

Construction and Publication of Maritime Linked Data Based on Lifecycle Model

Wei Guan^{1,a*} and Yiduo Liang^{2,b}

¹Research and Training Center, Dalian Neusoft University of Information, China

²Department of Information Management, Dalian Neusoft University of Information, China

^aguanwei@neusoft.edu.cn, ^bliangyiduo@neusoft.edu.cn

* The corresponding author

Keywords: Linked data lifecycle model; Domain modelling; Maritime linked data; Knowledge discovery

Abstract. The maritime data as one of the most important government data asset has increasingly attracted much of stakeholders' attention, with transparency and data reuse as two of the major aims. The term Linked Data refers to a set of best practices for publishing and interlinking structure data on the web. Using the methods and technologies within semantic web community, a number of maritime data has been putted on the web as linked data. However, the construction of existing maritime dataset is done without considering the concept of lifecycle, which naturally lies in the linked data itself. The paper makes attempts to propose an Ontology-driven Linked Data Lifecycle Model(OdLDLM) to assist publishing the maritime data on the web as linked data. We first conduct a systematic analysis of the existing lifecycle model, basing on which the modified OdLDLM is given, along with the description of each stage. Especially, we introduce domain model as the first step of our model and the key role of ontology in publishing the linked data is discussed. Taking the voyage data of maritime domain as example, we showcase in detail the process of how to use the proposed model to construct the maritime linked data. Also, a specific knowledge discovery case building on such data is given to demonstrate the additional value of linked data. The result shows that the OdLDLM is valid for publication of maritime linked data and convenient for later expansion and maintenance.

Introduction

In recent years, maritime open data as one of the most important government open data has sprung up around the world. However, much of the data published by people is unstructured textual information, and such information is always searched by using keyword-based search engines[1]. In July 2006, Sir Tim Berners-Lee first proposed the concept of Linked Data[2]. Linked data is a set of best practices for publishing and interlinking structured data on the web[3]. As a lightweight implementation of Semantic Web, linked data has attracted many scholars to devote to the related research field[4]. Using Linked Data practices and principles we can integrate generic data with smaller datasets[5]. Through the use of schema links, a level of semantic integration can be achieved. Meanwhile, by linking datasets, relations established between these datasets enrich highly the information contained in them. Thus, through predicate links, users can explore URI entities and data will get new meanings. In June 2009, Tim Berners Lee published an article, entitled "putting the government data online", and encourages governments to adopt Linked Data standards and release their government data on the web"[6]. Thus, the best practices have been adopted by a large number of data providers during the past years to publish the maritime data.

As a sea-faring nation, a large portion of Dutch history has been documented for taxing, toll, trade contracts and more. To further the digital history agenda, through linked data technology, De Boer et al. brings different individual datasets related shipping and recruitment together and present the Dutch Ships and Sailors(DSS) data cloud[7]. Basing on the DSS dataset, Balado, A. B. et al. try to discovery links between ships from Dutch maritime dataset and newspaper articles from historical archives[8]. As events have become central elements in the data representation from domains such as history, cultural heritage, multimedia and geography, Van Hage et al. proposes the Simple Event

Model(SEM) to model events in various domain[9]. Using the SEM, a new data set, Linked Open Piracy(LOP), is constructed on the Web of Data, which exposes descriptions of piracy attacks at sea. Thus, LOP can be seen as an Open Government Data initiative for intergovernmental data[10]. And in[11], Van Hage et al. continue to apply the SEM in a maritime safety and security use case for Situational Awareness.

As for the linked data itself is still changing all the time and has its own life-cycle. That is to say, the linked data has several different stages in a sequence. However, the datasets mentioned above do not take the concept of life-cycle into consideration. With the dataset becomes more and more large, especially the ontology may evolve into new formats in the future, the maintenance of the linked data will be increasingly difficult. Also, the inter-links between different datasets may become invalid and loses its additional value. Therefore, the concept of linked data life-cycle is proposed and up to now, a number of open data life-cycles exists[12]. Nevertheless, none of them are tailored to the specific needs of the maritime data. Moreover, some vital steps such as ontology modeling are ignored. Therefore, using the existing data life-cycle as a basis, we try to cover all necessary steps in the life-cycle of maritime data, in order to provide an available process which maritime data stakeholders can follow.

Ontology-driven Linked Data Lifecycle Model

By referring to the above Linked Data Life-cycle Model and investing the requirement of persons in maritime domain, we omit the redundant steps and propose the Ontology-driven Linked Data Lifecycle Model(OdLDLM), enhancing the role of ontology in linked data building process.

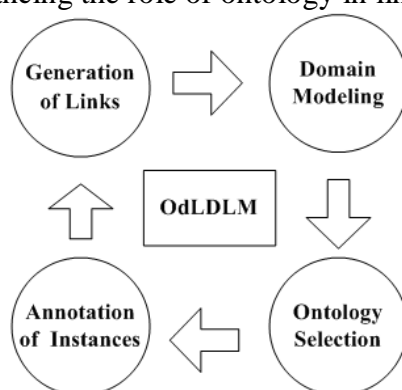


Figure 1. Ontology-driven Linked Data Lifecycle Model

The proposed life-cycle, shown in fig. 1, is made up of four sections, namely Domain Modelling, Ontology Selection, Annotation of Instances and Generation of Links. Here are explanations of each stage in detail:

Domain Modeling. The OdLDLM typically starts with the domain modeling. This is the process involving collecting and selecting the data. We use model tools to describe entities and its relationship in domain. This requires removing any private data or personal data, as well as identifying under which conditions will this data be published. Results of domain modeling can be mapped to ontology vocabulary later.

For maritime domain, voyage is the most common information to be recorded. A record of voyage usually consists of typical concepts such as Ship, Port, ShipCompany, Shipper and sailor etc.. With E-R model, we use a rectangle to represent a entity. A diamond is used to descript relationships between entities. With the aid of maritime domain expert, we could obtain the following E-R diagram.

Ontology Selection. Select appropriate vocabulary that meets your needs. If you want your data to be easily interlinked with other datasets, you need to reuse common vocabularies as much as possible. Reusing existing ontologies ensures that most people will use the most common vocabularies in the most common use cases. For maritime domain, there are three core RDF vocabularies below and will

be reused to model the most common types of linked data. DBpedia, GeoNames ontology and DSS(Dutch Ships and Sailors).

The goal of linked data is to allow anyone to describe anything. If we couldn't find an existing vocabulary term to describe the resource, we'd need to create a new vocabulary and resolve it on the web. We abstract the common concepts and define domain classes, including: Ship, ShipCompany, voyage, location, ShipType and Person. Besides, other subclasses are defined with the `rdfs:subclass-`

Of property. For example, a class 'province' is defined as a subclass of class 'location'. In order to show the relationships between concepts, the datatype property and object property are used. Datatype property is used to describe a feature of class, whose values are text or numbers. Attribute value of the object property is an individual object, which is used to describe the relationship between instances of the class.

Annotation of Instances. This is the actual act of opening up the data by annotating the instances with the selected ontology terms and publishing it on an appropriate RDF triples server such as Cliopatria or Fuseki.

Prepare the data by transforming data from its existing non-RDF format into RDF formats. Here, we chose the RDF/XML as the serialization formats and the dataset in an RDF database, which facilitates SPARQL access to your data.

URI is used to represent the resource. Any resource that is considered to be a meaningful 'thing' is assigned a URI. The URIs facilitate two aspects of linked data. First, they globally act as unique names for the things you want to represent, and second, they provide an easy way to access description online by using HTTP URIs. Here, we define our URIs within a publicly accessible HTTP namespace under our control and also use URIs that can be referenced.

There are many approaches to minting URIs and Many of our URIs fall into the following pattern: `http://{authority}/{container}/{item_key}`.

Authority. Generally, the authority part is a DNS machine name, optionally with a port.

Container. This part allows you to keep keys separate to each context.

Item. This part uniquely identifies a specific resource in that container.

We need a URI where we will publish our new vocabulary and call it whatever we like. For the purposes of show only, we use `http://www.website.maritime.mt/`, with the namespace prefix 'mt'.

Here, we publish the data mentioned above as linked data with the Cliopatria semantic server. Cliopatria is an RDF triple database that provides feedback on the produced RDF through a web interface.

Generation of Links. Data interlinking is the last step of linked data life-cycle. This allows the published data to have additional value, for the linking of data gives context to its interpretation. Technical method such as ontology alignment can be used to build links between different datasets with heterogeneous formats.

In the establishment of maritime ontology model, if the new class and existing class in other class are semantically equivalent, you can build an `owl:equivalentClass` binary relations on the two classes.

Place of establishment in the DBpedia Dataset describes a wide range of classes, which can be described as a super class of the 'location' class.

Using the object property, we could build the links between different datasets. For example, the range of `mt: PortOfUnloading` is `&dss;Port`, which makes the semantic link relationship between the two datasets.

Add the `owl: sameAs` or `rdfs: seeAlso` property to existing resource with its value referring to the same subject. DBpedia dataset is associated with a number of datasets to form a plurality of access links. Because of its extensive coverage, many datasets are also associated data through DBpedia datasets. The corresponding description of Dalian Port could be found in DBpedia dataset. By using the `owl: same As` property, Dalian Port can be linked with other resources.

Knowledge Discovery based on Maritime Linked Dataset

The maximum value of the linked data is to achieve knowledge discovery between different sets of data by establishing a link between the different resources. Maritime database stores a lot of voyage

information, involving many semantically related objects such as Port, Ship and ShipCompany etc.. After publishing maritime information as linked data, we could make knowledge discovery basing on such semantically related relationship.

A SPARQL 1.1 compliant endpoint is provided by the ClioPatria. By editing SPARQL queries, the data can amplify each other and new research possibilities open up. Here, with an appropriate SPARQL query, we can get a record of voyage resource stored in the ClioPatria as followed in fig. 2.

Local view for "http://www.website.maritime.mt/Voyage_014W"

Predicate	Value (sorted: default)
rdf:type	owl:NamedIndividual
mt:Voyage	mt:Voyage
mt:EstimatedArrivalDate	"2015-11-13"
mt:EstimatedDepartureDate	"2015-11-10"
mt:EstimatedSailingDay	"3"
mt:hasShip	mt:Ship_COSCO_BELGIUM
mt:placeOfArrival	mt:Place_Shanghai
mt:placeOfDeparture	mt:Place_Dalian
mt:PortOfLoading	mt:Port_Dalian
mt:PortOfUnloading	mt:Port_Shanghai

Figure 2. Voyage resource in ClioPatria

The voyage above has a ship named 'COSCO_BELGIUM'. By clicking the value link, the page redirects to a new page containing the ship information in fig. 3.

Local view for "http://www.website.maritime.mt/Ship_COSCO_BELGIUM"

Predicate	Value (sorted: default)
rdf:type	owl:NamedIndividual
mt:Ship	mt:Ship
mt:RegisteredPort	mt:Port_Hongkong
mt:ShipGrossTon	"153666"@en
mt:ShipLoading	"13145"@en
mt:ShipName	"COSCO BELGIUM"@en
mt:ShipNetTon	"65592"@en
mt:ShipYear	"2013"@en
mt:SubordinateCompany	mt:Company_COSCO

Figure 3. Ship resource in ClioPatria

Continue to click the 'SubordinateCompany' link in the page, the page will automatically jump to the company information page which the ship belongs to, showing in fig. 4.

Local view for "http://www.website.maritime.mt/Company_COSCO"

Predicate	Value (sorted: default)
rdf:type	owl:NamedIndividual
mt:ShipCompany	mt:ShipCompany
mt:CompanyAddress	"Hongkou District No. 678 East Daming Road, Shanghai"@en
mt:CompanyName	"China Ocean Shipping(Group) Company"@en
mt:CompanyTel	"021-65966666"@en
mt:CompanyWebsite	"http://ebusiness.coscon.com/wps/portal"@en
mt:hasCountry	mt:Place_China

Figure 4. Company resource in ClioPatria

In DSS dataset, a resource of China has been created and has a unique URI to be accessed to. By establishing a link between own maritime dataset and DSS with the `rdfs:sameAs` property, resource access across different sets of data can be achieved. Fig. 5 shows the description of China in DSS dataset.

Local view for "http://purl.org/collections/nl/dss/das/place-China"

Predicate	Value (sorted: default)
skos:preferred label	"China" ¹
rdf:type	dss:Place ¹
skos:inScheme	dss:DAS Place ConceptScheme ²

Figure 5. Place of China in DSS dataset

Conclusion

In this paper, an Ontology-driven Linked Data Lifecycle Model(OdLDLM) is proposed and used to publish the maritime linked data on the web. We showcase each step of the model, in which ontology plays an important role, making the expansion and maintenance of linked data more easier. Especially, we initiatively introduce ontology modelling into our model, which facilitates better understandings of domain knowledge. The value of linked data lies in additional knowledge discovery through an accessible http URI as the object in triples. With the new maritime linked dataset stored in Cliopatria, we accomplish semantic surfing on the web. The work here is significant to the broader Linked Data community since it presents a prime example of how collaboration between linked data lifecycle model and domain knowledge can lead to high-quality accessible digital maritime dataset. At the same time, additional enrichments between different datasets are possible at a later stage. We believe that the methodology can be reused in other multi-part datasets in future.

References

- [1] E. Kalampokis: *Proc. Panhellenic Conference on Informatics Chinese Navigation*(Thessaloniki, Greece, September 19-21, 2013). p.25.
- [2] Information on <http://www.w3.org/DesignIssues/LinkedData.html>
- [3] C. Bizer, T. Heath and T. B. Lee: *International Journal on Semantic Web and Information Systems*, Vol. 5(2009) No.3, p.1.
- [4] C.J. Xia, W. Liu, L. Zhao: *Journal of Library Science in China*, Vol. 38(2012) No.1, p.45.(In Chinese)
- [5] C. Bizer: *IEEE Intelligent Systems*, Vol. 24(2009) No.5, p.87.
- [6] Information on <http://www.w3.org/DesignIssues/GovData.html>
- [7] V. D. Boer, M. V. Rossum and J. Leinenga: *Proc. 13th International Semantic Web Conference*(Riva del Garda, Italy, October 19-23, 2014). p.229.
- [8] A. B. Balado, V. D. Boer, and G. Schreiber: *Proc. the 6th International Conference on Social Informatics* (Barcelona, Spain, November 11-13, 2014). p.254.
- [9] V. Hage, W. Robert: *Journal on Data Semantics*, Vol. 1(2011) No.3, p.167.
- [10] Information on http://data-gov.tw.rpi.edu/wiki/Open_Government_Data
- [11] W. R. V. Hage, G. D. Vries and G. Schreiber: *Proc. ACM International Workshop on Events in Multimedia*(Beijing, China, October 23-23, 2009). p.73.
- [12] J. Attard, F. Orlandi and S. Scerri: *Government Information Quarterly*, Vol. 32(2015) No.4, p. 399.