

Towards knowledge-based integration and visualization of geospatial data using Semantic Web technologies^{*}

Weiming Huang

GIS Centre, Lund University, Lund, Sweden

weiming.huang@nateko.lu.se

Abstract. Geospatial data have been pervasive and indispensable for various real-world application of e.g. urban planning, traffic analysis and emergency response. To this end, the data integration and knowledge transfer are two prominent issues for augmenting the use of geospatial data and knowledge. In order to address these issue, Semantic Web technologies have been considerably adopted in geospatial domain, and there are currently still some activates investigating the benefits brought up from the adoption of Semantic Web technologies. In this context, this paper showcases and discusses the knowledge-based geospatial data integration and visualization leveraging ontologies and rules. Specifically, we use the Linked Data paradigm for modelling geospatial data, and then create knowledge base of the visualization of such data in terms of scaling, data portrayal and geometry source. This approach would benefit the transfer, interpret and reuse the visualization knowledge for geospatial data. At the meantime, we also identified some challenges of modelling geospatial knowledge and outreaching such knowledge to other domains as future study.

Keywords: geospatial data, data integration, data visualization, Semantic Web, ontologies, rule-based inference.

1 Introduction

Geospatial information has received increasing attention from the mainstream IT world and become indispensable for various real-world applications of e.g. urban planning, traffic analysis and emergency response. In the geospatial community, the transfer, sharing and visualization of geospatial data mainly rely on a number of syntactic standards which shape the current solutions for spatial data infrastructure (SDI). Such standards are mainly from Open Geospatial Consortium (OGC), and most of them only guarantee interoperability on a syntactic level, whereas the semantics and knowledge are represented insufficiently. Therefore, we need a way for addressing the semantic challenges concerning geospatial data and knowledge [1]. Besides, the SDI - whose aim is

^{*} The PhD project is supervised by Prof. Lars Harrie and Dr. Ali Mansourian at GIS Centre, Lund University, and it is funded by China Scholarship Council and Lund University.

mainly for dissolving environmental and geospatial data held in silos – are still perceived as islands in mainstream IT [2], and this impedes the augmentation of the use of geospatial data to other domains. In this context, Semantic Web technologies, including the part concerning Linked Data, unveil a promising way for resolving these issues by embracing knowledge-based approaches which could foster better transfer, interpretation, expansion and reuse of geospatial data, information and knowledge.

Visualization is one of the most important and pervasive application areas of geospatial data and in geographic information systems (GIS); it allows users to explore, synthesize, present and analyze the underlying geospatial data in an interactive manner. However, the visualization of geospatial data unveils some long-standing challenges to both the providers and users. One such challenge is the data integration issue, which can be the integration between geospatial data and also between geospatial data and data from other domains. At the meantime, the visualization of geospatial data is also knowledge-intensive from a cartographic perspective for both the providers and users. For the providers, a wide range of cartographic theories is required to derive sensemaking and cartographically satisfactory applications; and for the users, the knowledge is required to interpret the presented data in a meaningful way. And sometimes the users need to reach a high level of cognitive consensus with the providers in order to better perceive the delivered information from the visualization applications.

Therefore, this PhD thesis mainly investigates how the Semantic Web technologies can foster better integration and visualization of geospatial data, and thereof aid the outreaching of geospatial data, information and knowledge to other domains. The scope of the PhD thesis is broad, and benefits brought up by Semantic Web technologies will be demonstrated in a few particular cases where some traditional geospatial problems are better solved with the Semantic Web. And in this framework, ontologies and rules are intensively used as two main paradigms for knowledge representation.

2 State-of-the-Art

The application of Semantic Web technologies has developed considerably in geospatial domain in the last decade as they address several long-standing challenges of e.g. data integration, semantic interoperability and knowledge formalization and provide a promising way to connect spatial data infrastructures (SDIs) with the mainstream to augment the application of geospatial data [2]. As a result, a vast number of geospatial datasets have been released as Linked Data, and some of them are serving as central hubs in the Linked Open Data (LOD) cloud¹. At the meantime, a number of vocabularies for representing geospatial data (see [3] for a comparison), the geospatial Linked Data query language GeoSPARQL [4] and several geospatial enabled RDF stores (e.g. Stardog² and Virtuoso³) have been developed. These theoretical and technical advancements have facilitated the publishing of geospatial Linked Data and the use of Semantic Web for geospatial knowledge representation and sharing.

¹ <https://lod-cloud.net/>

² <https://www.stardog.com/>

³ <https://virtuoso.openlinksw.com/>

2.1 Geospatial Linked Data

There is an ongoing trend of publishing geospatial data as Linked Data; initially Semantic Web researchers showcased the potential of Linked Data by transforming popular, third-party datasets to RDF, and then more Linked Data initiatives have been run by governmental agencies and large-scale data infrastructures [5]. For instance, Ordnance Survey (OS), the national mapping agency (NMA) in the UK, has released several geospatial datasets maintained by them as Linked Data [6]. In Europe, the e-Government and open data communities are increasingly adopting the Linked Data approaches, and this has motivated the Joint Research Centre (JRC) of the European Commission to investigate the potentials of publishing the INSPIRE-compliant geospatial data as Linked Data through the ARE3NA activity⁴. Varanka and Usery [7] argued that the map data (geospatial data) that are released in RDF according to corresponding ontologies can be treated as the knowledge base entailed by the map content. In this respect, we hold the same opinion, and argue that more geospatial knowledge can be represented upon the map knowledge base.

The increasing geospatial Linked Data have stimulated the studies of linking data from other sources to such data and exploiting such Linked Data in graphical interfaces. The visualization of Linked Data, in general, refers to the techniques of visually presenting the links between entities to facilitate the intuitive discovery of underlying information and knowledge [8]. For geospatial data, the spatial context is crucial for easing this perception and discovery process. Therefore, the visualization of geospatial Linked Data is generally in the form of map mashups, in which the data are spatially represented as thematic data on top various base maps. To this end, several tools for exploiting such data through visual and graphic interfaces have been developed. For instances, LOD4WFS [9] enables geospatial Linked Data to be queried through web feature service (WFS) protocol and visualized in GIS programs. Map4RDF [10] provides the possibility of editing the underlying data, and connecting to statistical data. Nonetheless, these tools generally use predefined and hard-coded visualization settings in the programs. However, in the context of Semantic Web, we can use a knowledge-based approach for the visualization by formalizing the knowledge concerning how the geospatial data (in Linked Data in this case) are visualized using ontologies and rules. In this way, the knowledge can be more readily be shared, interpreted and reused.

2.2 Geospatial knowledge representation using Semantic Web technologies

The capacity of knowledge representation of Semantic Web through leveraging ontologies and rules has been recognized in geospatial domain for many years and used in a number of studies. These studies span several research areas of e.g. visualization, geoprocessing and information retrieval. For instances, Janowicz et al. [1] proposed a framework for modelling the knowledge and semantics using ontologies and SWRL rules, and used the framework as a semantic enabled profile of current OGC-complaint

⁴ <https://inspire.ec.europa.eu/news/linking-inspire-data-draft-guidelines-and-pilots>

SDI. Hofer et al. [11] developed a knowledge base to support the composition of geo-processing workflow, in which the ontologies were used to formalize the geooperators, and SWRL rules are used for formulating the rules associated with the geooperators chaining. Keßler et al. [12] employed ontologies and SWRL rules for context-aware geographic information retrieval, where they used ontologies for organizing the semantically annotated data and rules for deriving inference for context detecting. Gould and Mackaness [13] formalized the knowledge for on-demand map generalization using ontologies to facilitate the knowledge to be shared, expanded and reused in mapping systems. Huang et al. [14] formalized the knowledge for both visualization scales for geospatial features and the relations between thematic data and base maps using ontologies to enable geometrically self-adapting web maps.

With regard to the visualization of geospatial data, we argue that the knowledge in this respect needs to be more formally modelled to facilitate the sharing and reuse of such knowledge and also outreaching such knowledge to other domains. Data portrayal is an indispensable part of visualization for geospatial data, and it subjects the semantic challenge as the current standards for modelling such information lack semantics, and this hampers the exchange and reuse of such information. This issue has also been identified by OGC, and thereof they initiated several testbeds to investigate a semantic portrayal solution. They developed ontologies for semantically modelling the information of style, symbol, symbolizer and graphic [15].

3 Knowledge-based visualization coupling ontologies and rules

The investigations performed by OGC Testbeds provide solid ground for the geospatial community towards knowledge-based visualization for geospatial data and the vision of shaping a *web of knowledge for geovisualization*. However, we argue that the modelling of conditional portrayal rules is deficient.

Conditional portrayal is prevalent for visualization, i.e. the symbol/symbolizer used for visualizing a feature depends on the visualization scale and attribute/geometric data associated with the feature. In the ontologies developed by OGC testbeds, the SPARQL ASK queries are recommended to model such conditions. However, such rule modelling approach has several limitations: (1) although SPARQL can be utilized for expressing rules in Semantic Web, the queries on their own are not commonly accepted as rule modelling for knowledge presentation and inference⁵; (2) the semantics could potentially be misinterpreted because the SPARQL ASK constraints are generally used to check whether certain conditions currently hold in a (scope of) knowledge graph and therefore facilitate verification and inconsistency⁶. To address this issue, we argue that, in the environment of Semantic Web, we can leverage the rule-based inference and knowledge representation capacities and thus augment the use of geospatial rules to other areas of the mainstream IT world.

⁵ <https://www.w3.org/2003/12/swa/dawg-charter>

⁶ <https://www.w3.org/Submission/spin-modeling/>

The most commonly used semantic rules in geospatial domain are of the type of SWRL, and this is mainly because the SWRL is supported by Protégé ontology editor and several rule engines and ontology reasoners. However, SWRL has several limitations for geospatial applications; for example, SWRL adopts the *open world assumption* and thereof only supports monotonic semantics, and in some geospatial applications, we need to tackle the *no data* or *voidable* situations that entail the handling of non-monotonic semantics. In contrast to SWRL, the object-oriented SPIN (SPARQL Inferencing Notation) rules, that combines concepts from object-oriented languages, SPARQL query language, and rule-based systems to model rules in the Semantic Web, has better expressiveness and several advantages for geospatial applications, e.g. SPIN rules can address the non-monotonic semantics and allow spatial predicates to be readily embedded in the conditions within spatially enabled RDF store. Therefore, we argue that the geospatial domain could appreciate the SPIN rules more (before its successor SHACL⁷ is better supported by tools).

Therefore, we have developed a new knowledge-based visualization framework in which the ontologies and rules (SPIN rules) are tightly coupled. Figure 1 shows the ontologies (knowledge base) used for data portrayal, and the SPIN rules are coupled with the *style* through a predicate *hasRules*.

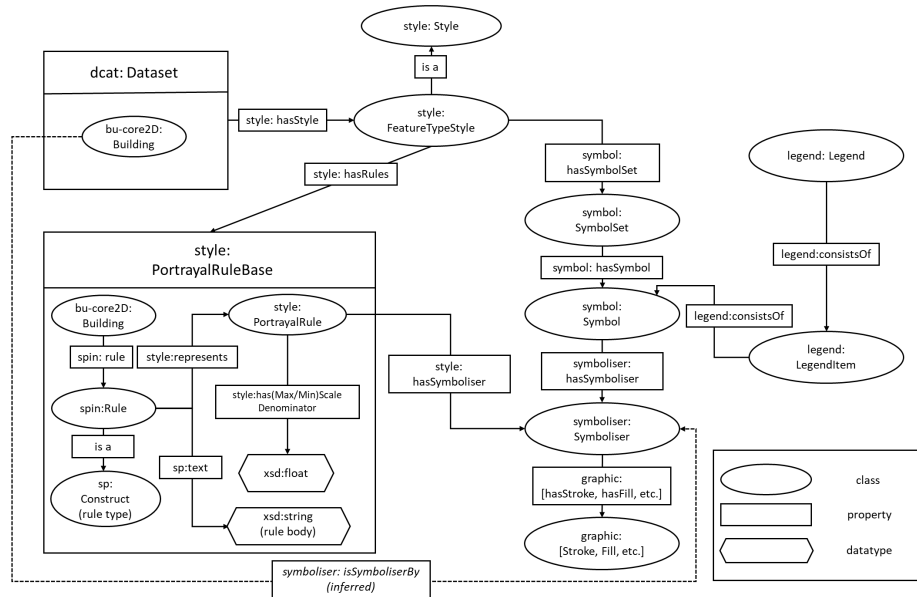


Fig. 1. The knowledge base for portrayal information.

Listing 1 demonstrates how a specific portrayal rule is formalized using SPIN in the syntax of Turtle. And this rule formulates that *if a building has been built for over 300 years and the rendering scale is larger than 1:10,000, then use the symboliser_0 to symbolize the building*. The INSPIRE vocabularies for 2D buildings are used in this

⁷ <https://www.w3.org/TR/shacl/>

case, and the age information is derived from the construction date. Furthermore, since the symbolizers used for portrayal can be different in different visualization scales. With a couple of such rules, the visualization programs can expose simple SPARQL queries to retrieve the symbolizers used for portrayal in different scales and for the features with different attribute values. In addition to this, we also designed ontologies and rules for the knowledge of geometry source, i.e. different geometries are used for each feature under different conditions and in different visualization scales (we designed the ontology for scale information as metadata for geospatial data). Such rules are also modelled in SPIN rules.

```
@[prefix definitions]
bu-core2D:Building a owl:Class;
    spin:rule[
        a sp:Construct;
        sp:text"""
        CONSTRUCT {?this symbolizer:isSymbolizedBy
                    portrayal:symbolizer_0}
        WHERE{
            ?this bu-base: AbstractConstruction.dateOfConstruction/
            bu_base:DateOfEvent.beginning ?built_up_time.
            BIND(year(now())-
                year(xsd:dateTime(?built_up_time)) as ?age)
            FILTER(?age>300)
            ?client_scale a scale:ClientVisualisationScale;
                scale:hasScaleValue ?rendering_scale.
            FILTER(?rendering_scale<=10000)
        }
    ].
```

Listing 1. An example of using SPIN rule to represent a portrayal rule in the syntax of Turtle.

4 Roadmap towards further knowledge-based visualization

At this stage, we have created the knowledge base for visualization of geospatial data through tightly coupling ontologies and rules. However, the information modelled here is still insufficient. In order to further accomplish the vision of *web of knowledge for visualization*, we need to incorporate more visualization/cartography knowledge which is often embedded in complex programs or mind of cartographers. And we believe that an infrastructure of such knowledge base would be of substantial help for knowledge transfer.

Recently, we have initiated a research cooperation with cycling researchers to visualize the cycling level-of-service in a spatial context (maps) to help the decision-makers to observe the cycling infrastructure situation in real maps rather than merely spreadsheets or sketch maps. However, the challenges arise in two respects: data integration

between the cycling data and geospatial features, and the transfer of visualization/cartography knowledge to the cycling researchers to generate competent geospatial visualizations. Simply put, the challenges lie in data integration and knowledge formalization, and this is where the Semantic Web technologies stand up. To address this issue, we are planning to employ a knowledge-based approach, in which on the one hand, the derivation of cycling level-of-service indexes is formalized using ontologies and rules; and on the other hand, the visualization/cartography knowledge concerning e.g. the color scale used for rendering, setting the width of the cycle lanes to make a slight separation from the vehicle lanes to avoid confusion for the users, and also to embed more information and knowledge into the legend to facilitate the users to perceive the content of the thematic map.

There are still some challenges that need to be resolved to realize the knowledge-based approach for visualizing cycling level-of-service, including:

- How to design ontologies to enable interoperability between geospatial road data and cycling data collected by the cycling professionals of e.g. the type of the cycle lane, the interaction between cycle lane and the adjacent vehicle lane? We preliminarily plan to use the relative positioning approach proposed by [14] to facilitate the information propagation from vehicle lanes to cycle lanes.
- What rule language would be sufficient for modelling the knowledge concerning both the derivation of cycling level-of-service indexes as well as corresponding cartographic knowledge of e.g. colour scale and feature displacement (which is in fact more complex than the cycling knowledge). To this end, we would investigate the successor of SPIN: SHACL, which includes the rule-based inference capacity as its advanced feature⁸. Also, we could also use SHACL for data validation, which is also indispensable in our cross-domain data and knowledge sharing.
- How should the approach be evaluated e.g. in comparison to traditional methods?

5 Conclusion

In this paper, we have presented a framework for shaping a knowledge-based approach for geospatial data integration and visualization, which can also be used for outreaching the geospatial data and knowledge to other domains. We have designed a knowledge representation approach tightly coupling ontologies and rules for the geospatial visualization knowledge on the aspects of scaling, data portrayal and geometry source. In the next steps, we will incorporate more visualization knowledge that could be used in different domains, and on such case that has been formulated is the visualization of cycling level-of-service in spatial context. We expect this case study would showcase the advantage of Semantic Web technologies in terms of data and knowledge sharing between different domains.

As stated earlier, the scope of this PhD project is broad and it investigates the benefits of Semantic Web technologies could bring up for geospatial applications. Hence, it is also interesting to investigate e.g. the use of Semantic Web (including the use of

⁸ <https://www.w3.org/TR/shacl-af/>

semantic rule) for real time integration between dynamic data (social media data) and static geographic data in the context of disaster management. Also, how the semantic gazetteers could foster better automatic construction of knowledge graph which includes spatial data is also an interesting topic to study.

References

1. Janowicz, K., Schade, S., Bröring, A., Keßler, C., Maué, P., Stasch, C.: Semantic enablement for spatial data infrastructures. *Transactions in GIS* 14, 111-129 (2010)
2. Schade, S., Smits, P.: Why linked data should not lead to next generation SDI. In: *Geoscience and Remote Sensing Symposium (IGARSS) 2012 IEEE International*, pp. 2894-2897. IEEE (2012)
3. Atemezing, G.A., Troncy, R.: Comparing vocabularies for representing geographical features and their geometry. In: *Terra Cognita 2012 Workshop* (2012)
4. Perry, M., and John Herring.: OGC GeoSPARQL-A geographic query language for RDF data. Open Geospatial Consortium technical report (2012)
5. Regalia, B., Janowicz, K., Mai, G., Varanka, D., Usery, E.L.: GNIS-LD: Serving and Visualizing the Geographic Names Information System Gazetteer As Linked Data. In: *European Semantic Web Conference*, pp. 528-540. Springer (2018)
6. Goodwin, J., Dolbear, C., Hart, G.: Geographical linked data: The administrative geography of great britain on the semantic web. *Transactions in GIS* 12, 19-30 (2008)
7. Varanka, D.E., Usery, E.L.: The map as knowledge base. *International Journal of Cartography* (in press)
8. Dadzie, A.-S., Rowe, M.: Approaches to visualising linked data: A survey. *Semantic Web* 2, 89-124 (2011)
9. Jones, J., Kuhn, W., Keßler, C., Scheider, S.: Making the web of data available via web feature services. *Connecting a Digital Europe Through Location and Place*, pp. 341-361. Springer (2014)
10. Leon, A.d., Wisniewki, F., Villazón-Terrazas, B., Corcho, O.: Map4rdf-faceted browser for geospatial datasets. In: *Proceedings of the First Workshop on Using Open Ddata, W3C*, 19–20 June 2012, Brussels, Belgium (2012)
11. Hofer, B., Mäs, S., Brauner, J., Bernard, L.: Towards a knowledge base to support geoprocessing workflow development. *International Journal of Geographical Information Science* 31, 694-716 (2016)
12. Keßler, C., Raubal, M., Wosniok, C.: Semantic rules for context-aware geographical information retrieval. In: *European Conference on Smart Sensing and Context*, pp. 77-92. Springer (2009)
13. Gould, N., Mackaness, W.: From taxonomies to ontologies: formalizing generalization knowledge for on-demand mapping. *Cartography and Geographic Information Science* 1-15 (2015)
14. Huang, W., Mansourian, A., Abdolmajidi, E., Xu, H., Harrie, L.: Synchronising geometric representations for map mashups using relative positioning and Linked Data. *International Journal of Geographical Information Science* 32, 1117-1137 (2018)
15. Fellah, S.: Testbed-12 Semantic Portrayal, Registry and Mediation Engineering Report. Open Geospatial Consortium technical report (2017)