

Received February 25, 2020, accepted April 15, 2020, date of publication May 4, 2020, date of current version May 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2992023

Towards Knowledge-Based Geospatial Data Integration and Visualization: A Case of Visualizing Urban Bicycling Suitability

WEIMING HUANG^{1,2}, KHASHAYAR KAZEMZADEH³,
ALI MANSOURIAN¹, AND LARS HARRIE¹

¹GIS Centre, Department of Physical Geography and Ecosystem Science, Lund University, 223 62 Lund, Sweden

²Center for Spatial Studies, Department of Geography, University of California at Santa Barbara, Santa Barbara, CA 93106, USA

³Division of Transport and Roads, Department of Technology and Society, Faculty of Engineering, Lund University, 221 00 Lund, Sweden

Corresponding author: Weiming Huang (weiming.huang@nateko.lu.se)

This work was supported by Lund University.

ABSTRACT Geospatial information plays an indispensable role in various interdisciplinary and spatially informed analyses. However, the use of geospatial information often entails many semantic intricacies relating to, among other issues, data integration and visualization. For the integration of data from different domains, merely using ontologies is inadequate for handling subtle and complex semantic relations raised by the multiple representations of geospatial data, as the domains have different conceptual views for modelling the geographic space. In addition, for geospatial data visualization—one of the most predominant ways of utilizing geospatial information—semantic intricacies arise as the visualization knowledge is difficult to interpret and utilize by non-geospatial experts. In this paper, we propose a knowledge-based approach using semantic technologies (coupling ontologies, semantic constraints, and semantic rules) to facilitate geospatial data integration and visualization. A traffic spatially informed study is developed as a case study: visualizing urban bicycling suitability. In the case study, we complement ontologies with semantic constraints for cross-domain data integration. In addition, we utilize ontologies and semantic rules to formalize geospatial data analysis and visualization knowledge at different abstraction levels, which enables machines to infer visualization means for geospatial data. The results demonstrate that the proposed framework can effectively handle subtle cross-domain semantic relations for data integration, and empower machines to derive satisfactory visualization results. The approach can facilitate the sharing and outreach of geospatial data and knowledge for various spatially informed studies.

INDEX TERMS Geospatial data integration, data visualization, ontologies, semantic constraints, semantic rules.

I. INTRODUCTION

Over the last decades, the massive use of geospatial information in various application areas (e.g., traffic analysis and energy simulation) has gradually revealed the indispensable role of geospatial information for interdisciplinary spatially informed research. Geospatial information is a key enabler for solving societal problems across disciplinary boundaries [1], and one of the most powerful information integrators to bridge diverse sources of information [2]. Although increasingly different types of geospatial data (e.g., authoritative and crowd-sourced geospatial data) have been generated

and disseminated e.g., through the Internet, readily utilizing such data in a meaningful way still remains a challenge, especially for experts from other domains in which geospatial information is indispensable.

Today's geospatial data analysis heavily relies on data synthesis, as data from a single source usually does not suffice [3]. Data integration is the problem of combining data residing at different sources, and providing the user with a unified view of the data [4]. Integrating geospatial information with information from other domains often entails the challenge of dissolving semantic heterogeneity [5]. Other domains, which are not geospatial *per se*, usually hold different conceptual views of the space that is delineated by geospatial data. Consequently, in different domains, the terminology

The associate editor coordinating the review of this manuscript and approving it for publication was Feng Xia.

varies to represent the geographic space. Such a situation induces significant difficulties for both data integration and the consumption of the integrated data.

Accomplishing semantic interoperability for geospatial data integration has been studied intensively; most commonly, ontological approaches are employed to explicitly represent and bridge the semantics in different domains or data sources (see e.g., [6]–[9]). The ontological approaches empower machines to compute the relations between concepts and properties residing in different ontologies, thereby enabling ontology-based data retrieval or transformation to (partially) achieve semantic interoperability. However, ontology-based approaches are inadequate for handling geospatial data with multiple representations, see e.g., [10]. Multiple representations are a special matching problem for geospatial data, as the concepts seem the same, but are not applied in the same way in data, due to differences in the geometric representations [11]. For example, one building object with a point geometry and another building object with a polygon geometry can both be categorized as *Building* in the ontology, but they are fundamentally different in terms of geometric representation and data usage. We have encountered this data integration problem in a case of evaluating urban bicycling suitability, an interdisciplinary study between the geospatial domain and the traffic domain.

Urban planners have been committing to improve urban infrastructures to improve their suitability for bicycling, which is environmentally friendly and beneficial for people's well-being [12]. As a result, traffic researchers have developed several indexes for evaluating transport performance and quality of bicycling experience. To this end, bicycling level of service (LOS) is a framework of quantifying bicycling performance [13]; in this framework, several different indexes have been developed. In this case, we intend to employ a network-based LOS index: the level of traffic stress (LTS) [14]. The rationale for choosing this index is that the network-based nature of this index implies that both links (segments) and nodes (junctions) are quantitatively evaluated to derive a comprehensive understanding of the network's suitability and connectivity. LTS produces four ratings ranging from LTS1 to LTS4 based on the types of network element and three key roadway attributes: (1) number of vehicle lanes; (2) speed limit; and (3) bike lane width (other factors include bike lane lockage, appearance of a centerline on the road, parallel parking, and the presence of traffic signal etc.). Table 1 demonstrates the means of deriving the LTS value for *mixed traffic* (a type of link in the bicycling network). For the full explanation of the LTS, see [14].

Some of the variables used for the LTS can be found in geospatial (GIS) road databases, while other variables (e.g., the appearance of a centerline on the streets, the appearance and size of median at the junctions, and the type of bicycling link (e.g., mixed traffic)) must be collected in the field. Therefore, data integration between geospatial databases and the field-collected data becomes a prerequisite. However, such data integration is not smooth, partly due to the semantic

TABLE 1. Criteria for level of traffic stress in mixed traffic.

Speed Limit	Street Width		
	2-3 lanes	4-5 lanes	6+ lanes
Up to 25 mph	LTS 1 or 2	LTS 3	LTS 4
30 mph	LTS 2 or 3	LTS 4	LTS 4
35+ mph	LTS 4	LTS 4	LTS 4

Note: use lower value for streets without marked centerlines or classified as residential and with fewer than 3 lanes; use higher value otherwise.

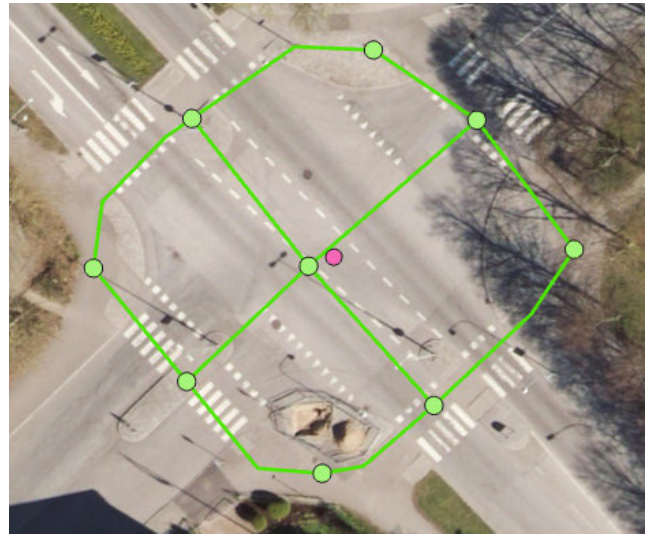


FIGURE 1. Illustration of different modelling approaches of a road junction in two levels of detail in the NVDB. Green links and nodes comprise the representation of this junction in the more detailed level, while the single red node represents the junction in the less detailed level. The background photograph is from ESRI's world imagery map.

heterogeneity and different conceptual views held by the two domains. One example is the modelling of links and nodes in the network. In Sweden, the national road database (*Nationell vägdatabas*, NVDB) models the road network in two levels of detail. In the more detailed level, the road links (network element that connects two nodes and represent a homogeneous path in the network) are comprehensively delineated, including multi-direction representations and the features of physically separated bikeways; the nodes in this detailed level are modelled according to the Swedish NVDB mapping rules [15], i.e., the detailed skeleton of the junctions are mapped by the vertices (points) and the links (poly-lines) between the vertices. In the less detailed (coarse) level, the links are modelled in a more generalized way, i.e., the lanes are aggregated and most of the dedicated bikeways are omitted; the nodes are modelled in a way that each junction is represented by a single node (point). Figure 1 illustrates an example of how a junction is modelled in two different levels of detail in the NVDB.

The multiple representations of geospatial data lead to semantic intricacies for traffic researchers. They need to integrate the records (each record represents either a junction or a road segment) in their field-collected data (spreadsheets) to geospatial features (instances). Traffic researchers need a comprehensive and detailed set of links, which corresponds

to the links in the more detailed level; whereas they view the junctions in the same way as they are modelled in the less detailed level (if present). It is unintuitive for them to link a junction record to a set of points and links. That is, in the conceptualization of the road network from the traffic domain, road junctions correspond to the data modelling approach in the coarse level of geospatial data, while links correspond to the more detailed level. Therefore, this becomes a cross-detailed-level data integration task. Such difficulty in geospatial data has led to mainly two types of compromise in previous studies: either the intersections are not explicitly represented on the map and the indexes of intersections are transferred to links [14], or the less detailed road dataset is used with manual editing, e.g., dedicated bicycling paths [16].

It is desirable to formally represent the subtle and complex semantic relations for sharing this knowledge and facilitate such cross-domain data integration missions, instead of experts from the domains having to discuss for each application of this kind. However, such cross-detailed-level semantic relations are difficult to capture merely using ontology alignment, because in most geospatial (network) ontologies, the level of detail information is modelled at dataset level (see e.g., [17]). For example, in the INSPIRE (infrastructure for spatial information in Europe) network ontology,¹ the concept of *node* is defined, while the above semantic relations between the different views of link and node in two different domains can hardly be captured using, for example, Ontology Web Language (OWL)² restrictions. For instance, OWL is not able to express the restriction *one junction instance from field-collected data should be linked to one node instance in the less detailed road network*. Therefore, we need a method to formally represent the semantic relations for data integration, and knowledge reuse.

Furthermore, another missing piece for performing this interdisciplinary case resides in the knowledge sharing and formalization of data usage, in which semantic challenges also often arise. This study entails the engagement of multiple analyses from different domains, including how to derive the LTS index and how to appropriately visualize the processed data on the map. In particular, geospatial data visualization (geovisualization) is a knowledge-intensive art and pertains to a wide range of cartographic knowledge, in which there are abundant semantic intricacies [18]–[21]. The knowledge from the two domains is usually embedded implicitly in complex software, or in the mind of domain experts. Traditionally, experts from one side have to either refer to literature or cooperate with the experts from the other side to accomplish such work [22]. Either of these ways is prone to misunderstanding due to the semantic heterogeneity between the domains. Moreover, such an informal way of knowledge sharing impedes the wide sharing, reusing, and expansion of that knowledge. Therefore, we also need methods to formally

represent the knowledge for data usage from the two domains to foster better communication and knowledge reuse.

The aim of this paper is to formalize knowledge from different domains for geospatial data integration and visualization for spatially informed studies using semantic technologies. Semantic technologies have been increasingly adopted in the geospatial domain [23], [24], and it possesses several knowledge representation paradigms that empower us to reinforce the bridges between different domains. The approach is showcased in the visualization of urban bicycling suitability with the LTS index, in which semantic heterogeneity is a significant impediment. Specifically, we leverage ontologies, semantic rules, semantic constraints, and linked data for data integration and visualization. The knowledge for data integration, derivation of LTS, and visualization is formally represented to foster better interpretability and reusability. Overall, the contributions of this paper are:

- 1) A framework for cross-domain and cross-detailed-level geospatial data integration is proposed, in which ontologies and semantic constraints are leveraged to represent complex and subtle semantic relations, in order to ensure the semantic correctness of data integration.
- 2) A knowledge base consisting of ontologies and semantic rules is developed for formalizing the knowledge of analysis (deriving the bicycling suitability index for a road network) and visualizing data on maps, which showcases the communication of knowledge from different domains for geospatial applications.
- 3) The knowledge for data analysis and visualization is represented at different abstraction levels, in order to ease cross-domain knowledge communications.
- 4) The knowledge base for data visualization is context-aware, i.e., the visualization varies in different contexts.

Following this introduction, Section II reviews related works in geospatial semantics for data integration and visualization. Section III presents an overview of the proposed approach, which is showcased in Section IV–VIII. Section IV provides information concerning the multi-source data and the study area of the case study; Section V and VI elaborate our proposed knowledge-based approach for geospatial data integration and visualization in the case study. Section VII presents the implementation details. Section VIII evaluates the proposed approach in the case study. The paper ends with a discussion (Section IX) and conclusions (Section X).

II. RELATED WORKS

The research in geospatial semantics has gained much attention in the last two decades, mainly in virtue of the increasingly widespread use of geospatial information in various domains. [25] viewed geospatial semantics in the context of semantic interoperability, and defined that geospatial semantics is about understanding GIS contents (geospatial information), and capturing this understanding in formal theories. We summarize research in geospatial semantics from three

¹<https://raw.githubusercontent.com/inspire-eu-rdf/inspire-rdf-vocabularies/master/net/net.ttl>

²<https://www.w3.org/TR/owl2-overview/>

perspectives, namely data integration, data processing, and data visualization.

The semantic challenge for geospatial data integration is well identified in the context of spatial data infrastructures (SDIs). [26] made a proposal to integrate the perspectives from GIS and computer science for establishing a European geographic knowledge graph as an SDI. [27] proposed a framework for semantically enabling SDIs, in which both the geospatial data and activities (discovery, registration, processing and visualization) are semantically annotated with ontologies. [23] utilized a linked data approach for geospatial data integration in SDIs, in which the data integration relations (e.g., matching relations) are formalized. [9] leveraged ontologies and logical reasoning for overcoming semantic heterogeneity in SDIs to foster better geospatial data exchange and reuse. [10] identified that many vocabularies have been defined within domains, whereas other domains are seldom taken into account; thus they proposed a methodology and tools for non-automatic, community driven ontology matching for geospatial data harmonization to facilitate the data reuse between datasets in the geospatial domain; at the meantime, they also identified that some subtle semantic relations can hardly be represented using ontologies. Ontology matching was massively used in these studies. In Section I, we illustrate a case, in which the semantic relations between geospatial and traffic domain can hardly be represented merely using ontology matching, thereby in this paper we propose to utilize semantic constraints to handle the subtle and complex semantic relations raised by the multiple geometric representations modelled in geospatial data. Furthermore, unlike addressing semantic heterogeneity in traditional SDIs, in this study we resolve the semantic challenge between geospatial datasets and a non-geospatial dataset (scientific dataset with no geometry information).

Semantics also plays a pivotal role in geoprocessing (geospatial data analytics), mainly for the purpose of semantically describing geoprocessing tools and establishing interoperability between them. A key research topic in this context is the workflow composition of geoprocessing, where semantic technologies have been increasingly utilized. [28] developed a knowledge base to support the composition of geoprocessing workflows, in which ontologies were used to formalize the geooperators (tools), and the semantic web rule language (SWRL) was used to formulate the rules associated with geooperator chaining. [29] formalized both geoprocessing tools and the requirements from the users using ontologies and SPARQL CONSTRUCT queries. These works only concentrated on describing geoprocessing tools at metadata level, e.g., input and output datatype of the tools, but not at the operation level, thereby the internal logics of the tools are not formalized. In this paper, we make a step further to formalize the geoprocessing tool at the operation level (cf. Section VI.A).

For geovisualization, it is commonly acknowledged that map making is an inherently human process that is difficult to automate, as computers are usually not capable of handling

perceptual properties of the data portrayal [18]. Nevertheless, cartographic knowledge can be formally represented to enhance computer aiding and the propagation of such knowledge. To this end, [19] distinguished extensive and intensive properties using machine learning techniques and formalized different types of properties using ontologies to help map making, as the cartographic rules applied to the two types of properties are fundamentally different. [17] designed an ontology for cartographic map scaling, as scale resides in the very core of cartography and is essential for geovisualization, and they formalized the cartographic scale information on the dataset level for representing the scale knowledge associated with geospatial datasets. [21] formalized the knowledge for both visualization scales for geospatial features and the relations between thematic data and base maps using ontologies, and structured the data accordingly in linked data to enable self-adapting web maps. [20] proposed to formally represent the knowledge of context-aware geovisualization in three aspects: cartographic scaling, data portrayal and geometry source, which are three prominent facets of geovisualization knowledge in the contemporary web mapping environment. They employed a semantic web technology based framework, in which linked data was leveraged as underlying data model, and ontologies and semantic rules (SPIN rules) were utilized for formalizing geovisualization knowledge in a both human- and machine-readable manner. However, the knowledge base designed in [20] relied on visualization settings at feature level, but did not include high-level cartographic knowledge, which made the knowledge representation work inefficient and difficult to be transferred to other applications. In this paper, we incorporate high-level cartographic knowledge into the knowledge base for geovisualization to ease the knowledge representation task and foster better transferability of the knowledge (cf. Section VI.B).

III. KNOWLEDGE-BASED GEOSPATIAL DATA INTEGRATION AND VISUALIZATION

This section provides an overview of the knowledge-based approach for geospatial data integration and visualization leveraging semantic technologies. The approach generally comprises two main parts: data integration and data visualization.

With regard to data integration, a semantic approach is employed. First, ontologies are designed to formally represent the semantics of the data from multiple sources (in this case the semantics of geospatial data with multiple representations and the field-collected data). Ontologies are formal representations of the knowledge within a domain of interest, which are defined by the concepts in the domain and the relationships between the concepts [30]. The ontologies can either be designed from scratch or (partially) reused from state-of-the-art standardized ontologies; the latter is encouraged whenever possible [31]. In the geospatial domain, many ontologies have been designed and standardized for the purposes such as data exchange and query. For example, in Europe, the INSPIRE directive has designed several

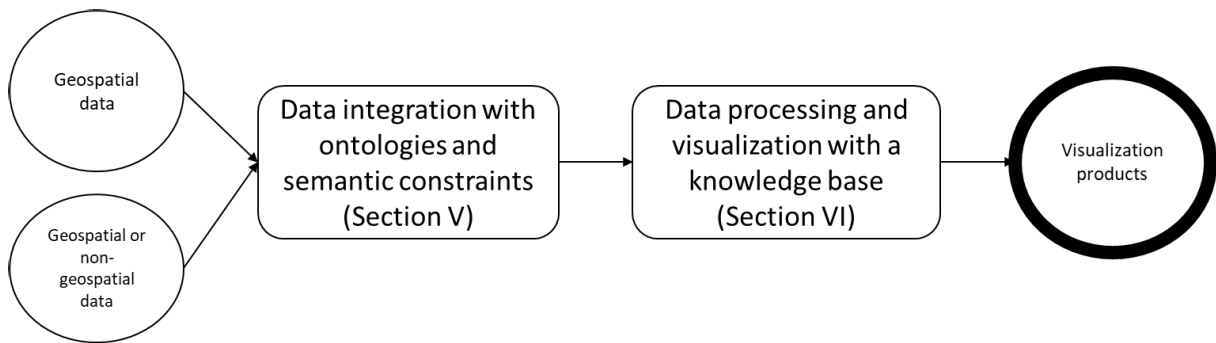


FIGURE 2. Workflow of the knowledge-based approach for geospatial data integration and visualization.

ontologies for representing geospatial data with different themes, e.g., road network [32]. Yet, for the bicycling LOS evaluation, there is no existing ontology to the best of our knowledge, thus we design the ontologies from scratch. The employed ontologies are then bridged via semantic relations from, for example, OWL, for data integration. However, for the relations that cannot be captured by semantic relations from OWL, we employ semantic constraints [33] to represent such subtle and complex relations. In the study, the complex semantic relations stem from the multiple representations of geospatial road network data. Data from different sources are then transformed to the semantic data model for linked data—Resource Description Framework (RDF) [34]—from their source data models, e.g., ESRI shapefiles for geospatial data. In order to explicitly represent the multiple representation relations of geospatial data, a multiple representation database (MRDB) is constructed before the data transformation to RDF. An MRDB organizes geospatial objects in different levels of detail, and the relations between the representations from different levels of detail are explicitly stored [35]. That is, the geospatial data in RDF have explicit relations between different representations of the geospatial objects. Then, the corresponding data instances, e.g., an intersection from the multi-scale road network and from the field-collected data, are matched. The matching relations are validated against the semantic relations represented by OWL constructs and semantic constraints. Such a knowledge-based geospatial data integration method is detailed in Section V.

For data visualization, the knowledge is formalized firstly to transform the integrated raw data to the phenomenon that is to be visualized, and the derived phenomenon is visualized according to the formalized visualization knowledge, i.e., how the geospatial data should be properly visualized on a map in a sense-making and cartographically satisfactory way. The knowledge for phenomenon derivation and data visualization usually stems from different domains. In our case, the knowledge concerning how to derive bicycling suitability indexes comes from traffic experts, and the knowledge for data visualization is from cartographers. The solicited knowledge is then formalized using ontologies and semantic

rules [20]. With the formalized knowledge encapsulated in ontologies and semantic rules, reasoners are able to derive phenomenon values and visualization means (e.g., styles and symbols) to develop the final visualization products. This knowledge-based geospatial data visualization approach is detailed in Section VI. A flowchart of our approach is illustrated in Figure 2.

All the ontologies, semantic constraints, semantic rules, and source codes used in this study can be found in a GitHub repository at <https://github.com/RightBank/Knowledge-based-integration-and-visualization>. We are, however, not permitted to distribute the data used in the study.

IV. STUDY AREA AND DATA

In this study, we showcase our approach in evaluating and visualizing the urban bicycling network in the center-west part of Lund, Sweden. The entire transport network is evaluated using LTS. According to the Swedish traffic regulation *Trafikförordning* (1998:1276), cyclists are legally allowed to ride in motor vehicle infrastructure even if a dedicated cycle path is available, unless bicycling is clearly prohibited. Therefore, we evaluate the dedicated bicycling infrastructure together with the motor vehicle infrastructure that is not prohibited for bicycling.

We utilize two main data sources: geospatial road networks (the NVDB) in two levels of detail with geometries (essential for visualization) and the information of lane numbers and speed limits, as well as the field-collected data containing other necessary information (variables) for LTS derivation. Figure 3 shows the multi-scale road network of this study area, and Figure 4 is a snapshot of field data collected by traffic researchers. In the more detailed level of the road network, there are 219 node features and 369 link features; in the less detailed level, there are 56 node features and 106 link features.

For the geospatial multi-scale road database, we create an MRDB. In the MRDB, the correspondence relations of network elements in two levels of detail are identified and stored. The relations between links in the road network are identified by the tool *Generate Rubbersheet Links*, and the relations between intersections are identified by the tool



FIGURE 3. Multi-scale geospatial road networks (NVDB) in the study area (center-west of Lund, Sweden). Left is the more detailed level of NVDB, and the right one is the less detailed level.

	A	B	C	D	F	G	H	I
1								
2		Local ID ▾	Location ▾	Length ▾	Traffic Type ▾	Speed(mph) ▾	N Lane ▾	Centreline ▾
3		S-1	Steglitsvägen	308	Mixed	18	2	N
4		S-2	Steglitsvägen - Örnvägen	85	Mixed	18	2	N
5		S-3	Hökvägen (Steglitsvägen - Örnvägen)	85	Mixed	18	2	N
6		S-4	Starvägen (Steglitsvägen - Örnvägen)	85	Mixed	18	2	N
7		S-5	Örnvägen (Steglitsvägen - Örnvägen)	85	Mixed	18	2	N

FIGURE 4. Snapshot of the data collected in the field by traffic researchers.

spatial join; both of the tools are from ESRI ArcGIS Pro 2.0.0. The identified relations (links) are thoroughly inspected to ensure their semantic correctness.

V. KNOWLEDGE-BASED GEOSPATIAL DATA INTEGRATION USING ONTOLOGIES AND SEMANTIC CONSTRAINTS

This section elaborates the knowledge-based data integration approach in the context of the case study. The approach is based on formal representations of data semantics and the correspondence relations between the means of representing geographic objects in the geospatial domain and the traffic domain. The ontologies and semantic constraints used in this study are all available in the GitHub repository of this paper.

A. SEMANTIC ENRICHMENT FOR GEOSPATIAL DATA

It is a common practice to leverage ontologies to formally represent the taxonomy in each domain, and use semantic relations (e.g., relations in OWL or SKOS³ ontologies) to bridge the domains for e.g., data exchange and integration.

For geospatial data, the ontologies for representing geospatial networks and the information regarding the level of detail

(cartographic scale) are necessary. In this study, we utilize the INSPIRE network ontology (*net* as prefix). This ontology defines the key concepts for geospatial networks, such as *Network*, *Link*, and *Node*. The geometric information is defined by incorporating the simple feature part of GeoSPARQL (*sf* as prefix)—a query language for geospatial linked data [36]. A *Network* instance can be associated with a number of *Link* and *Node* instances that are both *NetworkElement*, and *Link* and *Node* instances can also be connected to express the connectivity of the network. We create three subclasses of the class *Link*: *Bikeway*, *Motorway* and *CrossingLink* (the links comprise the junctions) for different types of links in the road network; and also three subclasses of the class *Node*: *Intersection*, *Roundabout* and *DeadEnd*. For the level of detail information, we partly reuse and complement the cartographic scale ontology (*scale* as prefix) from [16]; that is, the two properties of *hasMaxScaleDenominator* and *hasMinScaleDenominator* are defined to associate the datasets (geospatial networks with different levels of detail) with respective visualization scales, and the properties of *isMoreGeneralThan* and *isMoreDetailedThan* (inverse properties) are defined to represent the relations between datasets. The corresponding features identified in two levels of detail when constructing the MRDB (cf. Section IV) are associated by properties in

³<https://www.w3.org/TR/2008/WD-skos-reference-20080829/skos.html>

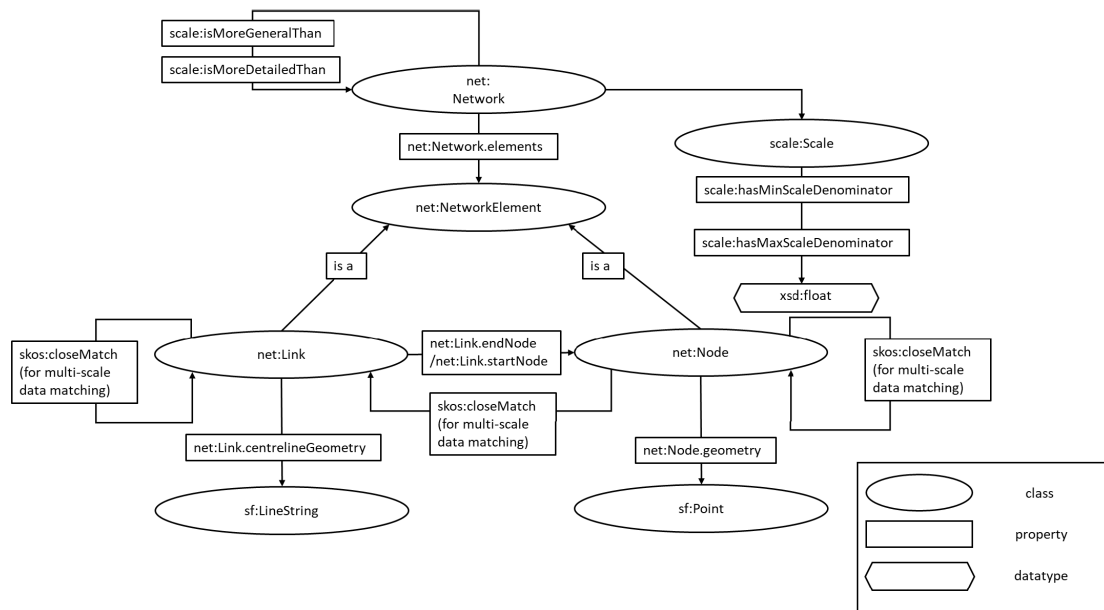


FIGURE 5. Core parts of the ontologies for semantically representing multi-scale geospatial networks.

the SKOS vocabulary, i.e., using *skos:closeMatch* relation to associate the corresponding features at different levels of detail (e.g., a node in the coarse level can be matched to a set of nodes and links in the detailed level). Figure 5 demonstrates the essential parts of the ontologies for semantically organizing the multi-scale NVDB in RDF.

B. SEMANTIC ENRICHMENT FOR FIELD-COLLECTED ROAD DATA

The data collected in the field for evaluating the bicycling network based on the LTS are recorded in spreadsheets (tables). Tables are a common way to store and exchange data, e.g., on the Web, whereas most of the tables' information is only understood by humans but not machines. In fact, the tables are sometimes even difficult to understand for humans, particularly in the interdisciplinary studies such as this case where the table data need to be understood by experts from another domain. Therefore, it is important to formally and explicitly represent the semantics in the tables, so that they can be unambiguously understood by both humans and machines. This is in line with the research topic of semantic table interpretation in the semantic web domain [37]. In our study, the meaning of the table data is unclear for the geospatial experts, and this hampers the data integration task. Therefore, we develop ontologies for representing the semantics in the traffic domain, especially for the data used for the LTS. Developing ontologies and building relations with geospatial ontologies can not only ease the cross-domain communication, but also facilitate the reuse and sharing of such knowledge. We develop ontologies from scratch, as no previous work has been accomplished in this regard. The ontologies are designed through several comprehensive discussions between traffic researchers

and knowledge engineers, and the ontology design approach *METHONTOLOGY* is employed to build glossary, concepts, and relations [38]. The ontologies are developed in two levels: the LOS level and the LTS level (the LTS is an index in the framework of LOS that includes a series of evaluation methods). The rationale for developing the multi-tier ontologies is that we model a part of visualization knowledge at the LOS level, i.e., the cartographic rules apply to all LOS indexes, including the LTS (see Section VI.B). In the upper level ontology of LOS, the common concepts and relations for deriving indexes (including the LTS) are defined, including the concepts of *LOSIndex*, *BicyclingNetwork*, *BicyclingNode*, and *BicyclingLink*, and the relations of *hasLOSIndexValue*, and *isMatchedTo* (for associating an instance of bicycling link or node to the instances of geospatial network element). The developed LTS ontology incorporates the concepts and relations used specifically for the LTS index, and is developed on the basis of the LOS ontology. Essentially, the concept *LTS* is defined as a subclass of *LOSIndex* with four instances of this class, i.e., *LTS1*, *LTS2*, *LTS3*, and *LTS4* (including rich semantics about the four levels); the types of bicycling network elements defined in LTS are incorporated: the concepts of *BikePathWithPhysicalSeparation*, *BikeLaneWithMarking*, *MixedTraffic*, and *PocketBikeLane* are defined as subclasses of *BicyclingLink* in different abstraction levels; the concepts of *Crossing*, *SignalisedCrossing*, *UnsignalisedCrossing*, *CrossingWithMedianRefuge* are defined as subclasses of *BicyclingNode* in different abstraction levels. An object property *hasLTSValue* (this relation is inferred based on semantic rules, see Section VI.A) is created as a subproperty of *hasLOSIndexValue*. Other properties needed for LTS derivation are also defined, e.g., *rightTurnLaneType*, *hasCentreline*, and *isConnctedTo* (denoting the connectivity between

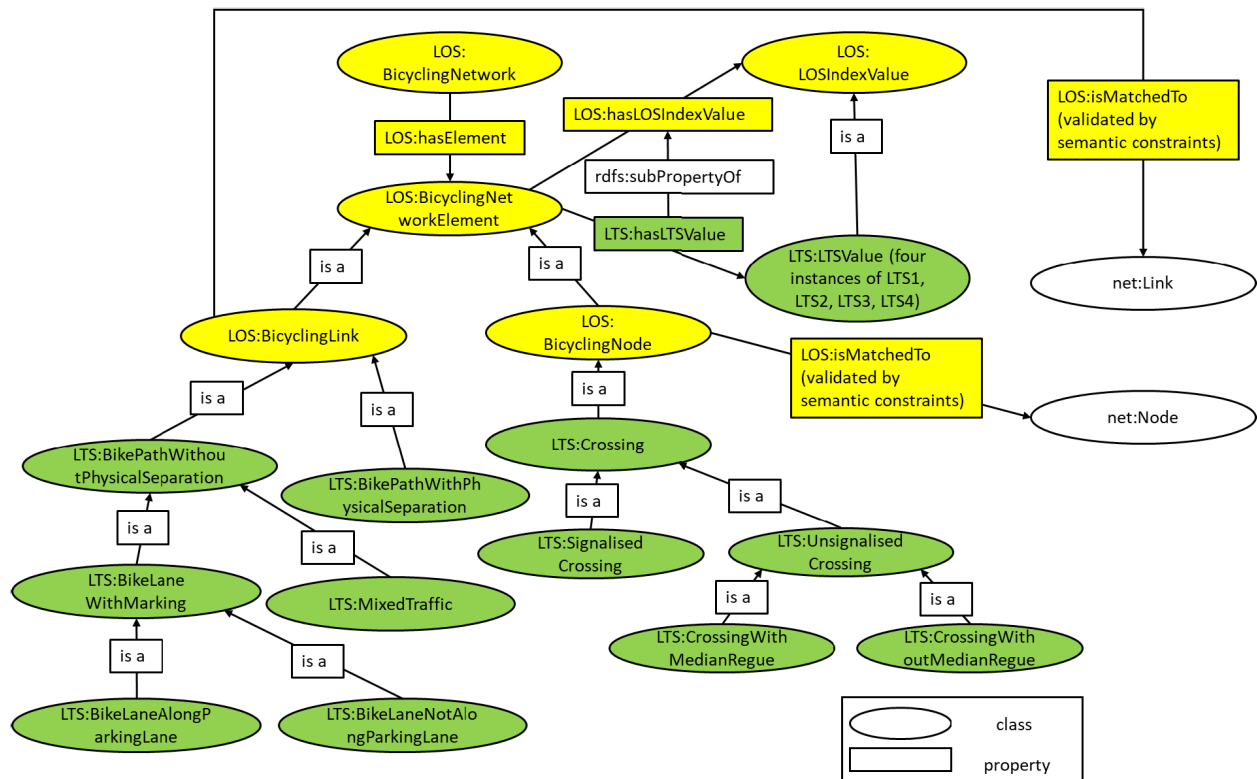


FIGURE 6. Core concepts and relations in the LOS and LTS ontologies, including their links with the INSPIRE network ontology. The elements in LOS ontology are with yellow color, and green color is for LTS ontology.

bicycling nodes and links). Figure 6 illustrates the concepts defined in different abstraction levels in the LOS and LTS ontologies.

In the LOS ontology, there are 4 classes, 3 object properties, and 2 data properties. The LTS ontology includes 17 classes, 2 object properties, and 7 data properties.

C. DATA INTEGRATION WITH SEMANTIC CONSTRAINTS

According to the semantic relations identified and discussed in Section I:

- 1) a bicycling link should be matched to at least one link in the detailed level geospatial road network
- 2) a bicycling node should be matched to exactly one node in the coarse level network data if that node feature is available in the coarse level, otherwise the node should be matched to one node in the detailed level (e.g., small junctions are only present in the detailed level with single points).

Since the level of detail information is defined at dataset level (the scale is associated with *net:Network* instance that represents the entire network in one level of detail), OWL, which is often used for representing data restrictions, is not capable of representing such subtle semantic relations and complex integrity constraints. Moreover, OWL was designed for reasoning, but not data constraints. OWL restrictions describe the reasoning to be applied based on them [39]. For example, assuming there is an *owl:maxCardinality 1* restric-

tion stating that one bicycling node (*LOS:BicyclingNode*) can only be matched to one geospatial node (*net:Node*) feature at maximum, and there are two *net:Node* instances matched, then an OWL reasoner will assume that the two *net:Node* instances must in fact represent the same real-world entity. Furthermore, OWL adopts the *open world assumption*,⁴ and thus, assuming an irrelevant instance (e.g., a building instance) is mistakenly matched to a bicycling node, then the reasoner will infer that the building instance is also a *net:Node* instance. Additionally, the *owl:minCardinality* will not report any integrity error of missing values, because more data may appear at any time to satisfy that restriction under the *open world assumption*.

Due to the limitations of OWL, there have been many efforts to develop data constraints for RDF graphs, see e.g., [33] for semantic environmental data validation. In this context, the shapes constraint language (SHACL) became a W3C (World Wide Web Consortium) recommendation in 2017 [40]. The W3C recommendation made SHACL the most promising technique for becoming the *de facto* standard of semantic data constraints. Primarily, SHACL is a language for validating RDF graphs, and can also be used for other purposes including, among others, data integration. SHACL has been increasingly adopted in various domains and applications, e.g., clinical information systems, and

⁴http://wiki.opensemanticframework.org/index.php/Overview_of_the_Open_World_Assumption

software regression testing [41], but has been seldom used in the geospatial domain. We argue that such semantic constraints have unexplored potential for geospatial linked data, which mostly do not adopt the *open world assumption*, and there is a significant need for the integrity assurance and data integration. Such need becomes more prominent in the spatially underpinned interdisciplinary studies, in which subtle and complex semantic relations between geospatial and other domains are often inevitable. This is also in line with the opinions from [10], who identified the problem of missing semantic relations for representing concept relations for multi-source geospatial data. In this context, semantic constraints can be leveraged to handle complex and subtle semantic relations.

We employ SHACL constraints for representing the subtle semantic relations for integrating the field-collected data and multi-scale road network data, i.e., the matching relations between *link* and *node* in the two domains. Listing 1 is a SHACL constraint (*sh* as prefix for namespace of SHACL) that is used for representing subtle semantic relations between *net:Node* and *los:BicyclingNode* for data integration. This constraint assures that an instance of *los:BicyclingNode* will be matched to an instance of *net:Node* in the coarse level of detail network if available, otherwise it must be matched to a *net:Node* in the detailed geospatial network. Once the constraints are violated, SHACL will generate reports to facilitate the identification of semantic mismatches [40]. Moreover, the subtle semantic relation is formally represented thanks to the expressive SHACL semantics and the SPARQL query embedded. With the formalization of such subtle semantic relations, this interdisciplinary study is eased, as such semantic constraints can be readily reused and expanded. Simply put, the bridge between the domains is reinforced than merely using ontologies. Note that the ontologies and semantic constraints provide a semantic framework for data integration to ensure semantic correctness for data integration, and the formally represented knowledge concerning how to incorporate multi-scale geospatial data into analysis can be readily interpreted and reused. By contrast, the matching between individual data objects (e.g., matching a record from field-collected data with a geospatial feature) is not automated by this framework. In this case, object matching (integration) is performed manually depending on the road name information, as the geometric information is not recorded in the field-collected data and distance-based object matching cannot be conducted. The results of the object matching process are validated against the ontologies and semantic constraints to spot the semantically incorrectly matched objects. In addition, the matching is revised according to the hints given in the error reports.

VI. KNOWLEDGE-BASED GEOSPATIAL DATA VISUALIZATION WITH ONTOLOGIES AND SEMANTIC RULES

Due to the interdisciplinary nature of the case study, evaluating bicycling suitability and then visualizing the

evaluation on the map entail the incorporation of knowledge from the two domains. We intend to develop knowledge bases to formally represent the knowledge from the two domains, i.e., derivation of the LTS and the map visualization. Such knowledge bases would foster better communication between the two domains, and they can be readily reused, rather than requiring domain experts to consult literature or cooperate each time.

In this study, we encapsulate the domain data analysis methods (derivation of LTS and visualization rules) using ontologies and semantic rules. Semantic rules (horn logic) are a prominent knowledge representation paradigm in the semantic web. They offer a knowledge representation model for both domain experts and developers; semantic rules are more manageable and understandable than procedural codes as they lessen the semantic gaps between domains [42]. The developed ontologies and semantic rules are all available in the GitHub repository of this paper.

A. KNOWLEDGE BASE FOR LTS

Deriving the LTS index values for different types of bicycling links and nodes is complex, as each type of the network element has its own method for its derivation. In this study, we formally represent the LTS derivation using semantic rules to foster rule-based reasoning. Such formalized knowledge can be understood by both humans and machines, and can be reused for the calculation of this index in other use cases.

The LTS derivation is formally represented using the object-oriented SPIN (SPARQL Inferencing Notation) rules, that combine concepts from object-oriented languages, the SPARQL query language, and rule-based systems to model rules in the semantic web [43]. SPIN rules are increasingly widely used as they are expressive and close to SPARQL, and also support non-monotonic reasoning. In fact, SHACL's advanced features include semantic rules, which are an upgrade of SPIN rules, whereas such advanced features are not in the W3C recommendation, and few reasoners currently support SHACL semantic rules. Therefore, we opt to still use SPIN rules, which can be readily migrated to SHACL rules in the future if necessary.⁵

As a proof-of-concept, in this study, we develop SPIN rules for a subset of LTS derivation scenarios, i.e., the bicycling network element types of *mixed traffic*, *bike path with physical separation*, *pocket lane*, and *unsignalized crossing with(out) median* that appear in the research area. The index derivation for each type is formalized into a few rules to cover all the logics, and overall 17 SPIN CONSTRUCT rules are developed to formally represent a part of the LTS derivation and enable the reasoner to infer the LTS value, i.e., infer the object property of *lts:hasLTSvalue* and associate each instance of *los:BicyclingNetworkElement* with an instance of *lts:LTSValue*.

⁵ <http://spinrdf.org/spin-shacl.html>

```

@prefix definitions]

ex:NodeIntegrationShape a sh:NodeShape ;

    #the shape constraint is for bicycling nodes
    sh:targetClass los:BicyclingNode ;

    #the shape constraint is targeted to this relation
    sh:path los:isMatchedTo ;

    # each bicycling node must be matched to exactly one geospatial node
    sh:class net:Node; sh:minCount 1; sh:maxCount 1;

    # a SPARQL constraint to represent the conditions incorporating level of detail information
    sh:sparql [ a sh:SPARQLConstraint;

    # message in case of error arises

    sh:message "a bicycling node should be matched to a node in the coarse level data if available."
    sh:select ""

    SELECT $this (?detailed_road_node as ?value)
    WHERE { $this $PATH ?detailed_road_node .

        ?detailed_network net: Network.elements ?detailed_road_node;

        scale:isMoreDetailedThan ?coarse_network.

        ?detailed_road_node skos:closeMatch ?coarse_road_node.}"" ;] .

```

Listing 1. A SHACL constraint that states a bicycling node must be matched to one node in the geospatial network of the coarse level of detail if available, otherwise it must be matched to one node in the detailed geospatial network.

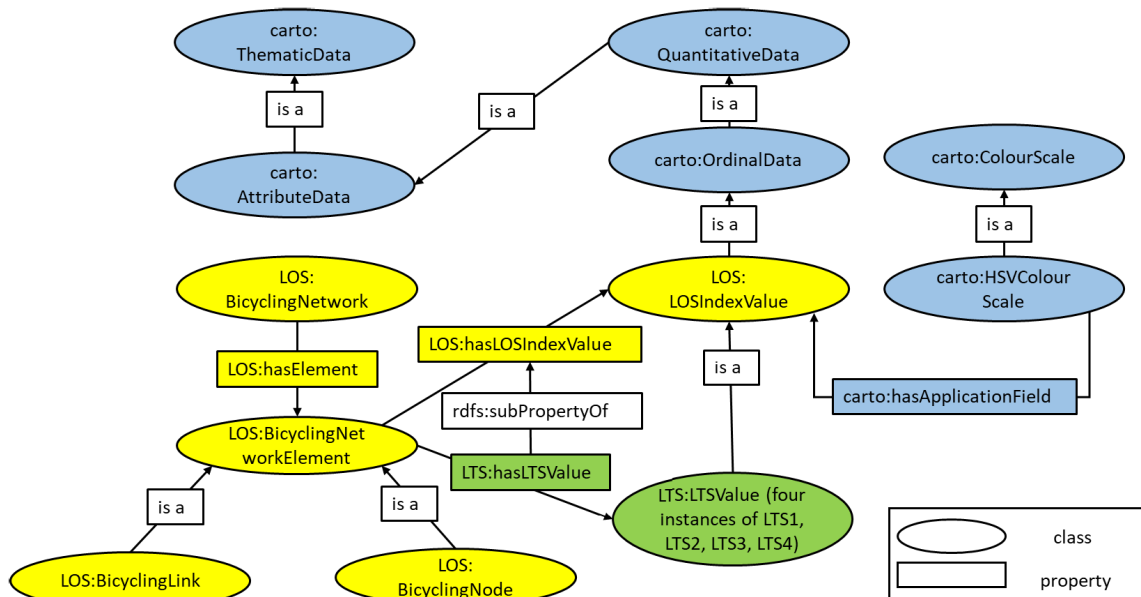


FIGURE 7. Core concepts and relations in the cartographic ontology and its relations with LOS and LTS ontologies. The cartographic ontology is annotated with blue color, the LOS ontology is with yellow, and the LTS ontology is with green.

B. KNOWLEDGE BASE FOR VISUALIZING GEOSPATIAL DATA

Similarly, the geovisualization (cartographic) knowledge can be formalized with ontologies and semantic rules, in order to facilitate the understanding of such knowledge in

interdisciplinary studies where information needs to be visualized on maps. To empower machines to understand cartography, it is commonly acknowledged that the cartographic knowledge framework should be formally represented using ontologies [44]. The authors of [20] designed ontologies and

semantic rules used for geovisualization, in which a data portrayal knowledge base comprises a major part. However, that work was for the purpose of web mapping, and did not incorporate high-level cartographic knowledge, i.e., common cartographic principles and rules. In this regard, [45] designed an ontology including many prominent cartographic concepts, e.g., cartographic method, and data types (e.g., according to measurement scale: nominal, interval, and ordinal data). We argue that although the data portrayal ontologies in [20] can explicitly represent the information of how features should be visualized under different conditions, the cartographic theories are unable to be readily utilized, which thus diminishes the automation level of knowledge modelling and representation. Therefore, in order to better leverage the cartographic theories, we complement the data portrayal ontologies in [20] with high-level cartographic knowledge. The visualization knowledge is mostly modelled with high-level cartographic concepts and relations, and then analyzed by semantic reasoning to transfer to lower-level data portrayal knowledge to render information on the maps.

For cartographic ontology, we reuse and extend the work of [45] (*carto* as prefix). We create the relation that *los:LOSIndexValue* is a subclass of *carto:OrdinalData* and *carto:ThematicData*, and we add the concept of *carto:ColorScale* with its subclasses *HSVColorScale*, *CMYKColorScale*, and *RGBColorScale* to represent the color scales in different color systems, as the color distinction is one of the most commonly used visualization practices. The defined concepts enable cartographers to model color scales to different types of thematic data. In this study, we model an HSV (hue, saturation, value) color scale for visualizing bicycling network elements with different LOS index levels, according to cartographic knowledge for ordinal data. We use a traffic signal color scale (green, yellow, and red), as the meaning of the colors in this scale is perceptible in the traffic domain. The defined color scale starts and ends at two certain HSV colors to represent the range (thereby defining the properties *carto:startsAtColor* and *carto:endsAtColor*), and the color scale instance is associated with the concept of *los:LOSIndexValue* using the property *carto:hasApplicationField* to denote the application field. Figure 7 illustrates the hierarchy and relations between the cartographic ontology, as well as the LOS and LTS ontologies.

Grounded upon the formalized concepts and relations in the ontologies, we then formalize generic cartographic rules using SPIN rules (with the prefixes of *spin* and *sp*). The color scale is evenly interpolated (one cartographic common rule) and then applied to different values of thematic ordinal data. Different values of line thickness are also applied to different types of links. The interpolation of the color scale is conducted in the three dimensions (hue, saturation, value) respectively. For real-time visualization, a portrayal rule base is created, consisting of four SPIN rules regarding using different symbolizers (basic units of visualization) under different conditions, and thus, how each feature should

be portrayed on the map can be deduced using semantic reasoning (cf. [20]). Listing 2 shows the symbolizers used for independent bikeways. In this rule, the color used for the thematic value (LTS value) is from the interpolation of the color scale modelled in the cartographic ontology. The interpolation of color scales for ordinal data (a type of thematic data) is formalized in a SPIN rule and derives the correspondence relations between each thematic value and color (with the property *carto:colorCorrespondsToThematicValue*). The interpolated colors modelled in the cartographic ontology are then transferred to a data portrayal rule (according to the data portrayal ontologies in [20]) in Listing 2 for assigning symbolizers to geospatial objects.

C. ABSTRACTION LEVELS OF DATA USAGE KNOWLEDGE

As described above, the knowledge concerning data usage, i.e., the derivation of LTS and its visualization is formalized. In this process, we create three knowledge representation abstraction levels, and different types of data usage knowledge are modelled at different levels. The three levels are: (1) cartographic common knowledge; (2) visualization knowledge for the LOS (theoretically it can cover all kinds of LOS indexes); (3) the particular LTS index level.

At the cartographic common knowledge level, the core concepts and relations of cartographic theories are modelled in the cartographic ontology. In principle many rules can be modelled at this level (e.g., cartographic rules for ordinal data) using semantic rules. In this study we showcase this by developing a rule of color scale interpolation at this level, i.e., the color scale is interpolated evenly according to the number of the ordinal thematic data (LTS1–4 in this case). As the subclass inheritance is formally defined in the ontologies (see Figure 7), all the semantic rules modelled at this level also apply to lower level leveraging ontological reasoning, i.e., the object-oriented SPIN rules modelled in the *carto:OrdinalData* level also apply to lower level concepts of *los:LOSIndexValue*, and thereby, *lts:LTSValue*.

At the second knowledge abstraction level—LOS level—we model all the application-specific visualization knowledge. An instance of *carto:ColorScale* (a color scale that fits the traffic phenomena visualization) is created with the application field of *los:LOSIndexValue*. After the color for each index value is retrieved through color scale interpolation, the semantic rules assign different colors to different LOS index values. For example, in the semantic rule in Listing 2, the property *los:hasLOSIndexValue* is used, which has a subproperty of *lts:hasLTSValue*; each link or node is associated with an LTS value through the property *lts:hasLTSValue*, thereby the upper level property *los:hasLOSIndexValue* can be used to retrieve the LTS values. In addition, all the line thickness rules are formalized at this level.

The derivation of the LTS is formally represented at the lowest level—the LTS abstraction level—to deduce the associations between the bicycling network elements and the LTS values ranging from LTS1 to LTS4 through *lts:hasLTSValue*.

```

@ [prefix definitions]
net:Link a owl:Class;
    spin:rule{
    a sp:Construct;
    style: representsPortrayalRule [portrayal_rule_URI]
    sp:text"""
    CONSTRUCT{ symbol:LOSSymbol symbolizer:hasSymbolizer ?symbolizer.
                ?symbolizer a symbolizer:LineSymbolizer;
                graphic:strokeColor ?color;          graphic:strokeWidth 3.
                ?this symbolizer:isSymbolizedBy ?symbolizer.}

    WHERE{
    ?bike_path a los:BicyclingLink;
    los:isMatchedTo ?this.
    ?this a net:Bikeway;
    ?bike_path los:hasLOSIndexValue ?LOSValue.
    # carto:colorCorrespondsToThematicValue was inferred during color scale interpolation.
    ?color carto:colorCorrespondsToThematicValue ?LOSValue.
    BIND(IRI(CONCAT("urn:symbolizer_for_independent_bikeway_",STR(?this))) as ?symbolizer) }"""

```

Listing 2. An example SPIN rule for assigning symbolizers for independent bikeways.

No visualization knowledge is defined at this level; they are instead transferred from upper abstraction levels.

The rationale of modelling the knowledge of data usage (analysis and visualization) at different abstraction levels is that, we believe it is unrealistic for cartographers to model how the data should be visualized for every single application, rather the applications can be aggregated to ease such knowledge representation work. The knowledge modelled at the LOS level can be used for every LOS index, as long as the subclass inheritance is explicitly represented. In this case, such knowledge transfer from upper level to lower level is showcased with the LTS—a particular LOS index. Thanks to the semantic reasoning capabilities, every semantic rule modelled at the upper level also applies to lower levels, therefore the number of knowledge abstraction levels can be increased if necessary, e.g., by adding another abstraction level of *traffic thematic data*, of which *los:LOSIndexValue* is a direct subclass.

D. CONTEXT-AWARENESS OF DATA USAGE

In principle, the data used in this case can be used for different analyses, e.g., the multi-scale geospatial road network data can be also used for traffic congestion analysis, in addition to the bicycling suitability analysis in this study. Therefore, the context information is crucial for spatially informed studies, informing the knowledge bases of the data usage contexts. Semantics plays a pivotal rule for modelling the context information [46]. The knowledge-based approach unlocks the opportunity of context-aware geospatial data visualization, i.e., the analysis or visualization methods can vary according to the data usage context.

In this study, the ontologies and semantic rules are context-aware. The visualization context is transferred to the knowledge base from the client side, and the context information

thereafter is involved in the semantic reasoning to deduce appropriate analysis and visualization means for the current context. Therefore, we create a light-weight visualization context ontology with the class of *VisualizationContext*; a *VisualizationContext* instance can be associated with a *carto:Phenomenon* instance through the property *visualizesPhenomenon*. In this case, *los:LOSIndexValue* is assigned as a subclass of *carto:Phenomenon*. The context information in the knowledge base can be updated according to the information transferred from the visualization client. In this case, the visualization context (*VisualizationContext*) instance visualizes (*visualizesPhenomenon*) the thematic data of *lts:LTSValue*. Then the rule-based reasoning deduces that the rules of LTS derivation and the visualization knowledge for LTS should be used (the color scale and rules for LOS).

With our knowledge-based approach, it becomes possible that a number of different knowledge bases for different contexts co-exist, and the context data is used for invoking the appropriate knowledge bases (ontologies and semantic rules) for data consumption and visualization.

VII. IMPLEMENTATION

A. DATA TRANSFORMATION

An MRDB is created based on the data from the multi-scale NVDB (see Section IV). The data (originally in ESRI shape-files) are transformed to RDF according to the INSPIRE network ontology with the correspondence relations of the features in two levels of detail (*skos:closeMatch*). The scale information is added at the network (dataset) level to denote the visualization scales and the level of detail information of the networks. The field-collected data recorded in spreadsheets are also transformed to RDF according to the LTS ontology (see Section V.B). The data transformations

are performed using R2RML⁶ transformation supported by Ontop.⁷

B. CROSS-DOMAIN DATA MATCHING

This step interlinks the road network element objects in MRDB with the bicycling link or node objects collected in the field using the relation *los:isMatchedTo*. The data matching is empowered by semantic constraints to tackle the subtle semantics of geospatial data raised by multiple representations. Two semantic constraints (in SHACL) are developed, of which one is for nodes, and the other is for links (see Section V.C). The correspondence relations (*los:isMatchedTo*) are identified manually depending on the road name information and validated against the SHACL constraints using the Jena⁸ and SHACL API⁹ in a Java environment. After this step, the data from the two domains are matched in a semantically correct way. All the RDF data (including MRDB, field-collected data and the cross-dataset links) are imported to the RDF store RDF4J.¹⁰

C. ENABLING RULE-BASED INFERENCE FOR DATA VISUALIZATION

In this study, we formally represent the knowledge concerning data usage, i.e., we use ontologies and semantic rules to derive LTS values as the evaluation metric and thereafter derive cartographically satisfactory visualization methods for the bicycling network depending on the LTS values. The semantic rules are developed by writing the domain logic into SPIN rules manually. The rules are also imported into RDF4J, which has the rule-based inference capacity. The LTS values and visualization means for the data objects are inferred over the data with the combination of ontological reasoning and rule-based reasoning.

D. VISUALIZATION TOOL

The results are visualized in a web-based environment and with a client/server architecture. A server is implemented using the Python web framework Django¹¹ to communicate with the knowledge base (data, ontologies, and semantic rules in RDF4J). The server sends SPARQL queries to the knowledge base, in which it asks the knowledge base to send all the geospatial objects (in the detailed road network) and the visualization means (symbolizer) of each object to the server. The server then parses the retrieved data (e.g., fetches the CSS values associated with the symbolizers) and wraps the data into JSON objects. The JSON objects are then sent to the frontend developed mainly using the web mapping library Leaflet.¹² The frontend (browser) parses the received data and visualizes the bicycling network according

to the visualization means (CSS values) encapsulated in the JSON objects. In order to enable the users to interactively understand the visualization, one could click the bicycling network elements and obtain further information (e.g., LTS value, and element type) in the popped-up RDF4J faceted browser. The users could also explore the knowledge base as every data object is dereferenceable in the faceted browser using URIs. The source code of the web-based visualization tool is available in the GitHub repository.

VIII. EVALUATION

In this paper, we propose a knowledge-based approach with semantic technologies (coupling ontologies, semantic constraints, and semantic rules) for geospatial data integration and visualization. The approach is used to solve a real-world geospatial data application—visualizing urban bicycling suitability—where data integration and visualization encounter complex and subtle domain semantics. In this context, we present a workflow for this interdisciplinary spatially informed study, including data integration, analysis, and visualization. We design several knowledge bases to cover all the aspects. Therefore, this approach is evaluated by the visualization results, which is a sink where semantics of the activities of data integration and processing are aggregated, interpreted, and visualized in a meaningful way [27].

Figure 8 is the visualization of the bicycling suitability (LTS) in the study area. The base map is a redistribution of OpenStreetMap (OSM) fed from the Mapbox API.¹³ In the visualization application, once a user clicks on an object, a pop-up could guide the user to explore further information in the faceted browser of RDF4J. Figure 9 shows the faceted browser with rich semantic information of a road link instance in the detailed level NVDB, and the inferred RDF statements (from ontological and rule-based reasoning) are also included, e.g., the relation of *isSymbolizedBy* that is deduced according to the ontologies and semantic rules.

It can be observed that with our approach, the cross-domain data integration and visualization are accomplished. With the constructed knowledge base for data analysis (LTS derivation) and visualization, essentially the client application (visualization tool) asks a question to the backend knowledge base (RDF4J in this case), “in this visualization context, what are the geospatial objects that should be rendered on the map and their visualization methods?”, the knowledge base will then provide the question with answers derived from the represented knowledge for data analysis and visualization. Furthermore, all the objects can be dereferenced, and more comprehensive information can be obtained (cf. Figure 9), which provides users with information beyond the graphic visualization. Such an application is difficult to develop with traditional web mapping techniques.

⁶<https://www.w3.org/TR/r2rml/>

⁷<https://ontop.inf.unibz.it/>

⁸<https://jena.apache.org/index.html>

⁹<https://github.com/TopQuadrant/shacl>

¹⁰<http://rdf4j.org/>

¹¹<https://www.djangoproject.com/>

¹²<https://leafletjs.com/>

¹³<https://www.mapbox.com/>



FIGURE 8. Visualization of LTS values on the map in the study area. The base map is a redistribution of OSM from Mapbox.

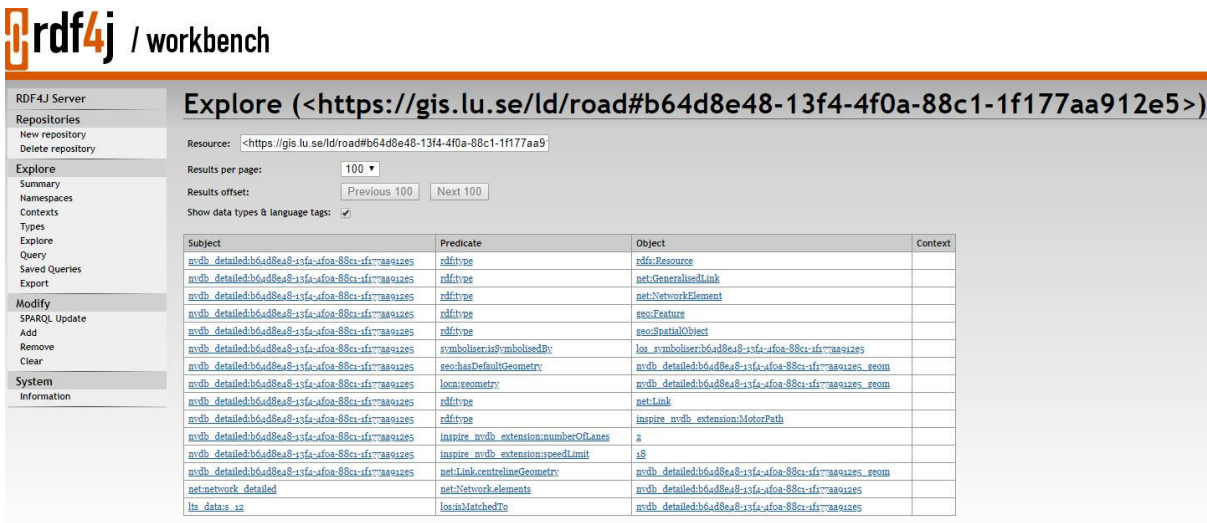


FIGURE 9. Information of a road link instance in the detailed level of road network in the RDF4J faceted browser.

IX. DISCUSSION

In this paper, we propose a knowledge-based approach for geospatial data integration and visualization using semantic technologies. We illustrate our approach in an interdisciplinary research application—visualizing urban bicycling suitability. In our study, we had many discussions between

traffic experts and geospatial experts. These discussions evidently unveiled that, in spite of massive use of geospatial information for decades in various areas, geospatial information still entails many intricacies for experts from other domains. Multiple representations of geospatial data seemed one of the most puzzling geospatial theories to

traffic researchers in this study. We initially planned to perform extensive discussions, so that either geospatial experts would understand the traffic theories and help them integrate the data and develop visualization products, or traffic researchers would (partially) grasp how the tasks should be accomplished. In this scenario, bespoke solutions could be developed, and, most likely, sufficient visualization products could be produced. Nevertheless, such a type of solutions has an intrinsic demerit: the knowledge communication emerging from the discussions and the theories embedded in the developed solutions (procedural codes) can hardly be transferred, interpreted, reused, and potentially expanded.

With our approach, the domain knowledge is formalized in a semantically-enriched and machine-readable manner. In principle, if one agrees with the modelled knowledge, the knowledge base can be readily used in relevant tasks (e.g., deriving LTS values and visualization for another study area, or deriving other types of LOS indexes) instead of domain experts having to sit down together each time or traffic experts having to consult geospatial literature to find appropriate methods. We argue that the knowledge-based approach would benefit the outreach and sharing of geospatial knowledge with a wider audience. This is in line with the research with regard to knowledge sharing using ontologies [47], whereas our approach enriches pure ontological approaches with semantic constraints and rules to cope with the complex semantic landscape in interdisciplinary studies. We believe our approach can be used in many other spatially informed studies in addition to the demonstrated case, as the approach offers a general framework for geospatial data integration and visualization with semantic technologies, particularly for handling multiple representations of geospatial data, which is an intricacy for data integration. The approach can also be used in other cross-domain and cross-detailed-level data integration tasks. For example, in spatiotemporal data integration, the events must be linked to the geospatial objects in a certain period of time, i.e., a geospatial object can have multiple (temporal) representations, and each corresponds to a certain period; the events should be linked to the corresponding representations in the time dimension. Another example is that during an emergency, the information of e.g., air pollution caused by fire is produced and should be linked to aggregated levels (e.g., county level), and some information is available at individual level (e.g., heritage building information); in order to analyze the affected heritage buildings during an emergency, cross-detailed-level data integration is necessary. In such cases, semantic constraints can be employed, as ontologies can hardly represent such restrictions.

Nevertheless, our approach also unveils several challenges. One prominent challenge is the modelling of knowledge, which is a demanding task. Generally, it is easier to train a domain expert with knowledge modelling (representation) than to equip a data scientist with domain knowledge [7]. This is in line with the extensively studied research topic in artificial intelligence, that is, knowledge elicitation (see e.g., [48]), which plays a significant role in expert system

development. A recent survey of geospatial expert systems demonstrated that the role of niched and standalone expert systems was downgraded, while the knowledge modelled for making integrated and complex spatial decisions clearly remains imperative [49]. The semantic landscape is increasingly complex, as more diverse sources of data are becoming available for geospatial analysis and visualization. Thus, the knowledge modelled for data integration and usage (visualization) will play a pivotal role to enhance the usability of geospatial information.

One may argue that quite a few semantic technologies are employed in our approach (e.g., ontology, semantic constraint, semantic rule, and linked data), which might be confusing for users. In fact, they are different types of knowledge representation paradigms that facilitate domain experts to formalize knowledge and thus make it explicit. Ontologies are the core in our approach, representing the essential conceptualizations of the domains. Built on the ontologies, semantic constraints and rules can be modelled, to represent more complex semantics, and to derive new facts (e.g., index values and visualization means). Therefore, once the employed ontologies have been decided upon, semantic constraints and rules can be readily developed by domain experts and grounded on the ontologies. In this way, domain experts are able to work with their own domain knowledge, rather than writing programs with procedure codes [20]. However, this approach does not apply to all applications due to the limited expressiveness of the knowledge representation paradigms of semantic technologies. There are certainly some analyses or visualization methods that cannot be formalized with ontologies and semantic rules. Nevertheless, it is possible to encapsulate a process that cannot be formalized with semantic technologies in, for example, a program, and semantically annotate its input and output to fit such processes into our knowledge-based approach. This method could be further investigated as a future work.

Another lesson learnt from this study is that the ontologies available for geospatial data have developed considerably, and the trend will most likely remain in the coming years. In this study, we reuse a number of state-of-the-art ontologies, e.g., cartographic scale ontology [17], data portrayal ontologies [20], and INSPIRE network ontology [33]. We acknowledge that such previous works provide solid ground for our work, and we also believe such ontology design works will benefit the outreach of geospatial information in the long run.

The contributions of this work can also be viewed from several perspectives. From a semantic web perspective, it has been long discussed and argued that as the open data has proliferated, the data available on the semantic web has increased dramatically in recent years, including geospatial data [50]. However, the representation of knowledge concerning how these data should be used is still sparse. This work advances the modelling of geospatial data and knowledge on the semantic web. The designed knowledge bases for geospatial data integration and visualization can be readily reused, and reached to wide audience. From the viewpoint of

geospatial semantics, this work sheds new insights into a new approach of representing geospatial knowledge for geospatial data integration and visualization. For data integration, this paper originally proposes to complement ontologies with semantic constraints to cope with complex and subtle semantic relations. With regard to data visualization, we propose a knowledge base with formalized geovisualization knowledge, which can generate visualizations from geospatial data in an automated manner.

X. CONCLUSIONS

This article proposes a knowledge-based approach for geospatial data integration and visualization with semantic technologies. Compared to other ontology-based approaches for (partially) accomplishing semantic interoperability for data integration, we reinforce the semantic bridge between the data from different domains using semantic constraints (SHACL constraints) to cope with complex semantic relations raised by multiple representations of geospatial data. In addition, we leverage semantic rules for modelling domain knowledge (analysis and visualization means) at different abstraction levels to enable machines to deduce the desired analysis results and visualization methods. The proposed framework is showcased and evaluated in a case study of visualizing urban bicycling suitability with the LTS index, in a study area in Lund, Sweden. The case study illustrates that the knowledge-based approach successfully overcomes semantic heterogeneity for cross-domain data integration with subtle and complex semantic relations. In addition, the knowledge modelled for data analysis as well as visualization effectively empowers machines to derive desired outcomes. This work provides a methodological framework for the sharing and outreach of geospatial data and knowledge to a wider audience for interdisciplinary spatially informed studies.

REFERENCES

- [1] W. Kuhn, "Core concepts of spatial information for transdisciplinary research," *Int. J. Geograph. Inf. Sci.*, vol. 26, no. 12, pp. 2267–2276, Dec. 2012.
- [2] K. Janowicz, "The role of space and time for knowledge organization on the semantic Web," *Semantic Web*, vol. 1, no. 1, pp. 25–32, 2010.
- [3] K. Janowicz, F. Van Harmelen, J. A. Hendler, and P. Hitzler, "Why the data train needs semantic rails," *AI Mag.*, vol. 36, no. 1, pp. 5–14, 2014.
- [4] M. Lenzerini, "Data integration: A theoretical perspective," in *Proc. 21st ACM SIGMOD-SIGACT-SIGART Symp. Princ. Database Syst.*, 2002, pp. 233–246.
- [5] K. Janowicz, S. Scheider, T. Pehle, and G. Hart, "Geospatial semantics and linked spatiotemporal data—Past, present, and future," *Semantic Web*, vol. 3, no. 4, pp. 321–332, 2012.
- [6] H. T. Uitermark, P. J. M. van Oosterom, N. J. I. Mars, and M. Molenaar, "Ontology-based geographic data set integration," in *Proc. Int. Workshop Spatio-Temporal Database Manage.*, 1999, pp. 60–78.
- [7] Y. Chen, S. Sabri, A. Rajabifard, and M. E. Agunbiade, "An ontology-based spatial data harmonisation for urban analytics," *Comput., Environ. Urban Syst.*, vol. 72, pp. 177–190, Nov. 2018.
- [8] Y. Shu, D. Ratcliffe, M. Compton, G. Squire, and K. Taylor, "A semantic approach to data translation: A case study of environmental observations data," *Knowl.-Based Syst.*, vol. 75, pp. 104–123, Feb. 2015.
- [9] M. Lutz, J. Sprado, E. Klien, C. Schubert, and I. Christ, "Overcoming semantic heterogeneity in spatial data infrastructures," *Comput. Geosci.*, vol. 35, no. 4, pp. 739–752, Apr. 2009.
- [10] L. van den Brink, P. Janssen, W. Quak, and J. Stoter, "Towards a high level of semantic harmonisation in the geospatial domain," *Comput., Environ. Urban Syst.*, vol. 62, pp. 233–242, Mar. 2017.
- [11] S. Volz, "Data-driven matching of geospatial schemas," in *Proc. Int. Conf. Spatial Inf. Theory*, 2005, pp. 115–132.
- [12] J. R. Pucher and R. Buehler, *City Cycling*, vol. 11. Cambridge, MA, USA: MIT Press, 2012.
- [13] H. C. Manual, *Highway Capacity Manual*, vol. 2. Washington, DC, USA: Transportation Research Board, 2000.
- [14] M. C. Mekuria, P. G. Furth, and H. Nixon, "Low-stress bicycling and network connectivity," *Mineta Transp. Inst.*, San Jose, CA, USA, MTI Rep. 11-19, May 2012.
- [15] L. Petterson. (2012). Regler för insamling och leverans av vägdata. Trafikverket, Borlänge, Sweden. [Online]. Available: <http://www.nvdb.se/globalassets/upload/styrande-och-vagledande-dokument/tdok-2013-0381-regler-for-insamling-och-leverans-ver-9.pdf>
- [16] R. Pritchard, Y. Frøyen, and B. Snizek, "Bicycle level of service for route choice—A GIS evaluation of four existing indicators with empirical data," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 5, p. 214, 2019.
- [17] D. Carral, S. Scheider, K. Janowicz, C. Vardeman, A. A. Krisnadhi, and P. Hitzler, "An ontology design pattern for cartographic map scaling," in *Proc. Extended Semantic Web Conf.*, 2013, pp. 76–93.
- [18] L. Harrie and R. Weibel, "Modelling the overall process of generalisation," in *Generalisation of Geographic Information: Cartographic Modelling and Applications*. Amsterdam, The Netherlands: Elsevier, 2007, pp. 67–87.
- [19] S. Scheider and M. D. Huisjes, "Distinguishing extensive and intensive properties for meaningful geocomputation and mapping," *Int. J. Geograph. Inf. Sci.*, vol. 33, no. 1, pp. 28–54, Jan. 2019.
- [20] W. Huang and L. Harrie, "Towards knowledge-based geovisualisation using semantic Web technologies: A knowledge representation approach coupling ontologies and rules," *Int. J. Digit. Earth*, pp. 1–22, Apr. 2019.
- [21] W. Huang, A. Mansourian, E. Abdolmajidi, H. Xu, and L. Harrie, "Synchronising geometric representations for map mashups using relative positioning and linked data," *Int. J. Geograph. Inf. Sci.*, vol. 32, no. 6, pp. 1117–1137, Jun. 2018.
- [22] D. Callister and M. Lowry, "Tools and strategies for wide-scale bicycle level-of-service analysis," *J. Urban Planning Develop.*, vol. 139, no. 4, pp. 250–257, Dec. 2013.
- [23] S. Wiemann and L. Bernard, "Spatial data fusion in spatial data infrastructures using linked data," *Int. J. Geograph. Inf. Sci.*, vol. 30, no. 4, pp. 613–636, Apr. 2016.
- [24] W. Huang, S. A. Raza, O. Mirzov, and L. Harrie, "Assessment and benchmarking of spatially enabled RDF stores for the next generation of spatial data infrastructure," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 7, p. 310, 2019.
- [25] W. Kuhn, "Geospatial semantics: Why, of what, and how?" in *Journal on Data Semantics III*. Berlin, Germany: Springer, 2005, pp. 1–24.
- [26] B. Bucher, E. Tiainen, T. E. von Brasch, P. Janssen, D. Kotzinos, M. Čeh, M. Rijdsdijk, E. Folmer, M.-D. Van Damme, and M. Zhral, "Conciliating perspectives from mapping agencies and Web of data on successful European SDIs: Toward a European geographic knowledge graph," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 2, p. 62, 2020.
- [27] K. Janowicz, S. Schade, A. Bröring, C. Keßler, P. Maué, and C. Stasch, "Semantic enablement for spatial data infrastructures," *Trans. GIS*, vol. 14, no. 2, pp. 111–129, 2010.
- [28] B. Hofer, S. Mäs, J. Brauner, and L. Bernard, "Towards a knowledge base to support geoprocessing workflow development," *Int. J. Geograph. Inf. Sci.*, vol. 31, no. 4, pp. 694–716, Apr. 2017.
- [29] S. Scheider, A. Ballatore, and R. Lemmens, "Finding and sharing GIS methods based on the questions they answer," *Int. J. Digit. Earth*, vol. 12, no. 5, pp. 594–613, May 2019.
- [30] F. Baader, D. Calvanese, D. McGuinness, P. Patel-Schneider, and D. Nardi, *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [31] N. F. Noy and D. L. McGuinness, "Ontology development 101: A guide to creating your first ontology," *Stanford Knowl. Syst. Lab.*, Stanford, CA, USA, Tech. Rep. KSL-01-05, 2001.
- [32] European Commission. (2017). *Guidelines for the RDF Encoding of Spatial Data, Technical Specification*. [Online]. Available: <http://inspire-eu-rdf.github.io/inspire-rdf-guidelines>
- [33] Y. Shu, "A practical approach to modelling and validating integrity constraints in the semantic Web," *Knowl.-Based Syst.*, vol. 153, pp. 29–39, Aug. 2018.

- [34] F. Manola, E. Miller, and B. McBride. (2004). RDF Primer. W3C Recommendation. [Online]. Available: <https://www.w3.org/TR/rdf-primer/>
- [35] C. B. Jones, D. B. Kidner, L. Q. Luo, G. L. Bundy, and J. M. Ware, "Database design for a multi-scale spatial information system," *Int. J. Geograph. Inf. Syst.*, vol. 10, no. 8, pp. 901–920, Dec. 1996.
- [36] M. Perry and J. Herring. (Sep. 2012). OGC GeoSPARQL—A Geographic Query Language for RDF Data. OGC Implementation Standard. [Online]. Available: https://portal.opengeospatial.org/files/?artifact_id=47664
- [37] Z. Zhang, "Towards efficient and effective semantic table interpretation," in *Proc. Int. Semantic Web Conf.*, 2014, pp. 487–502.
- [38] M. F. Lopez, A. Gomez-Perez, and N. Juristo, "Methontology: From ontological art towards ontological engineering," in *Proc. Ontol. Eng. AAAI Spring Symp. Ser.* Menlo Park, CA, USA: American Association for Artificial Intelligence, Mar. 1997, pp. 1–8.
- [39] H. Knublauch. *SHACL and OWL Compared*. Accessed: Feb. 24, 2019. [Online]. Available: <http://spinrdf.org/shacl-and-owl.html>
- [40] H. Knublauch and A. Ryman. (2017). Shapes Constraint Language (SHACL). W3C Recommendation. [Online]. Available: <https://www.w3.org/TR/shacl/>
- [41] S. Steyskal and K. Coyle. (2017). SHACL Use Cases and Requirements. W3c Working. [Online]. Available: <https://www.w3.org/TR/shacl-ucr/>
- [42] N. Bassiliades, "SWRL2SPIN: A tool for transforming SWRL rule bases in OWL ontologies to object-oriented SPIN rules," 2018, *arXiv:1801.09061*. [Online]. Available: <http://arxiv.org/abs/1801.09061>
- [43] H. Knublauch. (2011). Spin-Modeling Vocabulary. W3C Member Submission. [Online]. Available: <https://www.w3.org/Submission/spin-modeling/>
- [44] R. A. Smith, "Designing a cartographic ontology for use with expert systems," in *Proc. Special Joint Symp. ISPRS Tech. Commission IV AutoCarto Conjunction (ASPRS/CaGIS)*, 2010, pp. 1–6.
- [45] J. Brus, D. Zdena, J. Kanok, and V. Pechanec, "Design of intelligent system in cartography," in *Proc. 9th RoEduNet IEEE Int. Conf.*, Jun. 2010, pp. 112–117.
- [46] C. Keßler, M. Raubal, and C. Wosniok, "Semantic rules for context-aware geographical information retrieval," in *Proc. Eur. Conf. Smart Sens. Context*, 2009, pp. 77–92.
- [47] M. Jelokhani-Niaraki, "Knowledge sharing in Web-based collaborative multicriteria spatial decision analysis: An ontology-based multi-agent approach," *Comput., Environ. Urban Syst.*, vol. 72, pp. 104–123, Nov. 2018.
- [48] N. J. Cooke, "Varieties of knowledge elicitation techniques," *Int. J. Hum.-Comput. Stud.*, vol. 41, no. 6, pp. 801–849, Dec. 1994.
- [49] D. Demetriou, "A review of spatial expert systems: Do they still have a role to play?" in *Proc. 6th Int. Conf. Remote Sens. Geoinformation Environ. (RSCy)*, Aug. 2018, Art. no. 107730H.
- [50] L. van den Brink, P. Barnaghi, J. Tandy, G. Atezing, R. Atkinson, B. Cochrane, Y. Fathy, R. G. Castro, A. Haller, A. Harth, K. Janowicz, S. Kolozali, B. van Leeuwen, M. Lefrançois, J. Lieberman, A. Perego, D. Le-Phuoc, B. Roberts, K. Taylor, and R. Troncy, "Best practices for publishing, retrieving, and using spatial data on the Web," *Semantic Web*, vol. 10, no. 1, pp. 95–114, Dec. 2018.



WEIMING HUANG was born in Jinan, China, in 1990. He received the Ph.D. degree in geographical information science from Lund University, Sweden, in January 2020. During autumn 2019, he was a Visiting Researcher with the Center for Spatial Studies, University of California at Santa Barbara. He is currently a Researcher with Lund University. His Ph.D. research topic was knowledge-based geospatial data integration and visualization with semantic web technologies.



KHASHAYAR KAZEMZADEH received the B.Sc. degree in civil engineering, in 2011, and the M.Sc. degree in transportation engineering, in 2016. He is currently pursuing the Ph.D. degree in transportation with the Division of Transport and Roads, Department of Technology and Society, Faculty of Engineering, Lund University, Lund, Sweden. He is also a Teaching Assistant in the traffic engineering course with Lund University. His research interests include traffic flow, travel behaviour, and traffic safety. His Ph.D. research is mainly about how cyclists perceive comfort in mixed traffic with the focus on electric bikes.



ALI MANSOURIAN received the Ph.D. degree in geomatics engineering, in 2005. He is currently an Associate Professor with the Department of Physical Geography and Ecosystem Science, Lund University, Sweden. He is also a council member of Association of Geographic Information Laboratories in Europe (AGILE), from 2015 to 2023. He has been coordinating several international and national projects related to these topics. His main research interests are spatial data infrastructures (SDI), geospatial web services, semantic web, and spatial and temporal modeling using artificial intelligence techniques with applications in spatial epidemiology and health, disaster management, and urban planning.



LARS HARRIE was born in Gnosjö, Sweden, in 1968. He received the M.Sc. degree in geodesy from KTH, Stockholm, Sweden, in 1993, and the Ph.D. degree in geographic information science from Lund University, in 2001. Since 2006, he has been working with the Department of Physical Geography and Ecosystem Science and the GIS Centre, Lund University. During recent years, he has worked in the field of digitalization of the built environment processes. Since 2017, he has been a Professor in geomatics. His main research interests are algorithms for processing and presenting geographic data, and methods for improving the national spatial data infrastructure.

...