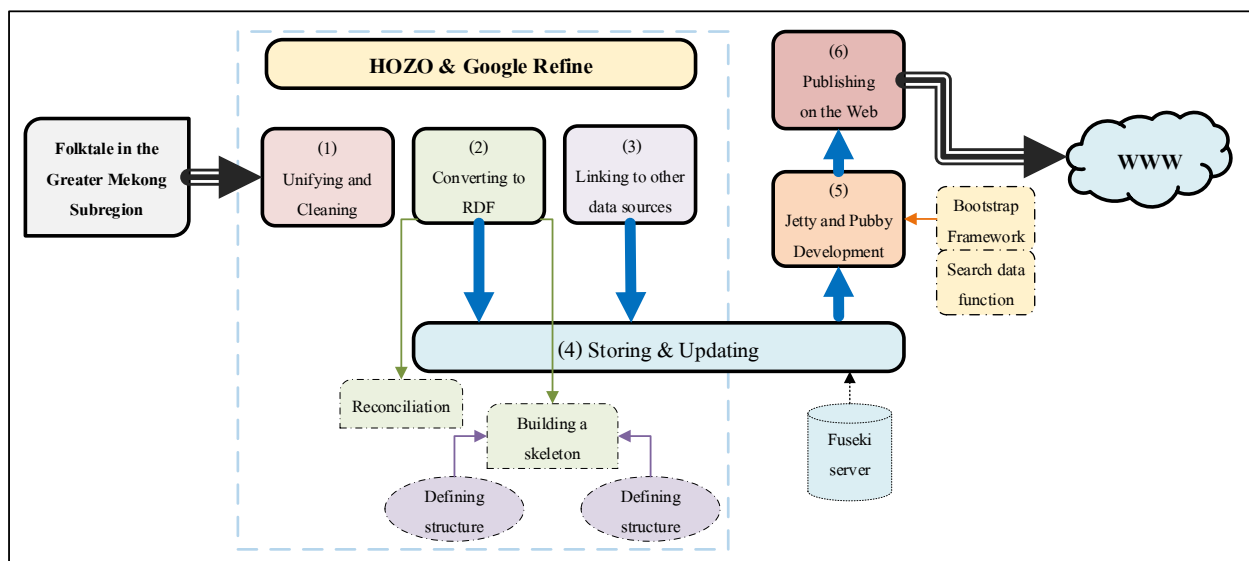


**Figure 9.** A framework of the search function in the Pubby web app

### 3-1-Linked Open Data Framework for the Greater Mekong Subregion Folktales Overview

This section provides a concise overview of the fundamental framework employed for work pieces, storage, data updates, and platform development pertaining to folktales in the Greater Mekong Subregion through the utilization of linked data and web services. The creation of a linked open knowledge framework, specifically tailored for folktales in the Greater Mekong Subregion, serves as a supportive foundation for this research endeavor. The framework encompasses six primary modules, namely: 1) Ingesting the metadata of Mekong countries' folktales [23] into the system, followed by data cleansing and ontology creation via an ontology editor. 2) Transforming the data into RDF format and converting the Mekong Subregion folktales data into RDF. 3) Establishing links with other data sources and configuring the data to be seamlessly connected with external sources like DBpedia and WikiData. 4) Utilizing a repository to store the outcomes obtained from the aforementioned framework [29]. 5) Developing a LOD platform by customizing the graphical user interface (GUI) using the Bootstrap framework to enhance aesthetics and functionality. This involves incorporating a function data search box within the platform, thereby eliminating the previous necessity for users to solely access data through URIs [1]. And 6) Publishing the framework's results on a website, allowing for the display and retrieval of folktale data in the Greater Mekong Subregion in adherence to linked data principles.

The Figure 10 shows a comprehensive LOD framework specifically designed for the folktales in the Greater Mekong Subregion. This framework delineates the systematic approach employed for organizing and structuring the folktales within a LOD environment, facilitating seamless integration and enhanced accessibility of the folktales' information.



**Figure 10.** A linked open data framework for the Greater Mekong Subregion folktales

### 3-2-Linked Open Data Platforms Evaluation

The concept of assessing LOD platforms, as introduced by Candela et al. [30], has been applied in this study to develop an evaluation tool encompassing the following steps:

Step 1: Item-objective congruence (IOC) procedure for expert opinion scale in relation to the LOD platform

For this stage, the assistance of three experts was sought, and a questionnaire comprising 23 items was administered. From these items, seven were considered for item-objective congruence (IOC) assessment, resulting in 21 valid responses. The item-objective congruence (IOC) values obtained from the analysis reached an average of 0.87.

Step 2: Expert opinion scale regarding the LOD platform

Following the evaluation of item-objective congruence (IOC), the 21 valid responses were utilized to construct a questionnaire seeking expert opinions on the LOD platform. In total, three individuals, consisting of two experts in the field of LOD and one expert in the domain of folktales, were involved in assessing their perspectives and opinions regarding the LOD platform.

The expert opinion scale for assessing the LOD platform involved the participation of three experts. The evaluation criteria employed as the key-indicator to determine the statistical data based on Likert's method [31], was as follows:

- 4.51 – 5.00 = The Most suitable;
- 3.51 – 4.50 = Well suitable;
- 2.51 – 3.50 = Moderately suitable;
- 1.51 – 2.50 = Slightly suitable;
- 1.00 – 1.50 = The Least suitable.

Figure 11 and Table 2 show a comprehensive overview of the results obtained from the expert's Linked Open Data (LOD) platform assessment. The assessment was based on the opinions expressed by the experts through an implemented rating scale, evaluating seven data quality criteria and the accuracy was rated as 4.83, indicating a high level of suitability, affirming that the platform delivers precise and reliable information. Reliability scored 4.00, denoting a well-suited level, showcasing the platform's trustworthiness and consistency in delivering data. Integrity was assessed at 4.66, signifying a the most suitable level, indicating that the platform maintains data integrity and coherence. Understandability was rated at 4.50, indicating a well-suited level, affirming that the platform presents data clearly and understandably. Collaboration was scored at 4.44, indicating a well-suited level, emphasizing the platform's capability to foster collaboration and information sharing. Accessibility scored at 4.72, signifying a the most suitable level, highlighting the platform's ease of access and availability to users. Interlinking was also rated as 4.83, reflecting a the most suitable level, illustrating the platform's effectiveness in establishing connections between diverse datasets. The overall average rating of 4.57 indicates a highly suitable evaluation across all criteria, demonstrating the platform's efficacy in meeting expert expectations and requirements. The LOD platform's evaluation was conducted by experts familiar with folktales in the Greater Mekong Subregion. These experts were requested to rate the platform based on seven criteria, namely unity, reliability, integrity, understandability, collaboration, accessibility, and connectedness. The experts unanimously found that the platform was The Most suitable across all seven criteria, affirming its effectiveness in organizing and providing access to knowledge about folktales in the region. The positive assessment by domain experts further solidifies the platform's potential as a valuable resource for scholars and researchers in digital humanities and cultural heritage.

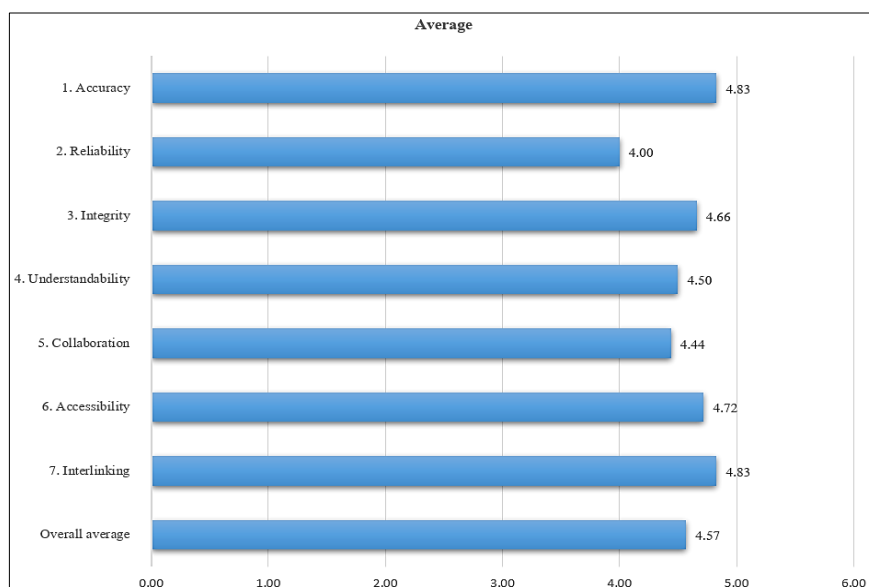


Figure 11. Average shown by items

**Table 2. Presents the results of the expert opinion scale regarding the LOD platform**

Assessment Items	Average	Meaning
<b>1. Accuracy</b>	4.83	The Most suitable
1.1 Accuracy of RDF documents	5	The Most suitable
1.1 Accuracy of RDF documents	5	The Most suitable
1.2 Accuracy of RDF document structure	5	The Most suitable
1.3 Accuracy of N-Triples documents	5	The Most suitable
1.4 The accuracy of the words displayed on the platform	4.33	Well suitable
<b>2. Reliability</b>	4	Well suitable
2.1 What is the reliability level of the KG?	3	Moderately suitable
2.2 What is the reliability level of SPARQL commands?	5	The Most suitable
<b>3. Integrity</b>	4.66	The Most suitable
3.1 Schema integrity	5	The Most suitable
3.2 Content integrity	4.33	Well suitable
<b>4. Understandability</b>	4.5	Well suitable
4.1 Clear resource description	4.33	Well suitable
4.2 Converting data to RDF understandable with self-describing URIs	4.66	The Most suitable
<b>5. Collaboration</b>	4.44	Well suitable
5.1 Avoiding empty nodes and editing RDF documents	4.66	The Most suitable
5.2 Use of external terminology	4.33	Well suitable
5.3 Relationship of proprietary terminology	4.33	Well suitable
<b>6. Accessibility</b>	4.72	The Most suitable
6.1 Possibility reference of resources	5	The Most suitable
6.2 Availability of the KG	4.66	The Most suitable
6.3 Provisioning of public SPARQL endpoint	5	The Most suitable
6.4 Provisioning of an RDF export	4.66	The Most suitable
6.5 Support of content negotiation	4.66	The Most suitable
6.6 Provisioning of KG metadata	4.33	Well suitable
<b>7. Interlinking</b>	4.83	The Most suitable
7.1 Accuracy of the link via owl:sameAs	4.66	The Most suitable
7.2 Accuracy of external URIs	5	The Most suitable
<b>Overall average</b>	4.57	The Most suitable

The weakness and limitations of the Pubby linked open data (LOD) web platform, the below were details and comparisons with the solution of this research developed and resolved. Limited GUI Customization—the Pubby web app did not have extensive GUI customization options. This could be a weakness for users who prefer highly tailored interfaces or organizations that wish to align the platform's appearance more closely with their brand. The new LOD platform developed in this research offers a user-friendly graphical user interface (GUI) with extensive customization options. This improvement allows users to have a more tailored experience and aligns the platform's appearance with their preferences or organizational branding. Lack of Advanced Query Options—the Pubby web app needs the potential development of a graphical user interface (GUI) enabling users to generate ontologically valid queries using dropdown lists based on available properties for each class. Meanwhile this may improve user accessibility, it might not cater to users who require more complex or specific queries. The need for advanced query options could limit the platform's utility for researchers who need to perform intricate data analysis [32].

The new LOD platform addresses the limitation of lacking advanced query options by implementing a search function that supports users' specific requirements. This enhancement caters to researchers who need to perform intricate data analysis, enabling them to access relevant data more efficiently. Challenges in Semantic Enrichment: The Pubby web app did not mention leveraging natural language processing (NLP) techniques for the semantic enrichment of the dataset. Semantic enrichment can be challenging, especially for diverse and contextually rich content like folktales. The quality of semantic enrichment depends on the accuracy of NLP algorithms and the availability of high-quality language resources, which might pose obstacles. To solve this problem, the new LOD platform incorporates semantic search to



enhance the dataset's contextual richness. By leveraging NLP algorithms, the platform facilitates the extraction of meaningful information from folktales text, providing valuable insights to users. Scalability and Performance: While the Pubby web app is not explicitly mentioned in the provided information, any LOD platform, including Pubby, may encounter scalability and performance issues as the dataset grows or as more users access the platform simultaneously. Ensuring efficient data retrieval and response times can be critical for a positive user experience. The new LOD platform emphasizes supporting scalability and high performance. This enhancement ensures efficient data retrieval and response times, contributing to a positive user experience, even as the dataset grows or more users access the platform simultaneously.

The research distinguishes itself from the existing literature by presenting a comprehensive approach to the development of the LOD platform for folktales in the Greater Mekong Subregion. It incorporates ontology analysis and design principles, addressing data quality and user experience. The new LOD platform's GUI customization, advanced query options, semantic enrichment, and scalability contribute to a more effective and user-friendly platform. Moreover, the study's methodology ensures the appropriateness of implementing the LOD platform, substantiating the platform's potential for organizing and querying folktales knowledge.

This research aims to develop a linked open data (LOD) platform for folktales in the Greater Mekong Subregion countries. The LOD platform was developed using Google OpenRefine, Fuseki Server, and The Pubby web app and this study was achieved through a research and development process specifically designed for this purpose. This study utilized the principles of ontology analysis and design, as proposed by Noy & McGuinness [33]. The development process encompassed seven key steps, including: 1) Clearly defining the objectives of the ontology scope, 2) Distributing terminology, 3) Classifying and establishing a hierarchy of concepts, 4) Defining relationships, 5) Specifying characteristics, 6) Determining perspectives on class characteristics, and 7) Generating sample data within classes. Wathanti et al. [20] also followed these ontology analysis and design principles, along with the design analysis and development of a linked open data platform, while creating an ontology of knowledge about development of ontology for the knowledge of Heet-Sib-Song custom. Additionally, a Pubby interface serves as an alternative front end for Linked Data navigation without needing SPARQL. Future developments include a GUI enabling users to generate ontologically valid queries using dropdown lists based on available properties for each class. These improvements aim to make JazzCats more accessible to experts and scholars in the digital humanities field [19].

To ensure a modern and user-friendly interface for data access, the Pubby web app was selected for development, specifically focusing on the graphical user interface (GUI) with the utilization of the Bootstrap framework. This choice aimed to enhance convenience and provide users with seamless access to data, whether through the platform's main page or via a search function. To evaluate the developed LOD platform, seven data quality criteria were employed, as proposed by Candela et al. [30]. These criteria consisted of accuracy, reliability, integrity, understandability, collaboration, accessibility, and connectedness. The platform was evaluated by experts, who found that it was The Most suitable for the purpose of organizing and querying knowledge of folktales in the Greater Mekong Subregion. The suitability assessment of the LOD platform was conducted based on these criteria. The methodology applied in this research study ensures the appropriateness of implementing the developed LOD platform for folktales in the Greater Mekong Subregion, as evidenced by the findings obtained through the proposed methodology. The results of this research show that it is possible to develop a LOD platform for folktales that is both effective and user-friendly. The LOD platform can be further applied to other content, such as traditional knowledge, cultural heritage, and historical documents.

In future research, there were three potential directions for future research on the development of LOD platforms personalized for folktales in the Greater Mekong Subregion countries: 1) Enhancement of Interlinking and Linked Data Integration: Further research could explore advanced techniques and methodologies for interlinking folktale data with external linked data sources. This would involve identifying additional relevant datasets and finding effective ways to integrate them into the LOD platform. The focus could be on improving the richness and interconnectedness of the platform's data by leveraging existing ontologies and knowledge graphs; 2) User Experience and Interface Design: Future research could explore user-centric design aspects to improve the user experience of the LOD platform. This could involve conducting user studies, gathering feedback from users, and incorporating their preferences and suggestions into the design and functionality of the platform. Exploring innovative visualization techniques, search functionalities, and user-friendly features could contribute to enhancing user engagement and satisfaction; 3) Semantic Enrichment and Multilingual Support: Extending the semantic enrichment of the folktales dataset and ensuring multilingual support would be another valuable research direction. This could involve leveraging natural language processing (NLP) techniques to extract semantic information from the folktales text, enriching the platform's ontology with additional domain-specific concepts, and addressing challenges related to language variations and translations within the Greater Mekong Subregion. These research directions aim to further enhance the development, usability, and data quality of LOD platforms for folktales in the Greater Mekong Subregion, advancing the understanding and accessibility of this cultural heritage.

## 4- Conclusion

The study focuses on developing a specialized Linked Open Data (LOD) platform for folktales in the Greater Mekong Subregion counties, and the platform was constructed using Google OpenRefine, Fuseki Server, and the Pubby web app, following a rigorous research and development process incorporating ontology analysis and design principles proposed by Noy and McGuinness. The LOD platform underwent evaluation by experts, employing seven data quality criteria, and received high suitability ratings for organizing and querying folktales in the Greater Mekong Subregion. The successful development of this platform showcases its potential application in preserving and exploring other forms of cultural heritage, traditional knowledge, and historical documents. Looking ahead, the research suggests three key research directions for personalized LOD platforms for folktales in the region, as follows: 1) Enhancement of Interlinking and Linked Data Integration: The study recommends exploring advanced techniques to interlink folktale data with externally linked data sources. Integrating additional datasets and leveraging existing ontologies will enhance data richness and interconnectedness within the platform; 2) User Experience and Interface Design: Improving user experience is crucial, and future research should focus on user-centric design through user studies and feedback collection. Innovations in visualization techniques, search functionalities, and user-friendly features will enhance user engagement and satisfaction; 3) Semantic Enrichment and Multilingual Support—the research suggests extending semantic enrichment through natural language processing (NLP) techniques to extract valuable information from folktale texts. Additionally, enriching the platform's ontology with domain-specific concepts and addressing language variations will ensure multilingual support. By pursuing these research directions, LOD platforms for folktales in the Greater Mekong Subregion can become more accessible, usable, and informative. Advancements in development, usability, and data quality will contribute to a deeper understanding and enhanced accessibility of this valuable cultural heritage. Moreover, the successful application of LOD principles to folktales paves the way for similar platforms in other cultural and knowledge domains, fostering collaboration and knowledge dissemination.

In conclusion, this research would significantly contribute to the broader landscape of linked open data, showcasing its potential in preserving and exploring cultural heritage and traditional knowledge since the LOD platform is a valuable resource for researchers, experts, and enthusiasts interested in the rich cultural narratives of the Greater Mekong Subregion, and its future enhancements promise an even greater impact in the digital humanities and in the near future.

## 5- Declarations

### 5-1-Author Contributions

Conceptualization, W.C. and M.B.; methodology, W.C., T.N., and M.B.; software, W.C. and T.N.; validation, W.C., P.M., and M.B.; formal analysis, W.C. and M.B.; investigation, W.C. and P.M.; resources, W.C., T.N., and M.B.; data curation, W.C. and T.N.; writing—original draft preparation, W.C., T.N., P.M., and M.B.; writing—review and editing, W.C., T.N., P.M., and M.B.; visualization, W.C. and T.N.; supervision, W.C. and M.B.; project administration, W.C. and M.B.; funding acquisition, W.C. and M.B. All authors have read and agreed to the published version of the manuscript.

### 5-2-Data Availability Statement

Data sharing is not applicable to this article.

### 5-3-Funding

This research was supported by the Research Fund for Supporting Lecturers to Admit High Potential Students to Study and Research on His Expert Program Year 2020 of Khon Kaen University and the Program Management Unit for Human Resources & Institutional Development, Research and Innovation-CU (grant number B05F630072).

### 5-4-Institutional Review Board Statement

Not applicable.

### 5-5-Informed Consent Statement

Not applicable.

### 5-6-Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

### 3- References

- [1] Cyganiak, R., & Bizer, C. (2007). Pubby—A linked data frontend for SPARQL endpoints. Available online: <https://github.com/cygri/pubby> (accessed on May 2023).
- [2] Coetzee, P., Heath, T., & Motta, E. (2008). SparqPlug: Generating Linked Data from Legacy HTML, SPARQL and the DOM. 22 April, 2008, Beijing, China.
- [3] Kumar, V. (2010). Linked Data. A best practice for better knowledge transaction. Knowledge Transactions. Documentation Research and Training Centre, Bengaluru, India.
- [4] Reda, R., & Carbonaro, A. (2018). Design and Development of a Linked Open Data-Based Web Portal for Sharing IoT Health and Fitness Datasets. Proceedings of the 4<sup>th</sup> EAI International Conference on Smart Objects and Technologies for Social Good, 43-48. doi:10.1145/3284869.3284890.
- [5] Tilahun, B., Kauppinen, T., Keßler, C., & Fritz, F. (2014). Design and Development of a Linked Open Data-Based Health Information Representation and Visualization System: Potentials and Preliminary Evaluation. JMIR Medical Informatics, 2(2), 3531. doi:10.2196/medinform.3531.
- [6] Ali, S., Zada, I., Mehmood, Z., Ullah, A., Ali, H., & Ullah, M. (2022). Publishing and Interlinking COVID-19 Data Using Linked Open Data Principles: Toward Effective Healthcare Planning and Decision-Making. Mathematical Problems in Engineering, 2022. doi:10.1155/2022/4792909.
- [7] D'Agostino, M., Samuel, N. O., Sarol, M. J., de Cosio, F. G., Marti, M., Luo, T., Brooks, I., & Espinal, M. (2018). Open data and public health. Revista Panamericana de Salud Pública, 42. doi:10.26633/rpsp.2018.66.
- [8] Dhayne, H., Farhat, H., & Kilany, R. (2019). Data Link Discovery Tools for Big Linked Data: A comprehensive study. BDCSIntell, 5-12.
- [9] Boilson, A., Weston, D., & Connolly, R. (2018). Evaluation of a European Data Analytic Framework. Proceedings of the 32nd International BCS Human Computer Interaction Conference, 32. doi:10.14236/ewic/hci2018.140.
- [10] Wu, P., Ma, F., & Yu, S. (2023). Using a linked data-based knowledge navigation system to improve teaching effectiveness. Interactive Learning Environments, 31(5), 3273–3284. doi:10.1080/10494820.2021.1925925.
- [11] Monika Rani, H. G., Sapna, R., & Mishra, S. (2018). An investigative study on the quality aspects of linked open data. Proceedings of the 2018 International Conference on Cloud Computing and Internet of Things. doi:10.1145/3291064.3291074.
- [12] Peña, O., Aguilera, U., & López-de-Ipiña, D. (2014). Linked open data visualization revisited: a survey. Semantic Web Journal, IOS Press, 1-16.
- [13] Dimitrakis, E., Sgontzos, K., & Tzitzikas, Y. (2020). A survey on question answering systems over linked data and documents. Journal of Intelligent Information Systems, 55(2), 233–259. doi:10.1007/s10844-019-00584-7.
- [14] Dimitrakis, E., Sgontzos, K., Mountantonakis, M., & Tzitzikas, Y. (2020). Enabling Efficient Question Answering over Hundreds of Linked Datasets. Information Search, Integration, and Personalization. ISIP 2019. Communications in Computer and Information Science, 1197, Springer, Cham, Switzerland. doi:10.1007/978-3-030-44900-1\_1.
- [15] Cuijuan, X., Wei, L., & Lei, Z. (2018). Implementation of a Linked Data-Based Genealogy Knowledge Service Platform for Digital Humanities. Data and Information Management, 2(1), 15–26. doi:10.2478/dim-2018-0005.
- [16] Hawkins, A. (2022). Archives, linked data and the digital humanities: increasing access to digitised and born-digital archives via the semantic web. Archival Science, 22(3), 319–344. doi:10.1007/s10502-021-09381-0.
- [17] Zeng, M. L. (2019). Semantic enrichment for enhancing lam data and supporting digital humanities. Review article. Profesional de La Informacion, 28(1). doi:10.3145/epi.2019.ene.03.
- [18] Hyvönen, E., Heino, E., Leskinen, P., Ikkala, E., Koho, M., Tamper, M., ... & Mäkelä, E. (2016). WarSampo Data Service and Semantic Portal for Publishing Linked Open Data About the Second World War History. The Semantic Web, Latest Advances and New Domains, ESWC 2016, Lecture Notes in Computer Science, 9678, Springer, Cham, Switzerland. doi:10.1007/978-3-319-34129-3\_46.
- [19] Nurmikko-Fuller, T., Bangert, D., & Abdul-Rahman, A. (2017). All the things you are: Accessing an enriched musicological prosopography through JazzCats. Digital Humanities, 1-4.
- [20] Wathanti, S., Chansanam, W., & Tuamsuk, K. (2020). Thai custom information sharing on the internet by linked data techniques. Journal of Critical Reviews, 7(8), 1398–1402. doi:10.31838/jcr.07.08.281.
- [21] Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., ... & Sheets, D. (2006, November). Tabulator: Exploring and analyzing linked data on the semantic web. Proceedings of the 3<sup>rd</sup> international semantic web user interaction workshop, 6 November, 2006, Athens, United States.

- [22] Regalia, B., Janowicz, K., & Mai, G. (2017). Phuzzy.link: A SPARQL-powered Client-Sided Extensible Semantic Web Browser. Proceedings of the Third International Workshop on Visualization and Interaction for Ontologies and Linked Data Co-Located with the 16th International Semantic Web Conference, CEUR Workshop Proceedings, Vol. 1947, ISWC 2017, October 22, 2017, 34-44, Vienna, Austria.
- [23] Kwiecien, K., Chansanam, W., Supnithi, T., Chitiyaphol, J., & Tuamsuk, K. (2021). Metadata Schema for Folktales in the Mekong River Basin. *Informatics*, 8(4), 82. doi:10.3390/informatics8040082.
- [24] Kozaki, K., Kitamura, Y., Ikeda, M., & Mizoguchi, R. (2002). Hozo: An Environment for Building/Using Ontologies Based on a Fundamental Consideration of “Role” and “Relationship”. *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web, EKAW 2002, Lecture Notes in Computer Science*, 2473. Springer, Berlin, Germany. doi:10.1007/3-540-45810-7\_21.
- [25] OpenRefine. (2023). A Free, Open Source, Powerful Tool for Working with Messy Data. Google, San Francisco, United States. Available online: <https://openrefine.org/> (accessed on May 2023).
- [26] Apache Jena Fuseki. (2023). The Apache Software Foundation. Forest Hill, New York, United States. Available online: <https://jena.apache.org/documentation/fuseki2/> (accessed on November 2023).
- [27] Tuamsuk, K., Kaewboonma, N., Chansanam, W., & Leopenwong, S. (2016). Taxonomy of folktales from the greater mekong sub-region. *Knowledge Organization*, 43(6), 431–439. doi:10.5771/0943-7444-2016-6-431.
- [28] Tuamsuk, K., Chansanam, W., & Kaewboonma, N. (2018). Ontology of folktales in the greater mekong subregion. *International Journal of Metadata, Semantics and Ontologies*, 13(1), 57–67. doi:10.1504/IJMSO.2018.096454.
- [29] Chansanam, W., Tuamsuk, K., & Chaikhambung, J. (2020). Linked Open Data Framework for Ethnic Groups in Thailand Learning. *International Journal of Emerging Technologies in Learning (IJET)*, 15(10), 140. doi:10.3991/ijet.v15i10.13337.
- [30] Candela, G., Escobar, P., Carrasco, R. C., & Marco-Such, M. (2022). Evaluating the quality of linked open data in digital libraries. *Journal of Information Science*, 48(1), 21–43. doi:10.1177/0165551520930951.
- [31] Likert, R. (2017). The Method of Constructing an Attitude Scale. *Scaling*, 233–242, Routledge, Milton Park, United Kingdom. doi:10.4324/9781315128948-23.
- [32] Saengrith, W., Viriyavejakul, C., & Pimdee, P. (2022). Problem-Based Blended Training via Chatbot to Enhance the Problem-Solving Skill in the Workplace. *Emerging Science Journal*, 6, 1-12. doi:10.28991/ESJ-2022-SIED-01.
- [33] Noy, N. F., & McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology*. Stanford University, California, United States.