

Linked Open Data Framework for Ethnic Groups in Thailand Learning

<https://doi.org/10.3991/ijet.v10i15.13337>

Wirapong Chansanam (✉), Kulthida Tuamsuk, Juthatip Chaikhambung
Khon Kaen University, Khon Kaen, Thailand
wirach@kku.ac.th

Shigeo Sugimoto
University of Tsukuba, Ibaraki, Japan

Abstract—The key significant worldview of the Semantic Web is Linked Open Data, another period of the World Wide Web that capacities to carry suggestions to information. An enormous number of both public and private foundations have distributed their information following the Linked Open Data philosophies, or have done as such with information from different associations. To this degree, since the generation and production of Linked Open Data are thorough designing procedures that require high consideration so as to achieve high caliber, and since experience has uncovered that current general guidance is not constantly adequate to be applied to each area, this paper presents a lot of guidance system for creating and distributing Linked Open Data with regards to ethnic groups in Thailand to outside (TEG-LOD Framework). This framework offers an exhaustive depiction of the undertakings to perform, including a run-down of steps, tools that help in accomplishing the errand, different alternatives for achievement of the assignment, and best practices and proposals. Also, this paper exhibits a pilot model on the generation and distribution of Linked Open Data about ethnic groups in Thailand, adhering to the available guidance, where the ethnic groups in Thailand are the property of the Princess Maha Chakri Sirindhorn Anthropology Center (SAC) have been made and distributed as Linked Open Data.

Keywords—Linked Open Data, Framework, Ethnic Groups in Thailand, Learning

1 Introduction

The resulted studies of ethnic groups in Thailand are published in books, articles, research reports, and master thesis and doctoral dissertation. Especially educational institutions those are located near the area where there are ethnic groups or minorities living in densities such as Chiang Mai University found that there is a lot of research on ethnic groups in Northern of Thailand. In addition, a research center has been set up to study about specific ethnic groups such as the Center for Ethnic Studies and Development, Chiang Mai University and Princess Maha Chakri Sirindhorn Anthro-

polology Centre (SAC) etc. Most of the research is presented in the form of printed publications and stored in the library of those institutions, so it is difficult to access information. Moreover, the content in this area should be conserved and published for the people in the country regardless of any ethnicity learning and take understanding the ethnic groups among others live together in this country and around the world [1]. Another problem with the ethnic groups in Thailand resource is that it has no link to an external dataset, which is as useless as using HTML to write a Web document, print it and send it by fax or air mail: the grace of HTML is the links, and without them it is a technology that makes no sense. Similarly, the merit of RDF is the links, which, coupled with the fact that triples are used to represent information, will make it possible that one day the whole web behaves like a universal database. So to use RDF and not link to other datasets, it is better not to use it: the vocabulary standard format, by itself, is perfectly valid. So the SAC database, having no links to other datasets or other datasets to it, is completely outside the global open data network (<http://lod-cloud.net/>) and therefore: It cannot be discovered. And also it cannot combine information with ease. In many cases, subject URIs connecting to ideas of information association frameworks are not accessible as Linked Data. When all is said in done, it might happen that a knowledge organization systems is not by and by accessible on the site, that they not utilizing semantic organizations, or that the web documentation is lacking. This is a basic confinement regarding the ease of use of a KOS by outsiders [2] For example, if the Thailand government has a URI for each consulate, the data could be integrated following the Linked Data philosophy and assuming that the negotiation was implemented correctly.

The advancement of new innovations and the progressing development of society have realized the presence of better approaches to create and share data [3], which includes more straightforwardness and wide access to data, particularly to information services. So as to stay aware of this improvement, a few tools have been actualized to view and process information, for the most part for open access or open information in spite of continually being open substance. Workgroups have been built up for these two subjects, for example, Open Access Foundation, Open Content Alliance or Data Documentation Alliance. Disregarding not being various, papers regarding the matter have not taken long to show up articles, books, and administrative writing just as associations that empower them. Fundamental, articles are those distributed under the backings of the Open Access Foundation since 2007 [4].

LOD generation and distribution are comprehensive designing methodology that request high consideration so as to achieve high caliber and, along these lines, some broad guidelines and best practices have been created to this date. In any case, Villazón-Terazas and colleague argue that, in spite of the fact that it is conceivable to have a general guidance, professionals ought to depend on various methods, innovations, and instruments for a specific area [5]. In addition, [4] Villazón-Terazas and associates argue that current guidance doesn't cover all the necessary strides with enough detail and including the related technologies.

The intrigue displayed in this paper is generally useful, as in they can apply to the expansive range of assorted situations. Notwithstanding, these directions have been created having as a primary concern qualities that are explicit and valuable to the

computerized humanities situation [6], [7]. These incorporate information authorizing, legitimate consistence, ethnic groups in Thailand and Open Data necessities [8], and real apparatuses to be utilized.

Moreover, the paper also shows an instantiation of the LOD creation and distribution methodology through the conversion into LOD of a dataset about ethnic groups in Thailand linked. The selected dataset of 73 ethnic groups in Thailand comes from the Princess Maha Chakri Sirindhorn Anthropology Centre (SAC) (Princess Maha-Chakri Sirindhorn Anthropology Center, 2015) includes data about ethnic groups in Thailand information.

This paper is prepared as follows. Section 2 illustrates related research endeavor. Sections 3 is a research methodology, Section 4 is an evaluation result, Section 5 is showing a framework, together with describe in detail. Lastly, Section 6 gives some discusses instructions academic and future work.

2 Related Works

Divergent works have investigated the advantages and capabilities of utilizing the LOD approach for coordinating and improve AEC information as uncovered by [9-12]. Other related work, for example, the work by Törmä and associates uncovered in [13], as of now calls attention to explicit research issues around there (e.g., connect type demonstrating and interface generation). In this fragment we assessment existing works in regards to the generation of LOD in the AEC field and furthermore existing general and cross-area writing about LOD generation and production. On the LOD side, we should make reference to key productions, for example, Heath and Bizer's book for driving the LOD generation and distribution process [14] and resulting works, for example, the results of the LOD2 project [15]. These speak to the beginning stage for following the way toward adding to the LOD activity; be that as it may, as contended in the presentation, existing general guidance doesn't give a particular degree of subtleties and doesn't consider explicit qualities of a specific space and related devices and systems. To that degree, a few assets may be particularly relying upon the current field, as has occurred in different regions (e.g., cultural heritage) where area specialists together with LOD engineers have obliged tools, strategies, and guidance to their particular necessities. In fact, we could recognize that LOD distribution in the area of ethnic groups in Thailand information is in its underlying stages. Since experience uncovers that practices are excessively broad and insufficient to be straightforwardly applied to each and every area, despite everything it needs methodological guidance supporting its improvement towards a settled and repeatable procedure and giving clear models in the current space.

3 Research Methodology

This segment shows the instruction for the generation of LOD for some existing data by explanation the different tasks to be performed in the process. The LOD generation procedure consists of eight tasks. After a data source which will be converted

to LOD is carefully chosen and access to that data source is acquired, the license has to be evaluated in order to define the terms of use. Next, the data source is examined in detail and a URI is identified. Subsequently, ontology for annotate the data is established and the data is transformed into the RDF format. Finally, generated data are linked to data from other LOD dataset. Ensuing, we explain each phase of the LOD generation procedure in detail.

3.1 Publish the dataset and the ontology on the web

The aim of this phase is to make available through the Web generation process main products, that is, the ontology and the RDF dataset. A SPARQL, the query language that we introduce in this study, allows users to express queries over LOD on the WWW [16]. This phase should cautiously follow existing principles and best practices in order to accomplish the preferred added value for the publisher. In specific, both the ontology and the RDF dataset should be published in a way that follows to the LOD principles. Furthermore, the publication process must be associated with the anticipated access strategies; to this finish, both the HTTP stack and LOD technologies provide the mechanisms of access control to do so. For example, the publisher could decide to enable access exclusively within a specific local network, to require credentials, and so on.

As well specified RDF repositories have many other alternatives for keeping RDF, such as using a relational database system or NoSQL database system [17] for an experimental assessment of current solutions.

We have chosen to put the RDF dataset into a specialized RDF repository; in particular, the data have been uploaded into an Apache Jena Fuseki Server. It is essential to have in cognizance that in this phase, the data are accessible on the local web. As well, the ontology developed for example has been published online [18].

In order to enable HTTP access to the data, a front-end of LOD has been chosen and configured. In specific, we have selected the Jetty and Pubby implementation of the LOD API specification. This frontend guarantees access via HTTP to our data and enables content mediation to allow users to request the data in various formats.

The last phase in the practice has been to enable access to the RDF store settled up in the first phase. For this, we have configured our Apache Jena Fuseki store to be manageable thru the SPARQL HTTP protocol and have enabled public access. This open access allows everyone to query our repository using the SPARQL language, but it is essential to note that this access could be limited using standard HTTP security mechanisms and a more specialized configuration of the repository. The file containing the ethnic groups in Thailand RDF data is also available online.

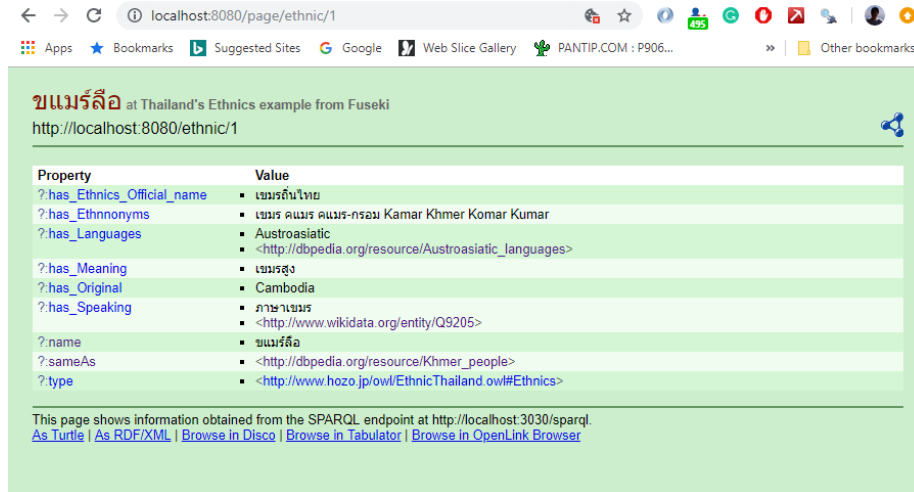


Fig. 1. Screenshot of Ethnic groups in Thailand LOD Pubby web page displaying information linking to the DBpedia and WikiData SPARQL endpoint

4 Evaluation

We have performed examines to assess the dependability of the important data, which have been produced by the proposed strategy [19]. We have directed count the exactness of the proposed system by physically checking whether the information recovered utilizing the accepted word are pertinent to the ethnic groups in the first source. This assessment procedure checks whether the recently returned source recovered utilizing the expected word in the ethnic groups LOD are in reality about indistinguishable ethnic groups from the first query string. In this testing, we re-start look inside the ethnic groups LOD utilizing terms as query strings and afterward measure the accuracy. The exactness in this examination point to the level of unwavering quality of the recently returned records. Something else, the exactness shows the portion of the recently returned records that are pertinent to the first query string. Subsequently, we have analyzed the pertinence of the recently returned records in the ethnic groups LOD to the first query string.

The figure 2 is an example of the RDF graph of the turtle file of the ethnic groups LOD. The base URI for the ethnic groups' LOD is <https://localhost:8080/ethnic/>, the base URI is shown as the prefix ethnic. The RDF graph was generated by the "Visualization of RDF graph Turtle, Microdata, JSON-LD, RDF/XML, TriG (kanzaki.com)".

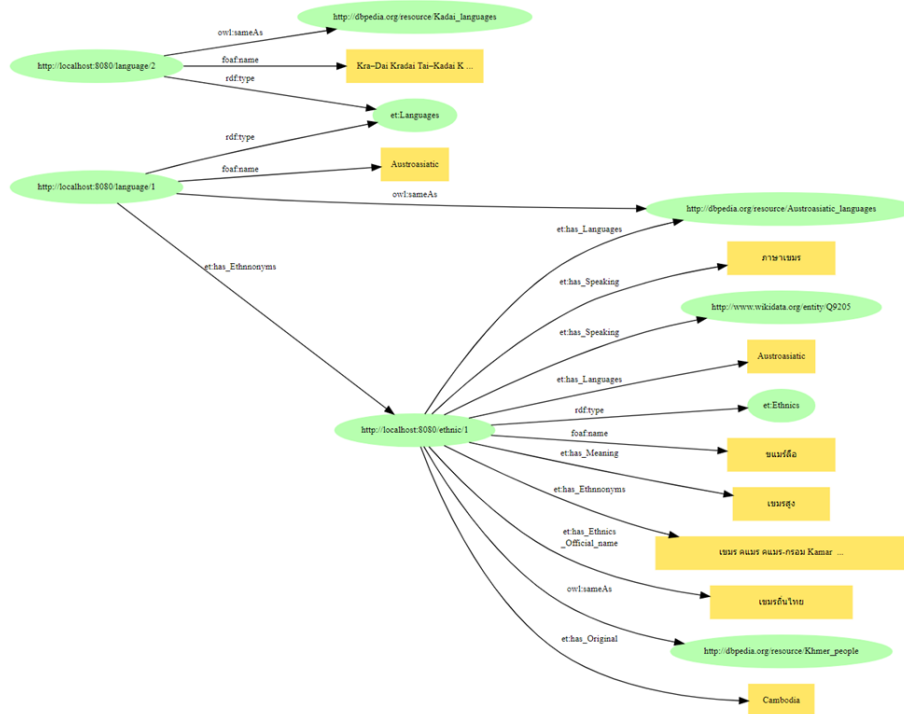


Fig. 2. A snap shot of an example Thailand ethnic groups RDF graph

5 A Thailand Ethnic groups Linked Open Data Framework for Learning: TEG-LOD

This segment, we shows the main of our framework for the production, storing and updating, and utilization of Thailand ethnic groups data using Linked Data principles in combination with Web-based services. We have constructed a framework, called Thailand ethnic groups Linked Open Data: TEG-LOD, that contributes support to this research, and which is summarized graphically in Figure 3. Our framework consists of the following five main modules: i) Unifying and Cleaning, which performs a gathering dataset to the system then cleaning and also creating a vocabulary by ontology editor tool; ii) Converting to RDF, transformation from Thailand ethnic groups data to RDF; iii) Linking to other data sources, which sets connections with others Linked Data and enables us to enrich Linked Data with attributes collected from DBpedia and WikiData; iv) Repository, which stores the obtained results after applying our framework components; and v) Publishing on the Web, which allows the display and querying of Thailand ethnic groups data using Linked Data principles in combination with Web-based services.

With the purpose of attendant the utilization of our framework, three situations have been defined: production and connection, storing and updating, and utilization.

These scenarios consider access limitations and needs related to data. The first situation uses three components (gathering, transforming, and linking) of our framework. The second situation of our framework proposal is storing and updating with RDF repository, which is executed in our framework through SPARQL endpoint. On the other hand, the third situation is associated with utilization, which is implemented in our framework through visualization and data query.

Next, we concisely define these modules, which are flexible and were verified with a case study in the digital humanities domain, as a working example, using the diagram shown in Figure 3 and data sources defined in following section.

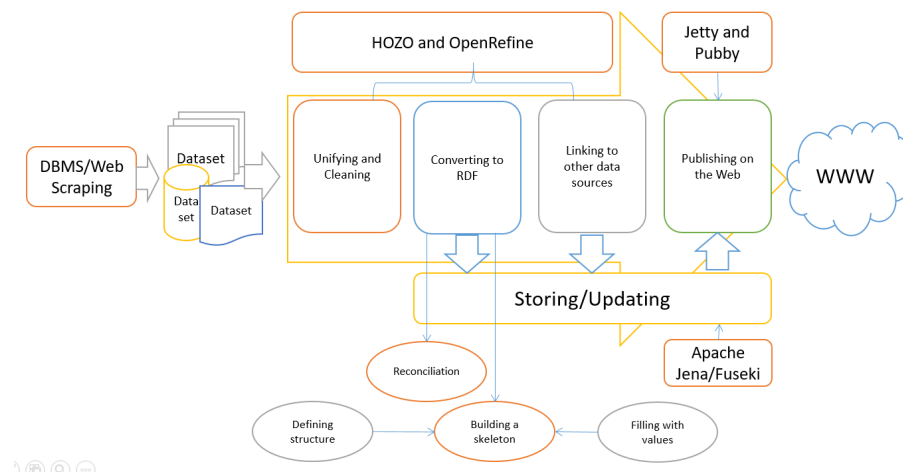


Fig. 3. A framework for production, storing and updating, and utilization's Thailand ethnic groups LOD

5.1 Production and connection

In this situation, gathering, transformation, and linking components are conducted. The appropriateness of the various segments in the work process relies upon two perspectives: i) unique information might be changed and distributed without impediments, or ii) there exist a few information get to limitations, yet halfway change of some applicable components might be executed. Next, we give subtleties identified with the parts of this situation.

Gathering data with web scraping: Any information source frequently contains certain references to other information types, and Thailand ethnic groups' information sources are no special case. These references are progressively important when we manage master area information sources. In any case, it is frequently hazardous for non-expert clients to comprehend and utilize these obscure information [20]. In this sense, as indicated by [21], the target of the Thailand ethnic groups gathering process is to recover such data and make it unequivocal. Our framework does the emergence of this point of view by methods for social affair new information from accessible

web scratching systems, and afterward, these information are utilized to enhance existing Linked Data.

Transforming Thailand ethnic groups data to RDF: A transformation process of Thailand ethnic groups data from the website of Princess Maha Chakri Sirindhorn Anthropology Centre (SAC), concretely CSV files, to RDF is performed by this module of our framework. We chose for the utilization of RDF as the ordinary structure for the Thailand ethnic gatherings datasets to be distributed, since we need to consolidate various organizations of our datasets (databases, HTML reports, CSVs, and so on.), to abstain from utilizing exclusive arrangements, and in light of the fact that we are looking for a Linked Data approach. As portrayed in the past area, RDF is one of the standard dialects in which data must be made accessible, as per Linked Data standards. The inspiration for this is it offers a few advantages, for example, the arrangement of an extensible outline, de-referenceable URIs, and as RDF joins are composed, safe blending (connecting) of various datasets [22].

In this change procedure, we suggest abusing every one of the upsides of Linked Data through a total change from the first Thailand ethnic gatherings information source to RDF, when there is finished access to information. Likewise, our methodology proposes two distinct other options, when information gets to is restricted.

We suggest leading this procedure by utilizing, reusing, or creating ontologies, despite the fact that Linked Data can be produced with or without the utilization of a specific vocabulary since simply changing information to RDF doesn't consolidate any semantics, as brought up by [23]. Besides, including ontologies in this procedure permits making express the significance of ideas in the datasets utilized, making a consolidated model for the considered datasets (utilizing normal and shared vocabularies) and making it simpler to peruse and get to Thailand ethnic groups data [24].

In order to conduct the transformation process, our framework contains one function, called RDF extension module in OpenRefine, which allow transforming CSV files into RDF according to standard vocabularies. This tool is based on: i) the open-source name OpenRefine and RDF extension plugin, with notable modifications and substantial enhancements to meet interoperability needs in other Thailand ethnic groups data formats and services; ii) Apache Jena (<https://jena.apache.org/>), a widely used Java framework for developing Semantic Web applications, tools, and servers.

In our framework, an OpenRefine (included RDF extension) work as Web applications, where ethnic groups and language family relations associated with ethnic groups features are transformed into RDF following the DBpedia SPARQL vocabulary through different configuration formularies (Web forms). This tool allow uploading ontologies related to a knowledge domain to generate RDF with explicit meaning of concepts in the datasets used (using common and shared vocabularies). Furthermore, OpenRefine and RDF extension have some optional competencies. These operations are carried out thanks to the integrated OpenRefine and RDF extension module and according to user specifications for the source and goal operation. An example of the RDF skeleton and RDF Preview of this transformation process using OpenRefine is shown in Figure 4.



Fig. 4. An OpenRefine transformation snap short

Using as above tool, we have created RDF data from different data sources. With respect to CSV files, they come from different local and global bodies related with the digital humanities domain.

In all these cases, RDF data are generated, using OpenRefine and RDF extension, respectively, according to common and shared vocabularies used in each domain, such as the Ethnic groups in Thailand ontology, Dublin Core Metadata (<https://dublincore.org/documents/dcmi-terms/>), Web Ontology Language (<http://www.w3.org/2002/07/owl/>) and cross-domain DBpedia (<https://wiki.dbpedia.org/services-resources/ontology>) ontology. Listing 1. Show how Thailand ethnic groups, DBpedia and WikiData information are integrated for our framework.

Linking data from Thailand ethnic groups data sources: The fourth Linked Data guideline is: "Incorporate connects to different URIs, with the goal that they can find more things." Therefore, an expanding number of literary datasets are distributed as RDF charts and connected to other outer datasets by equal assets distinguish in different datasets [25]. It uncovers that the estimation of information and its utility increment when it is increasingly interrelated with other information [14]. The aftereffect of this connecting procedure is frequently a rundown of owl:sameAs interfaces between elements of each dataset. These connections can be found utilizing a few devices that offer innovative help, for example, OpenRefine and RDF augmentation, which have likewise begun to remember some ethnic gatherings for the connections disclosure process. In this way, we have taken improvement in the advantages of OpenRefine to give extra focal points to our structure in the interlinking procedure.

This apparatus is utilized for finding joins between created RDF from Thailand ethnic groups information and other information on the Web of Data (DBpedia and WikiData).

Therefore, in our study case perspective, we have used OpenRefine and RDF extension for interlinking our generated RDF previously with information about ethnic groups name, meaning, address, custom, language family, and etc. from SAC (<https://www.sac.or.th/databases/ethnic-groups/>) and DBpedia (<https://wiki.dbpedia.org/>) and WikiData (<https://query.wikidata.org/>). For that, we have configured OpenRefine using different similarity metrics and, based on the “official name” of classes, this discovery process of links is accomplished. An example of the setting links between both RDF data sources is shown in Listing 1.

Listing 1. An RDF instance with a link to DBpedia and WikiData

```
<http://example.org/ethnic/1> a
<http://www.hozo.jp/owl/EthnicThailand.owl#Ethnics> ;
  foaf:name "ขแมร์ลือ" ;
  <http://www.hozo.jp/owl/EthnicThailand.owl#has_Meaning> "เขมรสูง" ;
  <http://www.hozo.jp/owl/EthnicThailand.owl#has_Ethnonyms> "เขมร คแมร คแมร-กรอม Kamar Khmer Komar Kumar" ;
  <http://www.hozo.jp/owl/EthnicThailand.owl#has_Ethnic_s_Official_name> "เขมรถิ่นไทย" ;
  owl:sameAs <http://dbpedia.org/resource/Khmer_people>
;
  <http://www.hozo.jp/owl/EthnicThailand.owl#has_Original> "Cambodia" ;
  <http://www.hozo.jp/owl/EthnicThailand.owl#has_Speaking> "ภาษาเขมร" , <http://www.wikidata.org/entity/Q9205> ;
  <http://www.hozo.jp/owl/EthnicThailand.owl#has_Languages>
<http://dbpedia.org/resource/Austroasiatic_languages> ,
"Austroasiatic" .
  <http://example.org/language/1>
<http://www.hozo.jp/owl/EthnicThailand.owl#has_Ethnonyms
> <http://example.org/ethnic/1> .
```

It is important to notification that OpenRefine focuses just on RDF data, sets owl:sameAs links, and deals with different name. However, within the Thailand ethnic groups information domain, many features are represented by complex ethnic groups and are collected in different types of ethnic groups resources (formats), for example, Our framework takes into account these issues by allowing the establishment of links between two interoperability universes, such as Linked Data and Web-based services.

The DBpedia and WikiData often maintain directories of public ethnic groups SPARQL endpoints built from Web-based services listed in their registries and play

the role of discovery node [26]. Therefore, the first step of the workflow starts by discovering ethnic groups in DBpedia catalog services and/or public lists of ethnic groups Web-based services. Once we have selected the DBpedia SPARQL related to our interest domain, the URL of a service is provided to our framework. Such URLs may have been discovered by a user by browsing some of the as above sites or may have been sent to the user by a peer, or may have been crawled automatically.

5.2 Storing and updating

After RDF generated, the framework stores them into our storehouse, an RDF triple store (see Figure 3), where yields can be questioned through an Apache Jena Fuseki SPARQL HTTP Endpoint. From the viewpoint of our contextual investigation, we have sent Thailand ethnic groups, since we have discovered that it offers an astounding tradeoff between load and query execution furthermore the help of the abilities that DBpedia and WikiData SPARQL gives (by Virtuoso). This triple store SPARQL graph name to query is available at <http://localhost:3030/ethnic4.ttl> repository. However, as mentioned previously, there exist other RDF triple stores that support SPARQL, such as Apache Marmotta, OpenLink Virtuoso, Strabon, and USeekM.

As the last point, following the above work process (see Figure 3), the acquired yields (RDF records) are put away in another diagram in the RDF triple store (Terse RDF Triple Language: Turtle document) related with our system, where results can be questioned. Essentially, when RDF information is overseen by outsiders, we suggest distributing these out-contributes the equivalent Linked Data vault where gathered.

5.3 Utilization

Data is ready to be discharged once they have been gathered, transformed, and linked (see Figure 3). The last target of connecting and opening this information is that clients can utilize semantics apparatuses and Linked Data advances in an organized manner to look, break down, imagine, or assimilate automatically the entire information accessible. Consequently, applications sending over the produced Thailand ethnic groups Linked Data is required to advantage this information and give agreeable GUIs to clients [24]. There are a few apparatuses and methods accessible to exploit Thailand ethnic groups Linked Open Data through graphical interfaces, for example, Jetty and Pubby, Drupal, and so on. Notice that apparatuses make web frontend and change Linked Data into Web-based services. In this manner, these services can be questioned utilizing the SPARQL convention legitimately through the referenced instruments.

In this situation, our framework choose a Jetty and Pubby, a help that tunes in to Web-based services demands and changes over these into the Apache Jena Fuseki Server. After the SPARQL query is handled, the Jetty and Pubby get the RDF result set from a triple store, encodes it as an XML archive, and returns it to the clients. In view of this tools, we execute an improvement in the query handling, since our framework permits setting SPARQL questions on an RDF triple store and indicating a refreshing of the outcomes on Web-based services consequently. This is an elective

manner by which our structure permits setting associations between two interoperability universes.

So as to empower access to this affiliation, our framework has a Web-based application dependent on Jetty and Pubby. This application enables us to show information distributed as Web-based services and query Linked Data sources utilizing SPARQL end-point. These SPARQL endpoint questions permit the generation of "elements" data since every datum inquiry shows a refreshed Thailand ethnic groups between identified with the acquired consequences of each inquiry.

Moreover, the conveyed design related to our system permits Thailand's ethnic groups query to be executed on the information. For instance, the SPARQL query in Listing 1 would get classes inside the ethnic groups of "Khmer", with names related to their official names.

By the by, the genuine capability of our TEG-LOD framework proposition is fascinating when we do a combination of the distributed datasets and, hence, connecting between both interoperability universes. As an occasion, a client may join ethnic groups and language family information and endeavor the SPARQL part connected with them. Along these lines, the client may recover all classes situated in the examination zone with their official name, their related depiction inside Thailand's ethnic groups (dataset in Linked Data), their relationship with various geology and language family arrangements present in the investigation region, and their settlement.

This query may effectively be adjusted to give examination concentrated on undeniable classes, related to solid ethnic groups, or various sorts. Also, the information can be utilized by controlling its ethnic groups part through DNA (deoxyribonucleic corrosive) examination, movement, and communicating in language investigation.

The mix of datasets and interoperability universes (Web-based services and Linked Data), upheld by the semantics gave by the connections, is one of the principle favorable circumstances of our TEG-LOD framework proposition. This makes it simpler to promote dress issues related to worldwide information [27], which require a cross-disciplinary approach, and permits opening information stockpiling tower encouraging their reuse across organizations and networks.

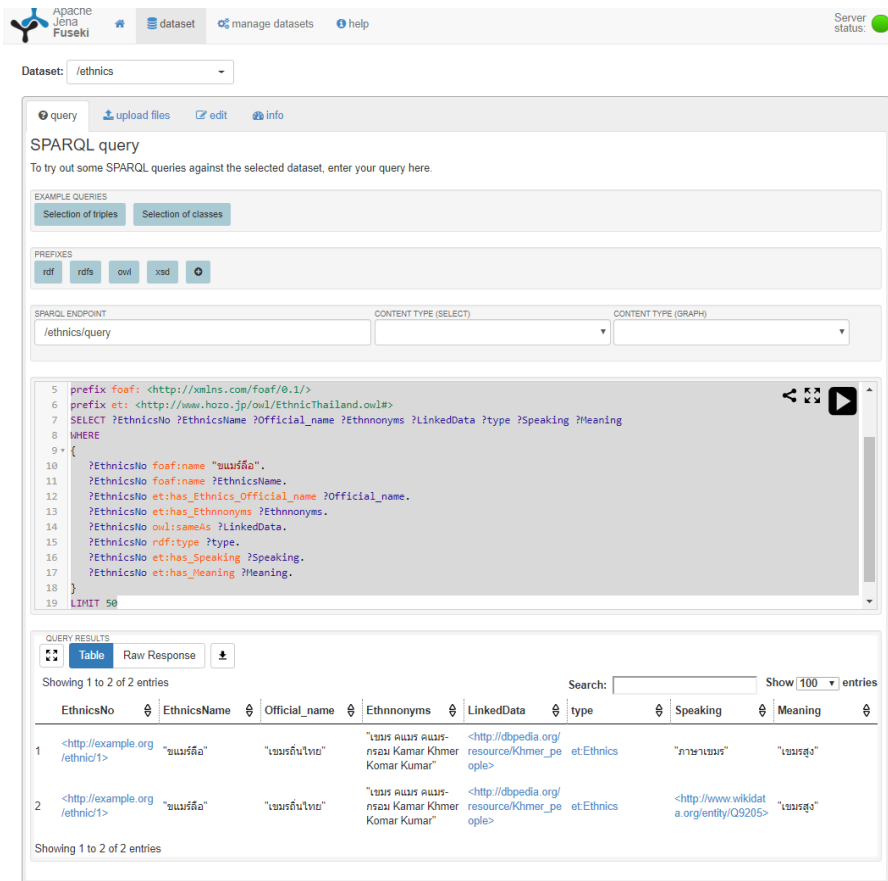


Fig. 5. A snap short result of SPARQL query sample

6 Discussion and Conclusion

As discussed in the Introduction segment, although some broad guidance for LOD generation and distribution exists, experience has indicated that such broad guidance isn't constantly adequate so as to be applied to each area. So as to conquer this issue, area situated guidance should be created. Such guidance will in general location space explicit qualities and give area related models, which help the network to all the more likely comprehend Linked Data advances and may prompt their quicker appropriation.

This paper introduces a lot of directions for LOD generation and distribution of the TEG-LOD system, together with one pilot model in the area of ethnic groups in Thailand for learners. By giving itemized portrayals of each assignment in the generation and production forms, these directions help both private and open associations that work with information about ethnic groups and related different areas in Thailand in

creating LOD formal prepared existing information and in distributing the created information as indicated by the most recent gauges.

This paper likewise exhibits a pilot model case of how to utilize the TEG-LOD system so as to produce and distribute ethnic groups in Thailand's information as five star LOD, specifically, the ethnic groups in Thailand's information from the SAC. This model helps the crowd from various associations to increase a better understanding of the procedures of LOD generation and distribution, subsequently guaranteeing the highest caliber of the yields of these procedures.

In this paper we have completed the entire Linked Data publication cycle: we start from gathering a dataset that we upload to a server and the server provides users with the HTML interface or data in RDF. Through this process we can appreciate some of the advantages of Linked Data. We can create links to external data and thus integrate information through the existing web infrastructure. This process is much simpler and more efficient than, for example, integrating the ethnic groups database with the Wikipedia database at hand. As this dataset uses vocabularies (ontologies) already known as FOAF and OWL, the automatic agent can easily process the information, since it already knows in advance the meaning of, for example, `has_part_of`. That is, the semantics of the information is computationally explicit, represented in a standard language (OWL), so that the internal logic of the program is lightened, moving to the ontology (in this case ethnic group: et vocabulary).

Keep in mind that all this is "built with hand", we have not taken into account neither the efficiency nor the security, since the objective was to have it working (In fact, the Pubby configuration file leaves much to be desired). To implement this in production, the Pubby configuration file would have to be changed at many points, the most important being the SPARQL endpoint and the mapping of external URIs to the URIs of the dataset (from `[http://localhost:8080/page/ethnic/1]` that requests the agent to `[http://www.sac.or.th/databases/ethnic-groups/]` that is in the dataset), use a triple store like Virtuoso, a web server like Tomcat, etc. In production, it should also be taken into account that there are different options for negotiating content (`http://linkeddatabook.com/editions/1.0/#htoc11`).

The TEG-LOD framework presented in this paper is aimed to help researchers, analysts, and experts, who interested in cultural heritage and digital humanities areas in exploiting Linked Data technologies. Since it is sensible to expect that such advancements are new to target experts, future work will manage to make a lot of services for encouraging the use of Linked Data innovations. Such services will help professionals in embracing these advances, and in this manner make focal points for their associations and cross assortment information area.

Lastly, we expect the SAC's committee in Thailand and stakeholders to actively take part and to exploit the advantages of Linked Data technologies by generating and publishing their data as five star LOD. To that extent, the TEG-LOD framework as demonstrated in this paper is a valuable resource to achieve this objective.

7 Acknowledgement

This research has been funded by the Digital Humanities Research Group (DHRG), Khon Kaen University.

8 References

- [1] Chaikhambung, J. (2017). Development of Semantic Ontology of Knowledge on ethnic groups in Thailand. (Doctoral dissertation). Khon Kaen University, Khon Kaen. <https://doi.org/10.20319/pijss.2018.41.105110>
- [2] Díaz-Corona, D., Lacasta, J., Nogueras-Iso, J., Latre, M., & Zarazaga-Soria, F. (2019). Profiling of knowledge organisation systems for the annotation of Linked Data cultural resources. *Information Systems*, 84,17-28., from <https://doi.org/10.1016/j.is.2019.04.008>
- [3] Daradkeh, Y. I., ALdhaifallah, M, and Namiot, D. (2016).On Data Sharing in Educational Processes. *International Journal of Emerging Technologies in Learning*. 11(4), 28-31, from <https://doi.org/10.3991/ijet.v11i04.5298>
- [4] Heus, P. & Gregory, A. (2010). Maximizing the potential of data: Modern IT tools, best practices, and metadata standards for sbe sciences. NSF SBE 2020 paper, 1-6.
- [5] Villazón-Terrazas, B. et al. (2012). Publishing linked open data - There is no one-size-fits-all formula. In the Proceedings of the European Data Forum. Copenhagen, Denmark.
- [6] Radulovic, F. et al. (2014). D4.2: Requirements and guidelines for energy data publication. Tech. Rep., READY4SmartCities Consortium December 2014.
- [7] Radulovic, F. et al. (2015). Guidelines for linked data generation and publication: An example in building energy consumption. *Automation in Construction*, 57, 178-187.
- [8] Localidata. (2014). Ciudades Inteligentes, Datos Abiertos (Open Data). Retrieved May 2019, from https://www.idee.es/resources/presentaciones/JIIDE14/20141106/SmartCities_CTN178.pdf
- [9] Pauwels, P., De Meyer, R. & Van Campenhout, J. (2011). Interoperability for the design and construction industry through Semantic Web technology. *Lecture Notes in Computer Science*, 143-158. https://doi.org/10.1007/978-3-642-23017-2_10
- [10] Abanda, F. et al. (2013). Exploring the relationships between Linked Open Data and Building Information Modelling. In Proceedings of the Sustainable Building Conference, Coventry University, UK, 176-185.
- [11] Madrazo, L. & Costa, G. (2012). Open product modelling and interoperability in the AEC sector. In Proceedings of the 1st International Workshop on Linked Open Data in Architecture and Construction (LDAC 2012), Ghent, Belgium, 4-6.
- [12] Törmä, S. (2014). Web of building data integrating IFC with the web of data. In Proceedings of the 10th European Conference on Product and Process Modelling-eWork and eBusiness in Architecture, Engineering and Construction, Vienna, Austria. <https://doi.org/10.1201/b17396-27>
- [13] Törmä, S., Oraskari, J. & Hoang, N. V. (2012). Distributed transactional building information management. In Proceedings of the 1st International Workshop on Linked Data in Architecture and Construction (LDAC 2012), Ghent, Belgium, 9-11.
- [14] Heath, T. & Bizer, C. (2011). *Linked data: Evolving the web into a global data space* (1st ed.). Synthesis lectures on the Semantic Web: Theory and Technology. Morgan & Claypool. <https://doi.org/10.2200/s00334ed1v01y201102wbe001>

- [15] Auer, S., Bryl, V. & Tramp, S. (2014). Linked Open Data-Creating knowledge out of inter-linked data: Results of the LOD2 Project, Vol. 8661, Springer, 2014. <https://doi.org/10.1007/978-3-319-09846-3>
- [16] Hartig, O. & Pérez, J. (2016). LDQL: A query language for the Web of Linked Data. *Journal of Web Semantics First Look*, 41, 9-29. <https://doi.org/10.1016/j.websem.2016.10.001>.
- [17] Cudré-Mauroux, P. et al. (2013). NoSQL databases for RDF: An empirical evaluation. In the Proceedings of the 12th International SemanticWeb Conference, Sydney, NSW, Australia, 310-325.
- [18] Rezgui, K., Mhiri, H. and Ghédira, K.(2014).An Ontology-based Profile for Learner Representation in Learning Networks. *International Journal of Emerging Technologies in Learning*. 9(3), 16-25, from <https://doi.org/10.3991/ijet.v9i3.3305>
- [19] Biligsaikhan, B, Takeo K, Fuminori K, & Akira M. (2013). Linked data driven multilingual access to diverse Japanese Ukiyo-e databases by generating links dynamically, *Literary and Linguistic Computing*, 28(4), 522–530, <https://doi.org/10.1093/lilc/fqt058>
- [20] Tandy, J., van den Brink, L., & Barnaghi, P. (Eds.) (2017). *Spatial data on the web: Best practices*. Retrieved from <https://www.w3.org/TR/sdw-bp/>
- [21] Lehmann, J., Athanasiou, S., Both, A., Rojas, A. G., Giannopoulos, G., Hladky, D., ..., & Zaslowski, V. (2015). Managing geospatial linked data in the GeoKnow project. In T. Narock & P. Fox (Eds.), *The semantic web in earth and space science: Current status and future directions* (Vol. 20, pp. 51– 78). Amsterdam, the Netherlands: IOS Press.
- [22] Omitola, T., Koumenides, C. L., Popov, I. O., Yang, Y., Salvadores, M., Szomszor, M., & Shadbolt, N. (2010). Put in your postcode, out comes the data: A case study. In L. Aroyo, G. Antoniou, E. Hyvönen, A. Teije, H. Stuckenschmidt, L. Cabral, & T. Tudorache (Eds.), *The semantic web: Research and applications* (Lecture Notes in Computer Science, Vol. 6088, pp. 318– 332). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-642-13486-9_22
- [23] Janowicz, K., Schade, S., Bröring, A., Keßler, C., Maué, P., & Stasch, C. (2010). Semantic enablement for spatial data infrastructures. *Transactions in GIS*, 14(2), 111– 129. <https://doi.org/10.1111/j.1467-9671.2010.01186.x>
- [24] Vilches-Blázquez, L. M., Villazón-Terrazas, B., Corcho, O., & Gómez-Pérez, A. (2014). Integrating geographical information in the linked digital Earth. *International Journal of Digital Earth*, 7(7), 554– 575. <https://doi.org/10.1080/17538947.2013.783127>
- [25] Feliachi, A., Abadie, N., Hamdi, F., & Ateazing, G. A. (2013). Interlinking and visualizing linked open data with geospatial reference data. In Proceedings of the 8th International Conference on Ontology Matching (Vol. 1111, pp. 237– 238). Sydney, Australia: CEUR.
- [26] Lopez-Pellicer, F. J., Renteria-Agualimpia, W., Noguera-Iso, J., ZarazagaSoria, F. J., & Muro-Medrano, P. R. (2012). Towards an active directory of geospatial web services. In J. Gensel, D. Josselin, & D. Vandenbroucke (Eds.), *Bridging the geographic information sciences* (Lecture Notes in Geoinformation & Cartography, pp. 63– 79). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-642-29063-3_4
- [27] Zhang, L. (2016). Study on the Application of Web Information Retrieval in the Teaching of Language Translation. *International Journal of Emerging Technologies in Learning*. 11(4), 114-119, from <https://doi.org/10.1016/j.is.2019.04.008>

9 Authors

Wirapong Chansanam is an assistant Professor at the Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand. He has been working as a lecturer for more than 10 years in the field of information technology, information science, and business computing at Chaiyaphum Rajabhat University and Khon Kaen University, Thailand. e-mail: wirach@kku.ac.th.

Kulthida Tuamsuk is an associate Professor at the Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand. She has been working as a lecturer for more than 45 years in the field of library and information science at Khon Kaen University, Thailand. e-mail: kultua@kku.ac.th.

Juthatip Chaikhambung is an assistant Professor at the Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand. She has been working as a lecturer for more than 10 years in the field of information technology and information science at Khon Kaen University, Thailand. (e-mail: cjutha@kku.ac.th).

Shigeo Sugimoto is a Professor in the Department of Library, Information and Media Science, University of Tsukuba, Japan. His research interests include Library and information science/ Humanistic social informatics. He is serving as an editorial member and reviewer of several international reputed journals. He is a member of many international affiliations. He has successfully completed his Administrative responsibilities. He has authored many research articles/books related to Library and information science/ Humanistic social informatics. "

Article submitted 2020-01-22. Resubmitted 2020-02-22. Final acceptance 2020-02-22. Final version published as submitted by the authors.