Construction of Shipping Linked Data Lifecycle Model and Its Application in Semantic Navigation

Yiduo Liang*
School of Software, Dalian University of Foreign Languages, Dalian, China

*Corresponding author e-mail: liangyiduo@dlufl.edu.cn

Abstract. As one of the most important government data assets, shipping data has attracted more and more attention from stakeholders. The term linked data refers to a set of best practices for publishing and linking structured data on the web. The application of linked data technology to open the shipping data can promote the transparency and reusability of data. Many shipping data have been placed on the network as linked data. However, the existing construction of shipping linked data does not consider the concept of lifecycle, and the concept of lifecycle naturally lies in the link data itself. Therefore, this paper attempts to propose a linked data lifecycle model to assist in publishing shipping data as linked data on the Web. We first conduct a systematic analysis of the typical lifecycle model. On this basis, a new linked data lifecycle model is proposed, and the role of each step of the model is also explained. In particular, we take domain modeling as the first step of the model and discuss the key role of it in publishing linked data. Taking the voyage data as an example, we show in detail how to use the proposed model to construct shipping linked data. Then, a specific semantic navigation case based on the above data is given to prove the added value of the linked data. The results show that the lifecycle model proposed in this paper is suitable for the publication of shipping data and is convenient for future expansion and maintenance.

Keywords: Lifecycle Model, Shipping Linked Data, Semantic Navigation

1. Introduction
Internationally, government data opening has become a research hotspot, and shipping data opening is an important content of government data opening research. Shipping information is always searched using a keyword based search engine, and much of the data that people publish are unstructured text information [1]. Linked data was first proposed by Tim Berners-Lee in 2006 [2]. Linked data is a lightweight implementation form of the Semantic Web and has attracted the attention of many scholars [3]. Linked data refers to a set of best operations to open structured data on the web [4]. Using linked data practices and principles, we can integrate generic data with smaller data sets [5]. Tim berners-Lee published a paper titled "bringing government data online" in 2009, hoping that data owners will use the technology of linking data to publish the data online [6]. Therefore, owners of a large amount of shipping data have used linked data to open their shipping data in the past few years.
As a sea-faring nation, a large portion of Dutch history has been documented for taxing, toll, trade contracts and more. To further advance the digital history agenda through linked data technology, De B. [7] integrated various scattered data sets related to shipping and recruitment, and proposed the DSS Linked Data Set. Based on the DSS dataset, Balado A. B. et al. [8] attempted to find links between ships in the Dutch shipping data set and newspaper articles in historical archives. Van H. [9] proposed the SEM model to model events in various scenarios. Using SEM, a new data set named "Linked Open Piracy (LOP)" is constructed, which publishes the description of piracy attacks at sea. Van H. [10] continue to apply SEM to maritime safety and security use case for situational awareness.

As datasets become larger and larger, especially when ontologies may evolve into new formats in the future, the maintenance of linked data will become more and more difficult. In addition, links between different datasets may become invalid and lose their added value. For linked data, it is still changing itself and has its own lifecycle. That is, linked data has several different stages in a sequence. Therefore, the concept of linked data lifecycle [11] is proposed, and so far, there are many kinds of lifecycle. However, they are not suitable for the specific needs of shipping data. In addition, some important steps, such as domain modeling, are ignored.

In the paper, we first briefly introduce the different stages of linked data lifecycle, and propose the shipping linked data lifecycle model in Section 2, trying to cover all the necessary steps in the shipping data lifecycle to provide available processes that can be followed by shipping data stakeholders. In particular, domain modeling is introduced as the first and primary step. Taking the shipping domain as an example, in Section 3, we describe in detail how to use the proposed model to build shipping linked data. In the fourth part, we show the specific application of semantic navigation based on the newly constructed data set, and show the effectiveness of the proposed lifecycle model. This paper summarizes the advantages of the model and the future work. The paper is concluded by the advantage of model and the next work in future.

2. Linked Data Lifecycle Model

2.1. Introduction of Typical Lifecycle Model

In Part 2, we give a brief introduction to the various stages of a typical linked data lifecycle model [12]. Instructions for each of the stages are as following.

- Extraction. Information represented in unstructured form should be converted to RDF graph model.
- Storage and Querying. When there is a large amount of RDF data, appropriate methods should be adopted to quickly and accurately save, index, and retrieve data elements.
- Authoring. Users should be given certain rights to create new structured data and correct incorrect data.
- Linking. Links of the same entities between the different dataset have to be established.
- Enrichment. Use higher-level structured forms to organize data and the data can be aggregated and queried more efficiently.
- Quality Analysis. Reasonable evaluation indicators should be designed to comprehensively evaluate the quality of online linked data sets.
- Evolution and Repair. Strategies for repairing these problems are required.
- Search, Browsing and Exploration. A convenient and intuitive way should be provided for users to browse, explore and search the structure information on the web.

2.2. Shipping Linked Data Lifecycle Model

After understanding the actual needs of shipping business personnel and referring to the above linked data lifecycle model, we propose a shipping linked data lifecycle model. The model omits redundant steps and emphasizes the important role of domain modeling in linked data’s construction.

The proposed lifecycle is made up of four sections, namely Domain Modelling, Ontology Creation, Data Transformation and Construction of Links. Here are explanations of each stage in detail.
• Domain Modeling. This model usually starts with domain modeling. This is a process that involves the collection and selection of data. Modeling tools can be used to show various entities and its relationships between entities in the domain. The results of domain modeling can be mapped to the ontology vocabulary later.

• Ontology Creation. Select the appropriate domain ontology vocabulary. If we want linked data to be easily linked to other data sets, common terms are needed to be reused as much as possible. If there is no a vocabulary that meets our needs, a new ontology is needed to be customized.

• Data Transformation. This is the process of transforming unstructured data into structured data by using selected ontology terms. We need to select an appropriate RDF triplet server to publish instance data and provide a SPARQL access port.

• Construction of Links. This is the last step in the lifecycle of linked data. We can use the method of ontology fusion to establish links between heterogeneous data sets in various formats. This gives the published data additional value because the links to the data provide the context for its interpretation.

3. Construction of Shipping Linked Data Set

3.1. Domain Modeling

In relational databases, we can define special terms to establish and describe various relationships between data elements, called database schemas. E-R diagram is a graph model for entities and relationships, such as 1:1, 1:n and m:n. In domain modeling, we could use E-R diagram to describe various entity concepts and the relationships between concepts in shipping domain.

For shipping 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.. A typical E-R diagram is given as follows in Fig. 1.

![E-R model of shipping company entity and ship entity](image.png)

**Figure 1.** E-R model of shipping company entity and ship entity

3.2. Ontology Creation

In essence, ontology is a set of domain-related vocabulary that provides schema definitions for describing structured data. People have developed a large number of ontology vocabularies for different application purposes, and they are widely distributed on the web. If people choose to use the terms that have never seen before, anyone can never reuse linked data. There are two ways to avoid this problem: reuse existing vocabularies as much as possible, and make sure that the URIs defining linked data vocabularies follow the linked data principles.

Reusing this vocabulary can easily realize knowledge sharing and semantic interoperability to the largest extent, reducing the workload of repeated definition of ontology. For shipping domain, three common ontology vocabularies can be used: DBpedia [13], GeoNames [14] and DSS [7]. If we cannot find an ontology vocabulary that meets our needs, then a new ontology vocabulary needs to be defined and at the same time it is necessary to ensure that the ontology can be parsed on the web. According to
E-R model, we could abstract the common concepts and define domain classes, including: Ship, ShipCompany, voyage, location, ShipType and Person. Besides, other subclasses are defined with the rdfs:subClassOf property. In order to show the relationships between concepts, property of datatype and property of object are used. Property of datatype is used to describe the 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. The typical properties in shipping domain are depicted in Table 1.

| Class             | Property                                                                 |
|-------------------|---------------------------------------------------------------------------|
| Ship              | ShipName, BuildingYear, ShipAge, GrossTon, NetTon, ShipScale, ShipLength,  |
|                   | ShipWidth, ShipFlag, Loading, RegisteredPort, SubordinateCompany, typeOfShip,
|                   | Shipper                                                                   |
| ShipCompany       | ShipCompanyName, Website, EnName, EstablishYear, CompanyTel,              |
|                   | CompanyAddress, SubordinateCountry, SubordinateHeadOffice                 |
| Voyage            | EstimatedDepartureDate, EstimatedArrivalDate, ArrivalTime, EstimatedSalingDay,|
|                   | hasShip, placeOfDespatch, PortOfLoading, PortOfUnloading, placeOfArrival   |
| Port              | PortName, PortTel, PortFax, Latitude, Longitude, CountryTimeZone,         |
|                   | PortTimeZone, PortCode, PortShortName, hasCountry, hasCity, TypeOfPort    |
| Person            | PersonName                                                                |
| Location          | LocationName                                                              |
| PortType          | PortTypeName                                                              |

### 3.3. Data Transformation

Here, we chose the RDF/XLM and Turtle as the serialization formats to convert data from various other formats to RDF format, which facilitates SPARQL access to the data. In this step, we visited the official website of COSCO Co. LTD and selected a piece of voyage information from the above website: Port of Dalian to Port of Shanghai. Information about port of Dalian and port of Shanghai could be found on the website of CNSS. From the e-Business website of COSCO Co. LTD, we get the ship information of COSCO BELGIUM.

URI is used to represent the resource. Any resource that is considered to be a meaningful thing is assigned a URI. For the purposes of show only, we use http://lop.dlmu.edu.cn/mtdata/, with the namespace prefix “mtdata”. An example of Dalian port resource URI is like “http://lop.dlmu.edu.cn/mtdata/Port_Dalian”. With the namespace prefix “mtdata”, the whole URI could be shortened by & mtdata/Port_Dalian. Here, we publish the above data as linked data and use the semantic server to store it. Publishing port instance “Port_Dalian” as linked data in Turtle format.

```turtle
mtdata:Port_Dalian rdf:type mtdata:Port ;
  mtdata:Latitude "38.55"^^xsd:double ;
  mtdata:PortTimeZone "GMT+8" ;
  mtdata:PortFax "86-411-82807148" ;
  mtdata:PortTel "86-411-82627147" ;
  mtdata:CountryTimeZone "8.0" ;
  mtdata:PortName "dalian" ;
  mtdata:PortShortName "DAL" ;
  mtdata:hasCountry mtdata:Place_China .
```

### 3.4. Construction of Links

If two classes are semantically equivalent, a owl: equivalentClass property can be used to describe this binary relations. An owl: equivalentClass example about ship types between two classes is showed below.

```xml
<owl:Class rdf:about="http://lop.dlmu.edu.cn/mtonto/typeOfShip"/>
```
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.

Property of object can be used to construct the semantic links between various datasets. For example, the range of mtdata:PortOfUnloading is &dss;Port, which makes the semantic link relationship between the two datasets.

We can use the owl: sameAs or rdfs:seeAlso attribute to establish an identity linkage between two resources with the same meaning but different names. The corresponding description of “Port of Dalian” could be found in DBpedia dataset. By employing the property of owl:sameAs, Port of Dalian can be linked with other resources.

The location where Dalian port located at is represented in anonymous resources, and could be linked to the place in Geonames Dataset.

The anonymous resource representing China could be linked to the DSS resources.

4. Semantic Navigation Case for Shipping Linked Data
The biggest advantage of linked data is to execute semantic navigation between various data sets by establishing all kind of the links between different entities. Shipping database stores a lot of voyage information, involving many semantic related objects, such as port, ship and ship company etc.. After obtaining the shipping linked data set, we can carry out the semantic navigation application based on semantic relationship of the shipping linked data. By using the SPARQL endpoint and editing the appropriate SPARQL query clause, a record of voyage entity stored in RDF server can be found shown in Fig. 2.
Figure 2. Structural information of voyage entity
As can be seen from the above figure, there is a ship entity named "COSCO_BELGIUM" in the voyage “Voyage_DalianToShanghai” instance. By clicking the semantic link of the ship entity, the page will be redirected to a new ship entity information page. See Fig. 3.

Local view for "http://lop.dlmu.edu.cn/mtdata/Voyage-DalianToShanghai"

| Predicate                  | Value (sorted: default)                                      |
|----------------------------|--------------------------------------------------------------|
| rdfs:label                 | "Voyage from Dalian to Shanghai"                            |
| mttonto:ship               | mttonto:Voyage                                              |
| mttonto:arrival            | mtt onto:Place-Shanghai                                      |
| mtt onto:departure          | mtt onto:Place-Dalian                                        |
| mttonto:portOfArrival      | mtt onto:Port-Shanghai                                       |
| mtt onto:portOfDeparture   | mtt onto:Port-Dalian                                         |
| mttonto:ship              | mtt onto:COSECO_BELGIUM                                      |
| mtt onto:personNumber      | "189"                                                        |
| mtt onto:timeOfArrivalAtDestination | "2019-03-14"                                |
| mtt onto:timeOfDeparture   | "2019-03-11"                                                |
| mtt onto:voyageCode        | "Voyage-DalianToShanghai"                                   |

Figure 3. Structural information of ship entity
Continue to click the company entity link named with “China Ocean Shipping Company”, the page will also directly jump to the company entity information page which the ship belongs to. See Fig. 4.

Local view for "http://lop.dlmu.edu.cn/mtdata/COSCO_BELGIUM"

| Predicate                  | Value (sorted: default)                                      |
|----------------------------|--------------------------------------------------------------|
| rdf:type                   | mtt onto:Ship                                               |
| mtt onto:shipOwner         | mtt onto:China_Ocean_Shipping(Group) Company                |
| mtt onto:shipType          | mtt onto:passenger_vessel                                   |
| mtt onto:registeredPort    | mtt onto:Port-Hongkong                                      |
| mtt onto:shipGross         | "153866"                                                    |
| mtt onto:shipID            | "COSCO_BELGIUM"                                             |
| mtt onto:shipNet           | "65592"                                                     |
| mtt onto:shipYearOfBuild   | "2013"                                                      |

Figure 4. Structural information of company entity
It can be seen that with the increasing number of semantic links, we can start from an instance to browse and invent more semantic related instance information. Such information is structured and easy to be understood and processed by the machine automatically. This is also the unique advantage for linked data which is distinguished with the traditional hyperlink technology.

5. Conclusion
Shipping data is one of the important contents of government data openness. Publishing shipping data in the form of linked data can realize automatic data processing and semantic browsing, and then exert
the potential value of data. In our paper, a new linked data lifecycle model is proposed and used to
construct and open the shipping linked data on the web. Using the shipping linked data stored in RDF
server, we can edit the appropriate SPARQL query clauses and complete semantic navigation based on
the semantic links between different objects. As we can see, with the continuous enrichment of
semantic links, a wider range of knowledge expansion and browsing can be realized across different
data sets.

Acknowledgments
This work was financially supported by the Doctoral Research Startup Fund Project of Liaoning
Province "Research on Implementation Strategies and Key Technologies of Data Opening in the
Shipping Field under the National Big Data Strategy"(Grant No. 2019-BS-060).

References
[1] Kalampokis, E., Tambouris, E., & Tarabanis, K. (2013). On publishing linked open government
data. In Proceedings of Panhellenic Conference on Informatics, P. 25-32, ACM.
[2] Linked Data. Available at http://www.w3.org/DesignIssues/LinkedData.html.
[3] Xia C.J., Liu W., Zhao L., et al.(2012). The Current Technologies and Tools for Linked Data:
A Case of Drupal. Journal of Library Science in China, 38(1), 49-57.
[4] Bizer C., Heath T., & Berners-Lee.(2009). Linked Data-the Story So Far. International Journal
on Semantic Web and Information Systems, 5(3), 1-22.
[5] APABizer, C. (2011). The Emerging Web of Linked Data. In Proceedings of the 2011
International Conference on Intelligent Semantic Web-Services and Applications, P. 87-92,
ACM.
[6] Putting government data online. Available at http://www.w3.org/DesignIssues/GovData.html.
[7] Boer, V. D., Rossum, M. V., Leinenga, J., & Hoekstra, R. (2014). Dutch Ships and Sailors
Linked Data. In Proceedings of the Semantic Web – ISWC 2014, P. 229-244, Springer
International Publishing.
[8] Balado, A. B., Boer, V. D., & Schreiber, G. (2014). Linking Historical Ship Records to a
Newspaper Archive. Social Informatics. Springer International Publishing.
[9] APAHage, W. R. V., Malaisé, V., Erp, M. V., & Schreiber, G. (2011). Linked open piracy.
Journal on Data Semantics, 1(3), 167-168.
[10] Van Hage, W. R., Malaisé, V., Vries, G. D., et al. (2009). Combining ship trajectories and
semantics with the simple event model (SEM). In Proceedings of the 1st ACM international
workshop on events in multimedia, P. 73-80, ACM.
[11] Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government
data initiatives. Government Information Quarterly, 32(4), 399-418.
[12] Auer, S., Lehmann, J., & Ngomo, A. C. N. (2011). Introduction to Linked Data and Its
Lifecycle on the Web. Reasoning Web. Semantic Technologies for the Web of Data.
Springer Berlin Heidelberg.
[13] Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., & Cyganiak, R., et al. (2009).
Dbpedia - a crystallization point for the web of data. Journal of Web Semantics, 7(3), 154-165.
[14] GeoNames Ontology. Available at http://www.geonames.org/ontology/documentation.html.
[15] COSCO Container Lines Co. LTD. Available at http://www.coscon.com/home.do? language=en.
[16] CNSS. Available at http://app.cnss.com.cn/portsch/search.php.
[17] e-Business of COSCO. Available at http://ebusiness.coscon.com/wps/portal/.