[DM-01-092] Geospatial Knowledge Graphs

Abstract

Geospatial Knowledge Graphs (GeoKGs) organize geospatial data and knowledge into graph structures, in which entities like places and events serve as nodes and their relationships form the edges. They are complemented with expressive metadata in the form of ontologies defining concepts (classes) and their relationships (properties). This structure underpins the powerful capabilities of GeoKGs in addressing challenges such as data integration, retrieval, and knowledge formalization. This entry first introduces the fundamentals of knowledge graphs, focusing on their implementation via Semantic Web technologies. It then explores GeoKGs, covering their advantages, relevant techniques, prominent examples, and a few key application areas. The entry concludes with an outlook on emerging trends, underscoring the convergence of machine learning and GeoKGs as a promising avenue for Geospatial Artificial Intelligence (GeoAI).

Keywords: ontology

Author & citation

Huang, W. and Zhu, R. (2025). Geospatial Knowledge Graphs. The Geographic Information Science & Technology Body of Knowledge (Issue 2, 2025 Edition), John P. Wilson (ed.). DOI: 10.22224/gistbok/2025.2.8.

Explanation

1. Background of Knowledge Graphs
2. Introduction to Geospatial Knowledge Graphs
3. Technical Development of GeoKGs
4. Prominent Examples of GeoKGs
5. Applications of GeoKGs
6. Emerging Trends

1. Background of Knowledge Graphs

Knowledge graphs (KGs) are an important instrument for data organization, management, and integration. A key driver for the adoption of KGs in both industry and academia is their powerful capability for data integration, which is a challenge that remains ubiquitous and longstanding across various domains. KGs are composed of entities and relationships. Entities can be concrete real-world objects like people and places or more abstract concepts such as events and organizations. Relationships are also typed and carry semantics. The fundamental units of a KG are triples, in the form of <head entity, relationship, tail entity>. A concrete example of a triple is <Eiffel Tower, within, Paris>. This entry focuses on KGs realized with Semantic Web technologies, while in principle the graph structure could be implemented in other ways and does not prescribe a specific technology stack.

KGs are underpinned by a set of methods, tools, and standards set out from the Semantic

Web research (Hitzler, 2021). Ontologies can be understood as the schema for KGs. Unlike schemas for relational databases, they are based on formal logic. This makes ontologies explicit in defining knowledge, which helps to reduce ambiguity and enables the automatic deduction of implicit knowledge (Guarino et al., 2009). If we have both the triples <Eiffel Tower, within, Paris>, and <Paris, capitalOf, France>, with an ontology and appropriate formal rules, the implicit knowledge < Eiffel Tower, within, France> can be deduced, even if this was not explicitly stated. Ontologies act as the primary catalyst for data integration in KGs because they are designed to be shared and reused (Hitzler, 2021). This allows multiple data sources to be integrated through a common understanding of concepts and their relationships, e.g., ensuring that the concept “Building” carries the same meaning across different datasets.

Triples in KGs are represented using the Resource Description Framework (RDF, www.w3.org/RDF/), a foundational data model and standard recommended by the World Wide Web Consortium (W3C) for data interchange on the Web. Within the RDF framework, a key mechanism for data integration is the use of Uniform Resource Identifiers (URIs). Each resource in a KG, be it a concept or relationship in its underlying ontology, or an entity (data instance) described according to that ontology, is assigned a unique URI. This unique identification allows KGs to be integrated when the same URIs are used to denote identical resources across different graphs, thereby linking them. Conceptually, such interlinked KGs form a larger, integrated KG (Hitzler, 2021).

The Semantic Web technology stack includes several other key standards and techniques crucial for the construction and utilization of KGs. SPARQL (https://www.w3.org/TR/sparql11-query/) serves as the standard query language for RDF-based KGs, acting as the primary mechanism for data retrieval, which is analogous to SQL's role in relational databases. SPARQL is widely implemented in RDF stores (also known as triple stores), which are specialized database management systems designed to store and manage RDF KGs. The implementation of ontologies that underpin KGs relies on standardized languages like the Web Ontology Language (OWL, https://www.w3.org/OWL/) and RDF Schema (RDFS, https://www.w3.org/TR/rdf-schema/). OWL and RDFS are widely supported by various ontology reasoners to deduce implicit knowledge. Furthermore, the W3C recommends the Shapes Constraint Language (SHACL, https://www.w3.org/TR/shacl/) for validating KG structure and quality, and for enriching KGs through customized, rule-based inferences.

2. Introduction to Geospatial Knowledge Graphs

This entry adopts a narrow, geospatial-centric definition of Geospatial Knowledge Graphs (GeoKGs), as it is challenging to find a KG entirely devoid of geospatial information, given that most real-world entities and events are inherently situated in space and time. To be considered a GeoKG in this narrow definition, a KG should contain explicit geospatial references, such as geographic coordinates, place names, or well-defined geometries. The primary purpose of the graph is to model how entities relate to each other spatially (and sometimes temporally).

The most important driver for constructing GeoKGs is the persistent need to integrate geospatial data from diverse sources. Traditional methods for geospatial data organization,

representation, and access, such as individual shapefiles or relational databases, often fall short in establishing meaningful and semantically rich linkages between these disparate datasets. This leads to isolated data silos where information lacks clear semantic connections and is difficult to discover or reuse effectively (Janowicz et al., 2022). This considerably hinders the utilization of geospatial data, as data integration is a prerequisite for most geospatial analyses. For example, effective disaster response in a wildfire event requires the integration of multi-source data, e.g., demographic information, real-time environmental data, and critical infrastructure details (e.g., roads, hospitals), for effective situation assessment and humanitarian relief (Zhu et al., 2021). In this context, GeoKGs provide a powerful and flexible infrastructure (graph structure) for integrating geospatial data from diverse sources and for connecting geospatial with non-geospatial data (e.g., by linking to general-purpose KGs like DBpedia (https://www.dbpedia.org/) and Wikidata (https://www.wikidata.org/), which are two representative KGs constructed from Wikipedia).

Besides data integration, the use of GeoKGs has also been motivated by other factors. A key driver is knowledge formalization, which involves making informal or implicit geospatial knowledge (e.g., procedural knowledge for composing geoprocessing workflows or for online map design) explicit, and understandable for both humans and machines in an unambiguous manner. Moreover, GeoKGs are increasingly recognized for facilitating geospatial data to adhere to the FAIR (Findable, Accessible, Interoperable, and Reusable) principles (Ma, 2022). In this regard, GeoKGs inherently support interoperability by formal and shared ontologies and promote reusability through rich and structured metadata. They also significantly improve the findability and accessibility of geospatial data, e.g., using URIs and established links between multiple (Geo)KGs.

3. Technical Developments for GeoKGs

The growing popularity of GeoKGs has driven technical advancements in their construction and utilization. A prominent development is the OGC standard GeoSPARQL (https://www.ogc.org/standards/geosparql/), which extends SPARQL and provides a lightweight ontology for representing and querying geospatial data within KGs. This ontology is intentionally designed to be lightweight, fostering straightforward extension and integration. Core to the GeoSPARQL vocabulary are the concepts of Feature and Geometry, the former of which represents geospatial objects, while the latter describes their geographic extents and shapes, typically encoded as a Well-Known Text (WKT) literal. A Feature instance (e.g., a particular building) then links to a Geometry instance (e.g., a polygon linked to WKT text) using the hasGeometry property (relationship). Figure 1 depicts a subgraph of a GeoKG representing that the Eiffel Tower is within Paris, reusing GeoSPARQL ontology.

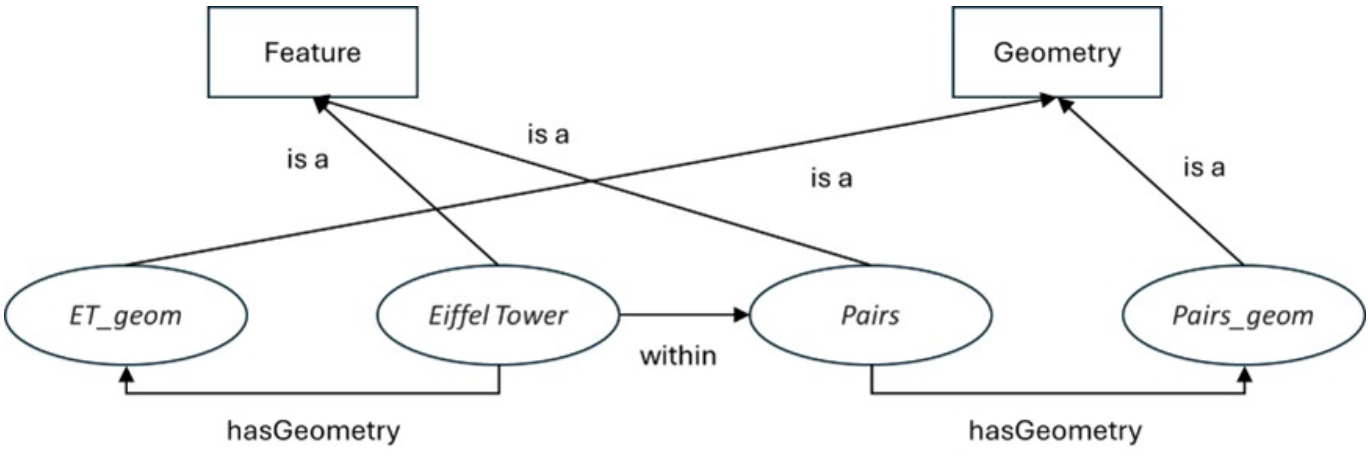


Figure 1. A subgraph of a GeoKG structured with the GeoSPARQL ontology, representing the statement “Eiffel Tower is located within Paris.” Italicized text denotes data instances, while regular text indicates ontology elements. Source: authors.

As a query language, GeoSPARQL defines several functions to enable querying based on spatial relationships within GeoKGs, using geometric information. These include functions for evaluating topological relationships (e.g., if a geometry contains, intersects, or overlaps another) and for non-topological computations such as calculating distances or creating buffers. A GeoSPARQL query illustrated in Listing 1 has the logic: "Find a café within a commercial area that is closest to a particular park." GeoSPARQL has been (partially) supported by several mainstream RDF stores, such as GraphDB (https://www.ontotext.com/), RDF4J (https://rdf4j.org/), and Stardog (https://www.stardog.com/) (Huang et al., 2019).

```
@[prefix definitions]

SELECT ?cafe ?cafeGeom

WHERE {

  ?cafe a ex:Cafe ;

        geo:hasGeometry ?cafeGeom .

  ?commercialArea a ex:CommercialArea ;

                  geo:hasGeometry ?commercialAreaGeom .

  ex:ASpecificPark geo:hasGeometry ?parkGeom .

  FILTER(geof:sfWithin(?cafeGeom, ?commercialAreaGeom))
}

ORDER BY geof:distance(?cafeGeom, ?parkGeom)

LIMIT 1
```

Listing 1. A GeoSPARQL Example. Source: authors.

The construction of GeoKGs, i.e., populating RDF triples from diverse sources and interlinking data instances, is largely an ETL (Extract, Transform, Load) process that can be implemented in various ways. In this regard, RDF mapping languages are particularly useful for translating other data models to RDF, such as R2RML (https://www.w3.org/TR/r2rml/) for relational databases and RML (https://rml.io/) for sources like CSV, JSON, and XML. Furthermore, “virtual GeoKGs” can be constructed from relational databases without materializing

RDF triples. These virtual GeoKGs can then be queried and utilized like materialized KGs, despite not being physically serialized into RDF. This is accomplished by SPARQL-to-SQL translation in real-time. This approach is beneficial, e.g., when integrating dynamic geospatial data such as traffic records. Ontop (https://ontop-vkg.org/) is a notable tool in this area, which supports GeoSPARQL queries over virtual GeoKGs (Bereta et al., 2019).

On the consumption and utilization of GeoKGs, dedicated tools have been developed for

their visualization and analysis, primarily in a spatial context. In this vein, a GeoEnrichment toolbox has been developed as a plug-in to enable direct querying of GeoKGs within ArcGIS (Mai et al., 2022). Such tools compose GeoSPARQL queries to access GeoKGs, with an interaction style similar to that of standard ArcGIS analytical tools. This allows the retrieved information to be used for visualization and further analysis within ArcGIS.

4. Prominent Examples of GeoKGs

Geospatial entities and their relationships are a natural integrator to consolidate data in various themes and from different sources, as everything happens at some geographic places during some period of time. Therefore, many GeoKGs have been developed in the past decade, which can be used to integrate and contextualize cross-domain datasets. GeoNames (https://www.geonames.org/) is an open gazetteer including over 25 million unique place names, together with their auxiliary information (e.g. place types, population, elevation, etc.), covering most countries and regions all around the world. LinkedGeoData (https://linkedgeodata.org/) is a KG version of OpenStreetMap. It consists of ~20 billion triples. Through the crosslink with GeoNames, places are enriched with more precise geometry and additional auxiliary information (e.g., opening hours). YAGO2 is a large-scale KG that contains enriched geospatial and temporal information, in which information from Wikipedia and GeoNames are combined to scope entities, facts, and events in the KG (Hoffart et al., 2013). YAGO2 was further extended in the YAGO2geo project with precise geometries from authoritative data sources in multiple countries (Karalis et al., 2019).

Although these aforementioned GeoKGs are useful, they were typically designed to utilize and integrate a limited set of data sources (e.g. LinkedGeoData is mainly from OpenStreetMap). KnowWhereGraph, a large-scale GeoKG integrating geospatial data from multiple sources, provides a new paradigm for building and accessing GeoKG (Zhu et al., 2025). First, it proposes a reusable ontology to facilitate the integration of geospatial data in different formats (e.g. remotely sensed images and geospatial vector data) using discrete global grids as the common locational unit (integrator). Second, KnowWhereGraph builds a stack of accessible tools, including GeoEnrichment plug-ins for ArcGIS and QGIS, customized disaster response platform, and a knowledge explorer search engine, for access from different user groups. Third, KnowWhereGraph links to general-purpose KGs like Wikidata, enriching itself with their vast repositories of factual knowledge.

5. Applications of GeoKGs

In merit of the rich data linking and semantic information carried by GeoKGs, they have been adopted in increasingly diverse applications. In this section, we discuss two application areas as examples to manifest the usefulness of GeoKGs, especially in terms of data integration and knowledge formalization.

GeoKGs, such as KnowWhereGraph, have been extensively used to help decision-makers respond to natural disasters thanks to their ability to provide situational awareness for any place on Earth. For instance, KnowWhereGraph was used by Direct Relief, a humanitarian organization, to determine where to send supplies in response to Hurricane Laura in 2020, a destructive Category 4 hurricane. In principle, it enabled decision-makers to quickly

retrieve relevant information, including demographic statistics, previous disasters, and health facilities, for regions affected by the storm. This process, which often takes hours or even days with traditional methods, can be accomplished in minutes with this GeoKG. Furthermore, it helped identify experts with local knowledge of the storm or the region by integrating both human and environmental information.

GeoKGs can formalize geovisualization processes by capturing expert knowledge on transforming raw geospatial data into graphics on maps. Huang et al. (2019) designed a geovisualization KG covering key Web mapping aspects like cartographic scale, data portrayal, and geometry source. This facilitates the interpretation, sharing, and reuse of the knowledge about how visualizations are produced, which is vital in scenarios like disaster response to ensure mutual understanding across diverse sectors.

6. Emerging Trends

Broadly, the term GeoKG encompasses both the graph-structured geospatial knowledge base (the artifact) and the methods, techniques, and standards for its realization and use. Learning this topic is challenging due to its interdependent technological stack and its nature as a rapidly advancing area.

The most prominent emerging trend is the synergy of machine learning (ML) and KGs in geospatial applications, offering transformative solutions to longstanding challenges in GeoKG construction and use. For example, ML can help: 1) extract geospatial entities (e.g., buildings, events) from diverse sources like imagery and textual reports; 2) integrate KGs by aligning their core components (concepts, relationships, instances); and 3) perform KG completion by predicting missing links (e.g., uncovering a previously unknown causal relationship between extreme weather and public health). Furthermore, the convergence of foundation models (e.g., Large Language Models) with GeoKGs opens new frontiers in GeoAI (Mai et al., 2024). This is a two-way enhancement: GeoKGs ground foundation models with structured geospatial knowledge for improved accuracy and interpretability, while these models significantly aid GeoKG construction, completion, and application (Pan et al., 2024).

References

Bereta, K., Xiao, G., & Koubarakis, M. (2019). Ontop-spatial: Ontop of geospatial databases. Journal of Web Semantics, 58, 100514.

Guarino, N., Oberle, D., Staab, S. (2009). What Is an Ontology?. In: Staab, S., Studer, R. (eds) Handbook on Ontologies. International Handbooks on Information Systems. Springer, Berlin, Heidelberg.

Hitzler, P. (2021). A review of the semantic web field. Communications of the ACM, 64(2), 76-83.

Hoffart, J., Suchanek, F. M., Berberich, K., & Weikum, G. (2013). YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. Artificial intelligence, 194, 28-61.

Huang, W., & Harrie, L. (2019). Towards knowledge-based geovisualisation using Semantic Web technologies: a knowledge representation approach coupling ontologies and rules. International Journal of Digital Earth, 13(9), 976–997.

Huang, W., Raza, S. A., Mirzov, O., & Harrie, L. (2019). Assessment and benchmarking of spatially enabled RDF stores for the next generation of spatial data infrastructure. ISPRS International Journal of Geo-Information, 8(7), 310.

Janowicz, K., Hitzler, P., Li, W., Rehberger, D., Schildhauer, M., Zhu, R., ... & Currier, K. (2022). Know, Know Where, KnowWhereGraph: A densely connected, cross-domain knowledge graph and geo-enrichment service stack for applications in environmental intelligence. AI Magazine, 43(1), 30-39.

Karalis, N., Mandilaras, G., & Koubarakis, M. (2019). Extending the YAGO2 knowledge graph with precise geospatial knowledge. In The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part II 18 (pp. 181-197). Springer International Publishing.

Ma, X. (2022). Knowledge graph construction and application in geosciences: A review. Computers & Geosciences, 161, 105082.

Mai, G., Huang, W., Cai, L., Zhu, R., & Lao, N. (2022). Narrative cartography with knowledge graphs. Journal of Geovisualization and Spatial Analysis, 6(1), 4.

Mai, G., Huang, W., Sun, J., Song, S., Mishra, D., Liu, N., ... & Lao, N. (2024). On the opportunities and challenges of foundation models for geoai (vision paper). ACM Transactions on Spatial Algorithms and Systems, 10(2), 1-46.

Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying large language models and knowledge graphs: A roadmap. IEEE Transactions on Knowledge and Data Engineering, 36(7), 3580-3599.

Zhu, R., Cai, L., Mai, G., Shimizu, C., Fisher, C. K., Janowicz, K., ... & Liu, Z. (2021, December). Providing Humanitarian Relief Support through Knowledge Graphs. In Proceedings of the 11th Knowledge Capture Conference (K-CAP '21). Association for Computing Machinery, New York, NY, USA, 285–288.

Zhu, R., Shimizu, C., Stephen, S., Fisher, C. K., Thelen, T., Currier, K., ... & Zalewski, J. (2025). The KnowWhereGraph: A Large-Scale Geo-Knowledge Graph for Interdisciplinary Knowledge Discovery and Geo-Enrichment. arXiv preprint arXiv:2502.13874.