Modeling linked open data for decision support in Thailand tourism

Kittima Mekhabunchakij
Suan Sunandha Rajabhat University, Thailand

Keywords
Linked Open Data, Linked Data Modeling, tourism ontology, tourism indicators

Abstract

Linked Data has become a standard for achieving semantic web. This paper presents an approach to Linked Open Data (LOD) modeling in development of Web of Data based on LOD standards and methodologies to manage the complexity of the domain of interest. An example from the domain Tourism exemplifies the approach applicable for LOD development and focuses on the design methods and techniques: designing relational data model from various standards and stakeholders in tourism industry; transformations between relational and RDF/OWL data models; and tourism ontology development mainly using Protégé. The approach uses the process of mapping relational data to Linked Data and key processes of LOD modeling and prototyping of Web of Data. The approach proposed is implemented as data visualization of key tourism indicators to illustrate the key processes of LOD modeling, and prototyping of Web of Data for the industry.

1. Introduction

Open Data is the concept that government and business can use a lot of information to manage organization, and has several advantages, including measurement of the impact of policy and a new knowledge from big data from various sources and formats (OKFN, 2015). Based on open Web standards, Linked Data enables data consumers to use generic tools to access, mash up, and visualize data, and its potential for easing access to public sector data is widely conceived (Bizer, 2009; Heath & Bizer, 2011). The concept of Linked Open Data (LOD) aims to provide information sources that can be easily combined together using standard tools (Bauer & Kaltenböck, 2011).

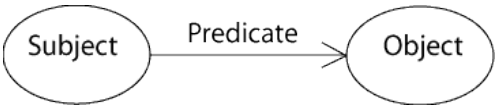
The Importance of tourism indicators. While as Tourism indicators are essential in supporting decision making processes, tourism domain is a highly complex and dynamic domain where decision-makers rely on decision support models to predict the future demand, or to analyze the relevant stakeholders. The data analytics about visitor arrivals and the destinations are important for the tourism industry for various tasks in decision making (Dwyer & Kim, 2003), such as understanding the contribution of tourism to the destination's economy, promoting the country (destination) as a place to live, invest in, do business with, etc., and marketing a destination by forecasting tourism demand and exploring potential source markets (Song & Li, 2008). Moreover, public policy makers and tourism planners can use the industry statistics to decide on planning tourism related infrastructure, such as accommodation facilities, food services, airports, and water treatment facilities (Dwyer & Kim, 2003; Sabou, Arsal, & Braşoveanu, 2013).

This paper presents an approach to LOD modeling in order to develop Web of Data to manage the complexity of the domain of interest. An example from the domain Tourism exemplifies the approach applicable for LOD development, focusing on the design methods and techniques: designing relational data model from various standards and stakeholders in tourism industry; transformations between relational and RDF/OWL data models; and tourism ontology development mainly using Protégé. Our implementation of TTLDV (Thailand Tourism Linked Data Visualization) is demonstrated and discussed.

2. Related Works

Linked Data

Linked Data, as a standard for achieving semantic web, describes a series of methods of publishing structured data using semantic web technologies (Manola, Miller, & McBride., 2004). Generally, Linked Data outlines the four principles: 1) URIs of identifying things (concepts), 2) HTTP URIs of addressing to those things, 3) provision of useful information using open standards (e.g. RDF), and 4) inclusion of links to other URIs (discovery of more things or concepts) (Berners-Lee, 2006). Publishing linked data enables machines to automatically discover more data from the data they already know.



Subject:	Measurement
Predicate (property):	*hasDestination*
Object:	Chiang Mai Province

Fig. 1. An example of RDF Graph

Resource Description Framework (RDF). RDF is a standard structure capable in describing data on the Web. The RDF data model is described in detail as part of the W3C RDF Primer (Manola, Miller, & McBride, 2004). RDF data structure called *Triple* consists of three parts: a resource (*Subject*), resource properties (*Predicate*) and the value of the property (*Object*) on the set of the *Triple* called RDF Graph. Fig. 1 illustrates an example triple of Measurement – *hasDestination* – Chiang Mai Province.

RDF has a great features that facilitate data integration. Although the format of the data is different. Making data collection easy and support the evolution of the pattern of data without changing all data into the same format (Berners-Lee, 2006). The RDF data model provides an abstract, conceptual framework for defining and using metadata. Its graph-based data model can ease the processing by applications.

RDF Schema (RDFS). RDFS is a semantic extension of RDF. RDF is domain-independent. The implementers can define their own terminology in a schema language, that is, RDF Schema (Brickley, Guha, & McBride, 2004). RDFS provides more flexible means to represent more complex relations of data concepts.

Linked Data for Tourism

The tourism domain is a highly complex and dynamic domain where decision-makers often rely on forecasting models to predict future demand or on decision support systems to analyze and compare the relevant stakeholders. The benefits of Linked Data technologies for the tourism domain refers to syntactic interoperability, as linked data technologies could enable a common technical infrastructure for sharing tourism data, and consequently facilitate the automatic and dynamic consumption of these data sets in decision supporting applications.

To solve the issues of data integration from heterogeneous sources, Sabou et.al. (2013) created TourMISLOD, a statistical data exploration applications with tourism indicators. Similarly, Vohnout, et.al. (2014) presents an approach to integrate several types of data sources in tourism (e.g. users' data, crowd-sourced data, and volunteered geo-information), enabling the reuse and sharing of existing tourism resources, channels and tools for tourism business support. Becker & Bizer (2009) presents DBpedia Mobile, a location-aware LOD mobile application, to support a tourist to explore a city through RDF links. The published content is interlinked with a nearby DBpedia resource.

Several other applications in tourism industry, such as cultural heritage data service (Mäkelä, et.al, 2012), and museum recommender (Ruotsalo, et.al. 2013), are emerging with Linked Data approaches.

3. Implementation of TTLDV

System Architecture

The system architecture of TTLDV (Fig. 2) has been designed based on the principles of Linked Data, with the Semantic Web layers (Koivunen, & Miller, 2001).

RDF Layer: Mapping Thailand Tourism Linked data

Ontology Layer. In design of an ontology for Thailand tourism indicators, the concepts in the TourMISLOD ontology are used, and then added with a number of indicator concepts particularly required by Thailand tourism industry.

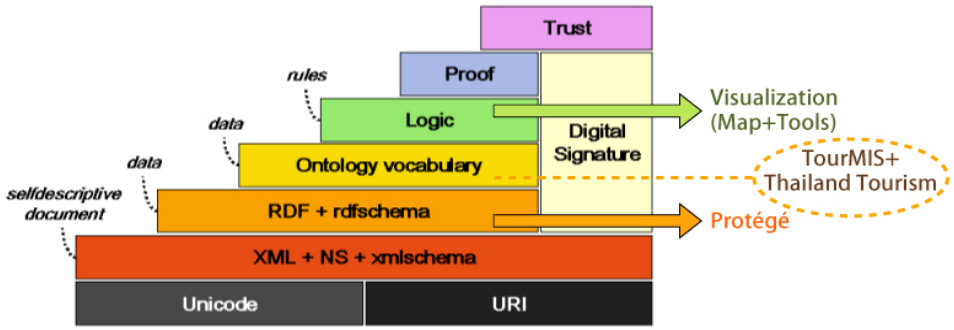


Fig. 2. TTDV System Architecture based on the Sematic Web layers (cf. Koivunen & Miller, 2001)

In our ontology development Protégé is used to model the domain of statistics and survey data, including 34 key indicators currently used by Tourism Authority of Thailand (TAT), and 17 tourism related economic and demographic indicators by the National Statistics Office of Thailand (NSO). Fig. 3 shows a concept hierarchy of tourism ontology. To cope with diverse concepts (entities, properties, and relationships) we introduce seven *Generic Tasks* to highlight the role of using LOD in our experimental platform.

The Logic Layer. This layer is about how to implement visualization of the defined tourism RDF data from the RDF layer. Apache HTTP server is used to host our PHP programs and Python library, and a RDF/JSON/JSON-LD repository.

Data Sources

This work is the first attempt in Thailand in bringing all statistical tourism data from a large number of public data sources (Mekhanunchakij, 2016), such as NSO, OGD (Open Government Data of Thailand), and TAT, into a LOD scheme for industry decision-making.

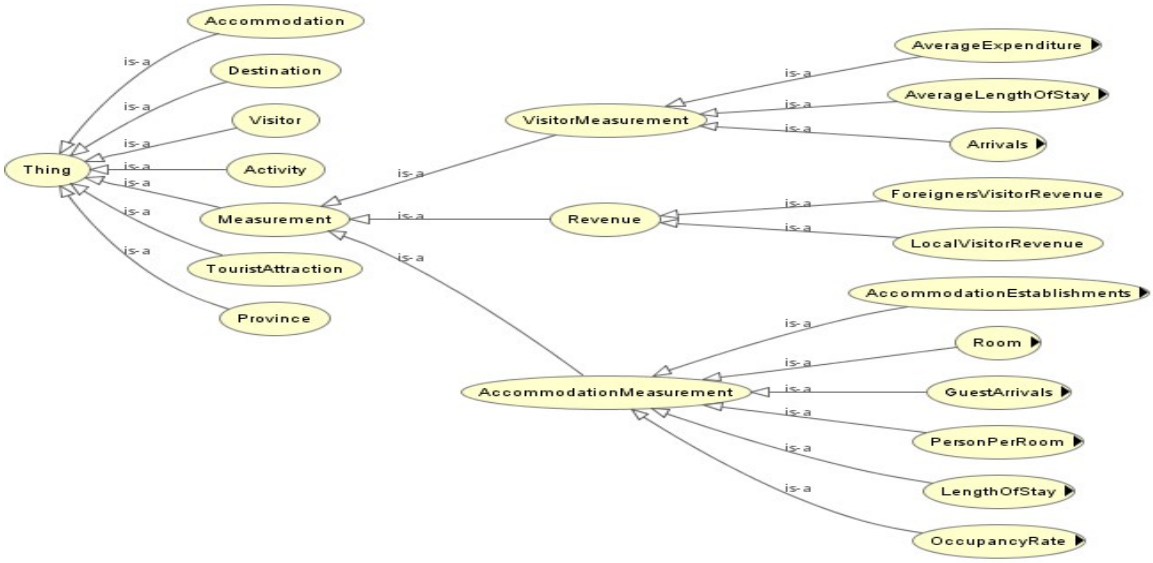


Fig. 3. A Concept Hierarchy of TTLDV Ontology

Relational Data to Linked Data

The process of mapping relational data to Linked Data is as follows:

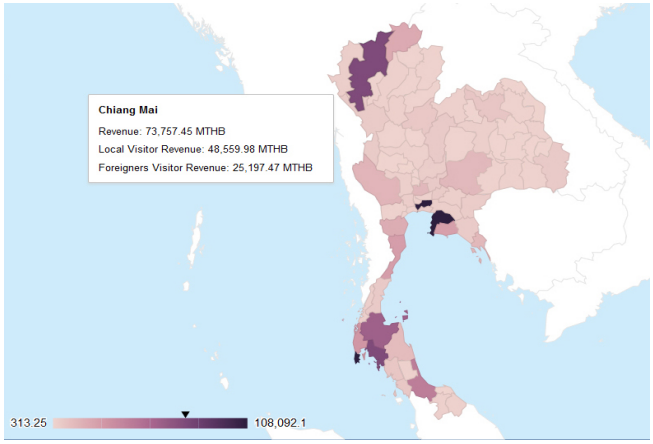
1. *Mapping the relational schema to an ontology:* (1) identifying "things" (or object types): for example, each relational table is mapped to an object type (e.g., "visitor", "destination", "measurement"), (2) identifying predicates (e.g., "v:visitorArrivals", "v:modeOfTransport").
2. *Creating RDF using the mapping:* Creating instances, assigning unique URIs. For example, each record in table "Visitor Arrivals" is mapped to an instance of type, "v: visitorArrivals", assigned with a custom URI ending with the destination code and the time period of the visitor Arrivals (primary key of the table).
3. *Managing the output RDF data:* Providing efficient translation process and JSON-LD / SPARQL query processing capability.

As for an example of the *Generic Task* no. 4, Fig. 5 illustrates three exemplified TTLDV Views using the resulting LOD model in linked-data integration, and JSON-LD-based RDF serialization techniques. Fig. 5(a) shows the revenue by location and a selected time period in geo-visualization with colors representing the revenue. Fig. 5(b) shows visitor statistics in peak periods and Fig. 5(c) a statistics of visitors' average expenditures per day.

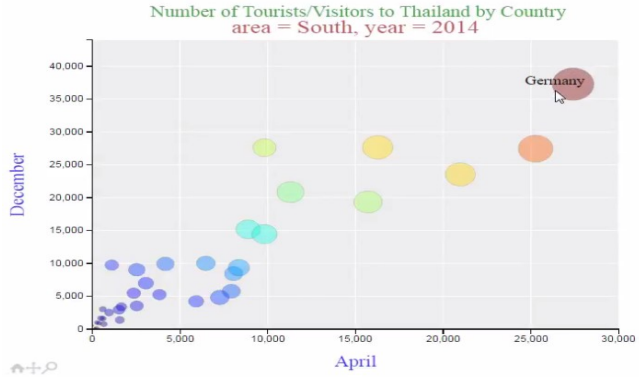
Generic Tasks. This research has tested the application by following key tasks.

1. Visitor arrivals, by location (Province) and time period
2. International visitor arrivals by country of origin, location and time period
3. Capacity of accommodation, by location and period
4. The revenues of inbound tourism, and the expenditures of outbound travelers, by location and period
5. The average revenues per visitor, by location and period
6. All of above, selection for Top 10, and Top 20 datasets
7. All of above, for year-to-year comparisons

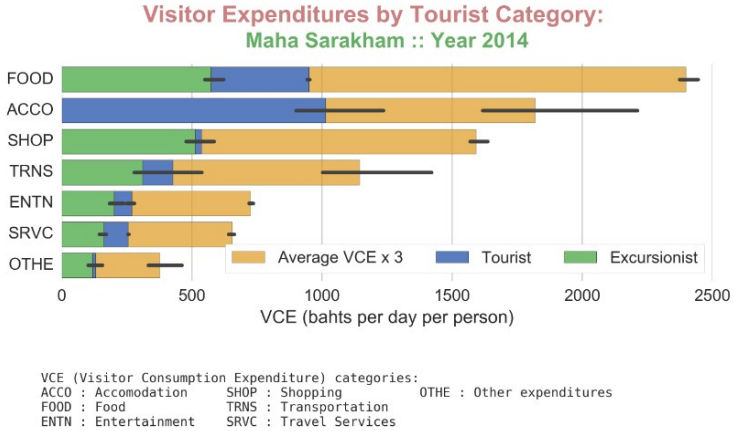
The Implementation. Charts for basic statistics are programmed in PHP using Google Charts API. For multivariate statistics that requires selecting, filtering, and drilling on data in visualization, the implementation is developed in Python using PyLD and mpdl3 libraries.



(a) Interactive map of tourism revenues by Province



(b) Visitor statistics in peak periods (December and April in 2014)



(c) Visitors' average expenditures per day (selected area, South, in 2014)

Fig. 4. Example uses of the resulting LOD for tourism data analysis

4. Discussions and Conclusions

This paper presents an approach to modeling LOD for the industry-specific decision support. The TTLDV implementation illustrates the key processes of LOD modeling. Because the data model and JSON-LD-based RDF serialization was undergoing itself, it needs to be redesigned for JSON-LD. Our experimental tool can be a starting point of further development of LOD-based methods and help foster the future growth of LOD and collaborations in the industry.

5. Future Work

Future work will focus on exploiting the approach of modeling additional dimensions of data, and an extension of the current ontology to cover new types of tourism concepts. For LOD visualization, further experimental development will extend the current tool to include more indicators and visual metaphors, and develop a semantic-web repository and services to handle more spatial and temporal dimensions with the SDMX standard.

Acknowledgements

This research is partially supported by the Institute of Research and Development, Suan Sunandha Rajabhat University. The author thanks Assoc. Prof. Dr. Veera Boonjing for his helpful comments and inspiring discussions, and TAT and OGA for their Open Data.

References

Becker, C. and Bizer, C., 2009. Exploring the geospatial semantic web with dbpedia mobile. *Web Semantics: Science, Services and Agents on the World Wide Web*, *7*(4), pp.278-286.

Berners-Lee, T., 2006. *Linked data-design issues*. [ONLINE] Available at: http://www.w3.org/DesignIssues/LinkedData.html [Accessed: 28 Jun 2016].

Bizer, C., 2009. The emerging web of linked data. *IEEE intelligent systems*, *24*(5).

Brickley, D. , Guha, R. V. , and McBride, B, 2004. RDF vocabulary description language 1.0: RDF Schema. *W3C recommendation*.

Dwyer, L. and Kim, C., 2003. Destination competitiveness: determinants and indicators. *Current issues in tourism*, *6*(5), pp.369-414.

Heath, T. and Bizer, C., 2011. Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, *1*(1), pp.1-136.

Koivunen, M. R. and Miller, E., 2001. W3C semantic web activity. *Semantic Web Kick-Off in Finland*, pp. 27-44.

Mäkelä, E., Hyvönen, E. and Ruotsalo, T., 2012. How to deal with massively heterogeneous cultural heritage data–lessons learned in CultureSampo. *Semantic Web*, *3*(1), pp.85-109.

Manola, F., Miller, E. and McBride, B., 2004. RDF primer. *W3C recommendation*, *10*(1-107), p.6.

Mekhabunchakij, K., 2016. Towards Modeling Linked Open Data for Decision Support: An example application of Thailand Tourism Linked Data Visualization, In *the 2016 Management and Innovation Technology International Conference*, pp. 88-92.

OKFN (Open Knowledge Foundation), 2015. *Open Data Handbook (Thai edition)*, Electronic Government Agency (Public Organization). [ONLINE] Available at: https://www.ega.or.th/th/content/890/877/. [Accessed: 29-Mar-2016].

Ruotsalo, T., Haav, K., Stoyanov, A., Roche, S., Fani, E., Deliai, R., Mäkelä, E., Kauppinen, T. and Hyvönen, E., 2013. SMARTMUSEUM: A mobile recommender system for the Web of Data. *Web semantics: Science, services and agents on the world wide web*, *20*, pp.50-67. *Systems* (pp. 201-204). ACM.

Sabou, M., Arsal, I. and Braşoveanu, A.M., 2013. Tourmislod: A tourism linked data set. Semantic Web, 4(3), pp.271-276.

Song, H. and Li, G., 2008. Tourism demand modelling and forecasting—A review of recent research. *Tourism management*, *29*(2), pp. 203-220.

Vohnout, P., Cerba, O., Kafka, S., Fryml, J., Krivanek, Z. and Holy, S., 2014, May. SmartTouristData approach for connecting local and global tourist information systems. In *IST-Africa Conference Proceedings, 2014* (pp. 1-6). IEEE.

Tongdhamachart, N., 2017. Challenges of Thailand Digital Economy Towards Community Tourism Development: Case Study of Baan Yafu, Maeyao District, Chiangrai Province. In *Actual Economy: Social Challenges and Financial Issues in XXI century (ACE-FIIS:2017)*. ACE.