Title
A Prototypical Geospatial Knowledge Graph And Spatio-Temporal Question Answering for Supply Chain Visibility

Authors
Dopler, Silvia
Scholz, Johannes

Publication Date
2021-12-01

DOI
10.25436/E2JS3V

Peer reviewed

A Prototypical Geospatial Knowledge Graph And Spatio-Temporal Question Answering for Supply Chain Visibility

Silvia Dopler[1][0000−0002−7843−0372] and Johannes Scholz[2][0000−0002−3212−8864]

[1] University of Applied Sciences Upper Austria, School of Business and Management, Logistikum, Wehrgrabengasse 1-3, 4400 Steyr, Austria
silvia.dopler@fh-steyr.at
[2] Graz University of Technology, Institute of Geodesy, Research Group Geoinformation, Steyrergasse 30/I, 8010 Graz, Austria
johannes.scholz@tugraz.at

Abstract. This paper elaborates on the geospatial semantics of Supply Chain Visibility (SCV) and how a Geospatial Knowledge Graph (GeoKG) helps answering spatio-temporal questions related to the supply chain. An ontology is developed to provide a generic semantic model for the SCV domain, based on existing ontologies and ontology design patterns. Secondly, an early-stage GeoKG is generated in a GeoSPARQL-enabled semantic triplestore to demonstrate the spatio-temporal querying capability. The focus of this work is on the semantic representation of spatio-temporal trajectories of material flow as this is the central aspect of SCV.

Keywords: geospatial knowledge graph, geospatial semantics, spatio-temporal trajectory, supply chain visibility

DOI: https://doi.org/10.25436/E2JS3V

1 Introduction

Spatial and temporal information about objects moving along supply chain networks (SCN) are essential parts of the definition of supply chain visibility (SCV) in [4]. A SCN is defined as "a network of connected and interdependent organisations mutually and co-operatively working together to control, manage and improve the flow of materials and information from suppliers to end users" [1]. Supply chain visibility (SCV) denotes the ability of a focal firm "to 'see' from one end of the supply chain to the other" [7]. More precisely, [4] defines the elements of SCV that we use as reference point of our study: "identity, location and status of entities transiting the supply chain, captured in timely messages about events, along with the planned and actual dates/times for these events." In the SCV context entities can be any objects (e.g. product, collection of orders, container, vessel, etc.) travelling along a SCN. Location refers to the position of each entity in the SCN and can be either static or dynamic [4] – e.g. as stock in a factory waiting for an order or loaded onto a truck on the way to the customer. These general considerations are consistent with the stop-move conceptualization of semantic trajectories [8]. Among other things, product flows, e.g. shipping of customer orders, are to be recorded and tracked in order to improve performance of the SCN under review [7]. In contemporary SCNs, various stakeholders with different data storage and applications collect data. In order to share the collected information in the SCN, data requires to be integrated overcoming institutional borders. Additionally, SCV applications typically require fast and reliable system architectures for the processing of (near) real-time data that evolve over time in their backends, need to be flexible enough to share data among participating stakeholders for different enterprise systems and fast enough in terms of reasoning over paths. These requirements can be met with the help of Semantic Web technologies [10, 9, 15] and graph databases [2]. Linked Data approaches, open standards and graph-based representation of data (i.e. RDF-Schema and OWL) can reduce efforts for information integration and sharing alike, while the value of the information increases [16]. This is due to the attribution with universally unique identifier systems (URI/IRI) and semantically enriched data which provide the basis for inferencing capabilities.

According to [6] a knowledge graph (KG) is a data graph, which is potentially enhanced with information of the intrinsic semantics, e.g. representations of a semantic, validating and/or emergent schema, identity, context and/or rules. In this paper we cover aspects of semantic schemas only which are represented by the World Wide Web Consortium (W3C) standards RDF-Schema (RDFS[3]) and the Web Ontology Language (OWL[4]) . Together they basically allow for defining classes, properties and axioms for RDF graphs. This information may be embedded into the data graph to enable inference for data enrichment by deductive knowledge [6]. If geographic information (GI) is modelled within the KG, we refer to it as GeoKG [17, 3].

We strive to emphasize the spatio-temporal dimension in the ontology and contribute to the spatial data science community by modelling a spatio-temporal supply chain, utilizing existing ontologies (or parts thereof) and ontology design patterns. This is of particular importance as spatially explicit models are preliminaries for GeoAI applications [9]. Hence, the paper applies GeoKG and geographic question answering with the objective to support SCV decisions in the future, based on spatio-temporal data and semantics.

The remainder of the paper is organized as follows. Section 2 describes the inductive methodology of ontology engineering and GeoKG development Section 3 presents the preliminary results of the work in progress and demonstrates how spatio-temporal information of a moving object in the SCN may be queried from the early-stage GeoKG. Section 4 discusses the results achieved and presents a critical outlook.

[3] https://www.w3.org/TR/rdf-schema/
[4] https://www.w3.org/TR/owl2-overview/

2 Methodology

The methodology followed in this paper is centered around the development of a domain ontology for SCV, based on existing work. This ontology is later on utilized in a GeoKG, that is populated with a test data set. Competency questions show the capability to support spatio-temporal question answering and reveal the functionalities and limitations thereof.

The requirement for the ontology is that it be kept as generic as possible for the SCV domain. For the authoring process we used the method of ontology engineering [5] and followed five steps: 1) defining scope and purpose, 2) describing use cases and competency questions that will help testing if the scope and purpose are met by the ontology, 3) building a list of terms that build the necessary vocabulary or lexicon of the ontology. This step involves the identification of existing vocabularies or taxonomies for re-use. We identified GeoSPARQL[5] and OWL-Time[6] to be appropriate in order to map the spatio-temporal aspects of the domain. For provenance we used DC terms[7] and SKOS[8]. The following step 4) includes writing informal explanations for all terms in natural language as a glossary. And lastly 5), the conceptualization step, where the formal OWL/RDFS ontology is developed. We made use of Protégé Desktop version 5.5.0 [9] and the plugin HermiT 1.4.3.456 reasoner. The namespaces and prefixes that are used for the ontology are depicted in fig. 1.

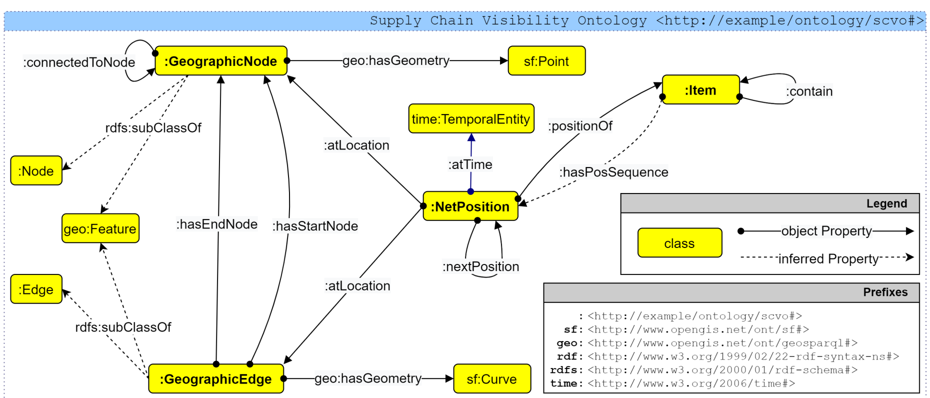

Fig. 1. Excerpt and graphical representation of Supply Chain Visibility Ontology

[5] http://www.opengis.net/doc/IS/geosparql/1.0
[6] https://www.w3.org/TR/2020/CR-owl-time-20200326/
[7] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
[8] https://www.w3.org/TR/2009/REC-skos-reference-20090818/
[9] https://protege.stanford.edu/

For constructing the GeoKG the authors utilized GraphDB Free[10], a semantic triplestore. Synthetic test data describing a supply chain are used for test purposes. The elements of the test data are mapped to RDF and lifted by the SCV ontology using the built-in RDFizer tool OntoRefine in GraphDB. The repository uses the OWL-Horst entailment ruleset for consistency checks and inference. The ontology is imported to the default graph of the graph dataset and data are imported to named graphs. To test and validate the resulting GeoKG the spatio-temporal competency questions serve as a natural language basis to code SPARQL queries.

3 Preliminary Results

An excerpt of the ontology is depicted in fig. 1. Since network elements are the carriers of the moving objects, the ontology models the basic structure of SCN (denoted as classes `:Node` and `:Edge`) and the moving entities (called `:Item`). To determine the trajectories, the spatio-temporal components of the item steps are determined via their recorded traces (denoted as `:NetPosition`). Depending on the required level of detail, the exact position on the transport network is recorded (e.g. via track & trace technology) or an association to a SCN element (e.g. Item enters/exits `:Node` or `:Edge`). The class `:NetPosition` acts as a mediator for the n-ary relations [14] that an `:Item` has to its location at a certain time stamp or interval (via `:atTime`). From this follows that data of class `:Item` have an indirect spatial dimension via the `:atLocation` relation from associated `:NetPosition`. The class `:Node` represents the stakeholders and physical locations involved in a SCN and are linked via objects of class `:Edge`. The latter may represent abstract connections, like arbitrary business relations, or transport routes, e.g. from raw material supplier to production site. Considering that not every `:Node` (stakeholder of a SCN) necessarily needs to be a carrier of an `:Item` (e.g. the office of a logistics service provider), we introduced a defined class `:GeographicNode`, that is a subclass of `:Node` and has a `geo:hasGeometry` relation to some `sf:Point` object. Accordingly, an `:Edge` that has a geographic extent is defined in the sublass `:GeographicEdge` and is linked to some `sf:Curve` geometry. Having a `geo:hasGeometry` relation makes `:GeographicNode` and `:GeographicEdge` automatically a `geo:Feature` of the GeoSPARQL ontology and allows distinction between spatial and non-spatial objects in the resulting GeoKG.

To test and validate the resulting GeoKG the competency questions defined during the ontology engineering process serve as a natural language basis to code GeoSPARQL queries. We anticipated the following spatio-temporal competency questions (QC) that need to be answered and help making informed decisions:

CQ 1 What is the spatio-temporal trajectory of an item?
CQ 2 What is the mean total travel distance and time of an item?
CQ 3 At what time did a certain item reside in a certain SCN position or area?

[10] https://graphdb.ontotext.com/documentation/free/

Listing 1.1. CQ 1 SPARQL query

```
PREFIX scvo: <http://example/ontology/scvo#>
PREFIX items: <http://example/base/items>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?item ?pos ?location ?time
WHERE { FILTER (?item = items:product5)
    ?item scvo:hasPosSequence|scvo:hasPosSequence/
        scvo:nextPosition+ ?pos .
    ?pos scvo:atLocation ?location .
    ?pos scvo:atTime ?time .
} order by ?item ?time
```

We carried out a test run and constructed the early-stage GeoKG. The test data containing 2,728 rows and a varying number of attribute columns have been mapped to RDF/Turtle syntax and lifted using the concepts of the ontology. To complement the mapping & lifting process additional data manipulation and a SPARQL INSERT query was conducted to implement the ontologically defined connection between items and their network positions (`:NetPosition`, see fig. 1). Each `:NetPosition` is associated with an `:Item` via `:positionOf` as well as with a `:Node` or `:Edge` via the `:atLocation` relationship. The resulting GeoKG holds 87,445 triples, where 23.1 % (20,242) have been explicitly inserted and 76.9 % (67,203) are inferred by the OWL-Horst ruleset of the repository and the defined axioms of the ontology. To demonstrate the early-stage capability of the GeoKG to answer spatial questions, we use a query that is associated to CQ 1. The query in listing 1.1 searches the KG for all positions of a certain item and their spatial and temporal properties. A visual representation of the spatio-temporal track is depicted in fig. 2.

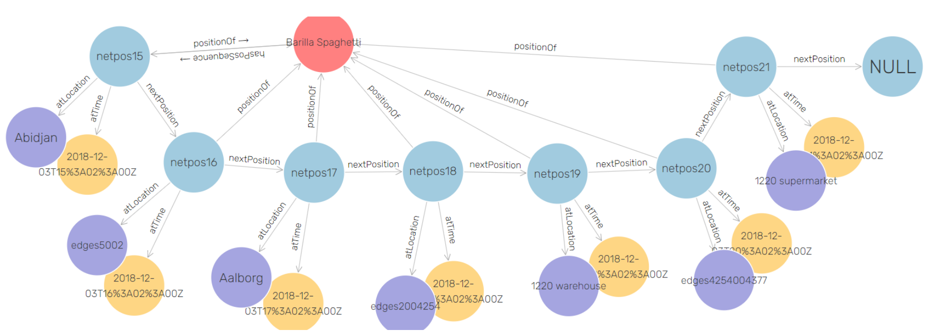

Fig. 2. Item's (red) graph view of its spatio (violet) -temporal (yellow) trajectory (blue)

4 Discussion and Outlook

The paper presents an early stage spatio-temporal ontology excerpt and GeoKG for enhancing SCV for physical objects. Information and monetary flows were not considered here, but indicate a direction for further development. The GeoKG is implemented in GraphDB Free and tested with a synthetic test data set. In order to evaluate the capabilities to answer spatio-temporal queries related to a supply chain, we used competency questions defined by the authors. The results of CQ1 SPARQL Query (listing 1.1) reveal that the spatio-temporal history of the token can be extracted from the GeoKG. With regard to the applicability of the GeoKG, further efforts need to be made towards ensuring data consistency and quality, e.g. introducing further ontology axioms, using Shapes Constraint Language (SHACL) or Shape Expressions (ShEx). Future research questions include the spatio-temporal aggregation of resulting data sets, based on geographic and semantic properties (see e.g. [17, 11]) as indicated by CQ 2 and 3. Further ontology design patterns concerning semantic trajectories in the spatio-temporal context ([8, 13]) and the supply chain domain have to be looked at and checked for their content, whether they represent or contain suitable complements or equivalent classes compared to the SCV ontology presented here. Moreover spatio-temporal semantic reasoning based on a GeoKG is of interest to support comprehensible decision making processes especially for the application field SCV. The market of commercial SCV platforms just emerges, as identified by the market research company Gartner, but still lacks prescriptive analytical capabilities [12] where exchange with representatives from geoscience and industry would be useful and desirable. Continuing efforts could help reduce the carbon footprint of global supply chains, as the GeoKG enables the collaboration between stakeholders of supply chains - in order to minimize transport distances and to utilize transport resources more efficiently.

References

1. Aitken, J.M.: Supply Chain Integration Within the Context of a Supplier Association: Case Studies of Four Supplier Associations. Phd thesis, Cranfield University School of Management, Cranfield (1998), http://dspace.lib.cranfield.ac.uk/handle/1826/9990
2. Arenas, M., Gutierrez, C., Sequeda, J.F.: Querying in the age of graph databases and knowledge graphs: Last revised 25 june 2021, http://arxiv.org/pdf/2106.11456v2
3. Ballatore, A., Bertolotto, M., Wilson, D.C.: Geographic knowledge extraction and semantic similarity in OpenStreetMap. Knowledge and Information Systems **37**(1), 61–81 (oct 2012). https://doi.org/10.1007/s10115-012-0571-0
4. Francis, V.: Supply chain visibility: lost in translation? Supply Chain Management: An International Journal **13**(3), 180–184 (2008). https://doi.org/10.1108/13598540810871226
5. Hart, G., Dolbear, C.: Linked Data: A Geographic Perspective. CRC Press, Hoboken (2013), http://gbv.eblib.com/patron/FullRecord.aspx?p=1114012

6. Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G.d., Gutierrez, C., Gayo, J.E.L., Kirrane, S., Neumaier, S., Polleres, A., Navigli, R., Ngomo, A.C.N., Rashid, S.M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., Zimmermann, A.: Knowledge graphs, http://arxiv.org/pdf/2003.02320v3
7. Holcomb, M.C., Ponomarov, S.Y., Manrodt, K.B.: The relationship of supply chain visibility to firm performance. Supply Chain Forum: An International Journal **12**(2), 32–45 (2011). https://doi.org/10.1080/16258312.2011.11517258
8. Hu, Y., Janowicz, K., Carral, D., Scheider, S., Kuhn, W., Berg-Cross, G., Hitzler, P., Dean, M., Kolas, D.: A geo-ontology design pattern for semantic trajectories. In: Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M.Y., Weikum, G., Tenbrink, T., Stell, J., Galton, A., Wood, Z. (eds.) Spatial Information Theory, Lecture Notes in Computer Science, vol. 8116, pp. 438–456. Springer International Publishing, Cham (2013). https://doi.org/10.1007/978-3-319-01790-7_24
9. Janowicz, K., Gao, S., McKenzie, G., Hu, Y., Bhaduri, B.: Geoai: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. International Journal of Geographical Information Science **34**(4), 625–636 (2020). https://doi.org/10.1080/13658816.2019.1684500
10. Janowicz, K., Scheider, S., Adams, B.: A geo-semantics flyby. In: Reasoning web. Semantic technologies for Intelligent Data Access, pp. 230–250. Springer (2013)
11. Mai, G., Janowicz, K., Cai, L., Zhu, R., Regalia, B., Yan, B., Shi, M., Lao, N.: Se-kge: A location-aware knowledge graph embedding model for geographic question answering and spatial semantic lifting. Transactions in GIS **24**(3), 623–655 (2020)
12. de Muynck, B., West, C.: Magic quadrant for real-time transportation visibility platforms: Updated 19 october 2021, published 14 april 2021 (2021), https://www.gartner.com/doc/reprints?id$\bar{1}$-27LUWGAU&ct$\bar{2}$11007&st$\bar{s}$b&utm_campaign$\bar{2}$021-Q2
13. Nogueira, T.P., Braga, R.B., de Oliveira, C.T., Martin, H.: Framestep: A framework for annotating semantic trajectories based on episodes. Expert Systems with Applications **92**, 533–545 (2018). https://doi.org/10.1016/j.eswa.2017.10.004
14. Noy, N., Rector, A.: Defining n-ary relations on the semantic web: W3c working group note 12 april 2006 (2006), http://www.w3.org/TR/2006/NOTE-swbp-n-aryRelations-20060412/
15. Paulheim, H.: Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic web **8**(3), 489–508 (2017)
16. Pellegrini, T., Sack, H., Auer, S.: Linked Enterprise Data. Springer Berlin Heidelberg, Berlin, Heidelberg (2014). https://doi.org/10.1007/978-3-642-30274-9
17. Yan, B., Janowicz, K., Mai, G., Zhu, R.: A spatially explicit reinforcement learning model for geographic knowledge graph summarization. Transactions in GIS **23**(3), 620–640 (jun 2019). https://doi.org/10.1111/tgis.12547