Representing Geospatial Knowledge in Narratives

VALENTINA BARTALESI, Institute of Information Science and Technologies "Alessandro Faedo" (ISTI) of the National Research Council of Italy (CNR), Italy
NICOLÒ PRATELLI*, Institute of Information Science and Technologies "Alessandro Faedo" (ISTI) of the National Research Council of Italy (CNR), Italy and Department of Information Engineering (DII) of the University of Pisa, Italy

This paper explores the representation of geospatial knowledge within narratives through a Semantic Web approach. We introduce the NOnt+Space (NOnt+S) model, an extension of the CIDOC CRM-based Narrative Ontology, which allows the representation of narratives and their geospatial aspects. By leveraging standards such as CRMgeo and GeoSPARQL, NOnt+S ensures systematic and interoperable geospatial representation in narratives, enabling geospatial queries on knowledge graphs. We present an assessment of NOnt+S utilising data from the H2020 MOVING European project (2021-2024), which collected knowledge about European mountain value chains intended as Cultural Heritage. We have represented this knowledge as geospatial narratives using NOnt+S. GeoSPARQL queries and semantic reasoning applied to the created KG reveal the ontology ability to infer new geospatial knowledge. Our work contributes to the ongoing efforts in the Semantic Web community to integrate and represent geospatial information within narratives, promoting collaboration and interoperability across various scientific domains.

CCS Concepts: • **Computing methodologies → Ontology engineering**; **Spatial and physical reasoning**.

Additional Key Words and Phrases: Semantic Web, Knowledge Graph, CRMgeo, GeoSPARQL, Geospatial Narratives, Digital Humanities

ACM Reference Format:
Valentina Bartalesi and Nicolò Pratelli. 2024. Representing Geospatial Knowledge in Narratives. *ACM J. Comput. Cult. Herit.* 0, 0, Article 000 (2024), 16 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

A narrative serves as a method to articulate life experiences meaningfully [42] and forms the conceptual basis of collective human understanding [48]. Humans utilise narratives, or stories, to elucidate and communicate the meaningful aspects of abstract experiences, express a specific viewpoint on a particular domain of interest, and share meanings among diverse communities [29]. A widely accepted psychological thesis supporting the centrality of narrative in human life states that humans make sense of reality by organizing events into narratives [10, 46]. Many scientific communities have datasets that are rich repositories of narratives, often containing geospatial information about places and territories, e.g., ecologists, scholars working in economic sociology, and agricultural modelling experts.

Authors' addresses: Valentina Bartalesi, valentina.bartalesi@isti.cnr.it, Institute of Information Science and Technologies "Alessandro Faedo" (ISTI) of the National Research Council of Italy (CNR), via G.Moruzzi 1, Pisa, Italy, 56124; Nicolò Pratelli, nicolo.pratelli@isti.cnr.it, Institute of Information Science and Technologies "Alessandro Faedo" (ISTI) of the National Research Council of Italy (CNR), via G.Moruzzi 1, Pisa, Italy, 56124 and Department of Information Engineering (DII) of the University of Pisa, via G. Caruso 16, Pisa, Italy, 56122.

ACM 1556-4673/2024/0-ART000
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Our research is focused on the representation of geospatial data in narratives using a Semantic Web approach and following the Linked Data paradigm [24]. In this context, we consider narratives as semantic networks of events, possibly associated with texts or multimedia objects [30]. Using a semantic approach to represent narratives and their geospatial information offers several benefits. First, this approach provides a structured and interconnected way to organize and maximize the interoperability of data representation using a formal model. Additionally, formal narratives provide an effective framework for adding context to geospatial information. Formal narratives also facilitate semantic integration of data from diverse sources. This enables the harmonization and linkage of knowledge, including geospatial information, across different datasets based on a formal model, thereby creating cohesive narratives. Moreover, applying semantic reasoners [32] to narrative Knowledge Graphs (KGs) allows for the inference of new knowledge. Indeed, using a semantic reasoner combined with GeoSPARQL - a standard for representing and querying geospatial Linked Data from the Open Geospatial Consortium (OGC) [8] - allows discovering geospatial knowledge that is not explicitly present in the source data and may not be apparent in traditional, non-graph representations. Finally, representing geospatial knowledge as Linked Data facilitates sharing and integration across various applications, platforms, and organizations. This enhances interoperability and collaboration, crucial for overcoming data silos and supporting cross-disciplinary analysis.

In this paper, we present an extension of the Narrative Ontology (NOnt) [31] for representing geospatial knowledge, which we have named NOnt+Space (NOnt+S). The Narrative Ontology is based on the ISO standard International Committee for Documentation (CIDOC) Conceptual Reference Model (CRM) [17]. We developed NOnt+S using classes and properties from the CRMgeo vocabulary [22, 25], a model compatible with both CIDOC CRM and GeoSPARQL. We assessed our model using a dataset collected within the H2020 MOVING (MOuntain Valorisation through INterconnectedness and Green growth) [34] project. In this paper, we report and explain the advantages of formally representing geospatial knowledge in the context of narratives.

2 STANDARDS FOR REPRESENTING GEOSPATIAL KNOWLEDGE IN NARRATIVES

In the domain of the Semantic Web, numerous frameworks have emerged to formally represent the concept of "event", which is the core element of narratives [5]. Notable among others are the Event Ontology [39], the Linking Open Descriptions of Events (LODE) [43], the Event-Model-F ontology [41], and the Simple Event Model (SEM) [47]. Broader frameworks that include the representation of events are the CIDOC CRM [17], the Europeana Data Model [16, 33, 38], and the DOLCE upper-level ontology [21].

Taking into account the previous works, we created a CRM-based vocabulary to formally represent narratives, called Narrative Ontology (NOnt for short) [31]. This choice to use the CRM as reference ontology was essentially made to guarantee and maximise the interoperability of NOnt. However, drawing particular inspiration from the LODE and SEM models, we tailored NOnt to also effectively represent the factual constituents of events (e.g., persons, locations, concepts, etc.) [4]. Once NOnt was developed and a stable version had been released, we decided to extend it to also represent geospatial knowledge. Indeed, throughout history, descriptions of geospatial elements as well as maps have consistently provided support for stories and storytelling, e.g., in representing places mentioned in a story or the influence of territory on an author [12]. Integrating geospatial elements into narratives is crucial for enhancing the reader's experience. Indeed, territorial descriptions provide a spatial context that allows readers to better visualise the geographical settings, follow characters on a journey, and understand the cultural and historical significance of locations. Furthermore, geospatial elements evoke emotions by connecting readers to the physical environments of significant events. The inclusion of geospatial descriptions adds depth to the storytelling process, making it more engaging, immersive, and educational, depending on the genre and context of the narrative.

In the Semantic Web context, several pivotal vocabularies ensure a geospatial standardised approach. For example, GeoSPARQL [8] enhances SPARQL [37] ability to represent and query geospatial Linked Data in RDF

[36], a standard language for expressing data models. Furthermore, qualitative reasoning models such as DE-9IM [45] and RCC8 [27] - which allow representing topological relationships between spatial entities - were partially standardise and implemented in OGC and GeoSPARQL, respectively. Meanwhile, stRDF [28], and its query language stSPARQL were proposed as frameworks for handling geospatial data that change over time with RDF. Finally, several formal models have been developed to establish a linkage between the CIDOC CRM and GeoSPARQL [18, 35, 40]. Among these models, CRMgeo [25] has been proposed for approval by CIDOC CRM Special Interest Group (SIG)[1] as an official compatible model that allows the integration of GeoSPARQL and CIDOC CRM. In particular, CRMgeo aims to integrate geoinformation using the conceptualisations, formal definitions, encoding standards and topological relations defined by the Open Geospatial Consortium. To achieve this goal, CRMgeo links the CIDOC CRM to GeoSPARQL.

To extend NOnt with geospatial knowledge (NOnt+S), we chose CRMgeo as reference ontology. This choice was based on the following two main reasons: (i) CRMgeo is a compatible model of CIDOC CRM, which is the reference vocabulary of NOnt. Therefore, CRMgeo ensures a high level of interoperability with NOnt; (ii) CRMgeo conforms to the OGC standard GeoSPARQL, which provides an extension to the SPARQL query language for handling geospatial data and supports both qualitative and quantitative spatial reasoning systems[8, 20].

3 NONT+SPACE ONTOLOGY

In this section, we present the extension of NOnt to represent geospatial knowledge, i.e., NOnt+S. To develop NOnt+S, we mainly rely on classes and properties from the CRMgeo ontology and GeoSPARQL. Indeed, the CRMgeo connects its classes and properties to the ones provided by GeoSPARQL. This allows geoinformation systems to perform spatiotemporal analysis based on the semantic representation proposed by the CIDOC CRM.

As a notational convention, CIDOC CRM class identifiers begin with an 'E' followed by a number. Similarly, properties have both a name and an identifier that adhere to the same convention, with property identifiers starting with a 'P' and then a number. According to the conventions used by the CIDOC CRM, the classes declared in the CRMgeo were given both a name and an identifier. The identifiers for classes start with SP, followed by a number. The properties also have a name and an identifier based on the same conventions. The identifiers for properties start with Q followed by a number.

Firstly, we introduce the CRMgeo and GeoSPARQL classes and properties that we used in developing NOnt+S. Note that we added to NOnt+S only two properties defined by us, i.e., a property that links a narrative to a country and another that links a narrative to a Local Administrative Unit, i.e. territorial units consisting of municipalities in the 28 EU Member States[2]. The first property is called "is about country" and relates a narrative to a country. Using this property, we are able to state that a particular narrative as a whole describes or refers to a specific country. We defined this property as a subproperty of ecrm:P129_is_about, having the class Narrative of the NOnt+S as domain and ecrm:E1_Entity as range. The second property introduced by us is called "is about LAU" and relates a narrative to a Local Administrative Unite (LAU) 2. This property allows linking a narrative to a more specific territory than a country. It is defined as a subclass of ecrm:P129_is_about and has as domain the class Narrative of the NOnt+S and as range ecrm:E1_Entity.

The first two main and more general classes that we linked to extend NOnt with a geospatial representation are Event and Place. In the CIDOC CRM, an event and a place are instances of the classes ecrm:E5_Event and ecrm:E53_Place, respectively. The property ecrm:P7_took_place_at links an event to its place of occurrence. In the CRMgeo, ecrm:E53_Place has as subclass crmgeo:SP2_Phenomenal_Place, which *"comprises instances of E53 Place (S) whose extent (U) and position is defined by the spatial projection of the spatiotemporal extent of a real-world phenomenon that can be observed or measured. The spatial projection depends on the instance of SP3 Reference Space*

[1]https://cidoc-crm.org/node/206
[2]https://ec.europa.eu/eurostat/web/nuts/local-administrative-units

onto which the extent of the phenomenon is projected". An instance of the class crmgeo:SP2_Phenomenal_Place represents any place identified by an IRI (Internationalized Resource Identifier) in a standard gazetteer, such as Geonames for modern places [1] or Pleiades for ancient places [2]. Finally, the class crmgeo:SP3_Reference_Space *"comprises the (typically Euclidian) Space (S) that is at rest (I) in relation to an instance of E18 Physical Thing and extends (U) infinitely beyond it. It is the space in which we typically expect things to stay in place if no particular natural or human distortion processes occur"* (e.g., the space inside and around the Earth).

In NOnt+S an instance of ecrm:E5_Event is directly linked to an instance of crmgeo:SP2_Phenomenal_Place by the property ecrm:P7_took_place_at. A crmgeo:SP2_Phenomenal_Place is linked to a crmgeo:SP5_Geometric_Place_Expression, which *"comprises definitions of places by quantitative expressions. An instance of SP5 Geometric Place Expression can be seen as a prescription of how to find the location meant by this expression in the real world, which is based on measuring where the quantities referred to in the expression lead to, beginning from the reference points of the respective reference system. A form of expression may be geometries or map elements defined in a SP4 Spatial Coordinate Reference System that unambiguously identify locations in a SP3 Reference Space"*.

Since crmgeo:SP2_Phenomenal_Place and crmgeo:SP5_Geometric_Place_Expression are subclasses of geosparql:Feature and geosparql:Geometry respectively, we used the following properties also to directly link instances of SP2 to instances of SP5:

- geosparql:hasDefaultGeometry, which links a geosparql:Feature to a geosparql:Geometry. As default geometry is intended the geometry that should be used for spatial calculations in the absence of a request for a specific geometry;
- geosparql:hasGeometry, which links a geosparql:Feature to a geosparql:Geometry, which represents its spatial extent.

A crmgeo:SP2_Phenomenal_Place is linked through the property ecrm:P1_is_identified_by to a ecrm:E41_Appellation, which provides a name for the place in a natural language. A crmgeo:SP5_Geometric_Place_Expression is linked through the property ecrmgeo:Q9_is_expressed_in_term_of to a crmgeo:SP4_Spatial_Coordinate_Reference_System that provides spatial coordinate reference system that is used by the geometry.
Then, the crmgeo:SP4_Spatial_Coordinate_Reference_System is linked to a crmgeo:SP3_Reference_Space through property the crmgeo:Q7_describes. A crmgeo:SP5_Geometric_Place_Expression is also linked to its serialisation format through the property geosparql:hasSerialization, which has two subproperties corresponding to the following different kinds of literal:

- geosparql:asWKT, linking to a WKT literal;
- geosparql:asGML, linking to a GML literal.

The Well-Known Text (WKT) is a markup language for describing vector geometry objects [3] [14], and the Geography Markup Language (GML) is an XML grammar defined by the OGC to express geographical features [11, 13].

Thus, since NOnt+S is a GeoSPARQL-compliant model, it allows representing the latitude and longitude of a geographic point as a WKT or GML literal, e.g., the longitude and latitude of the city of Pisa (Italy) can be represented in WKT format using the Turtle (Terse RDF Triple Language) syntax [9] as follows:
<http://www.opengis.net/def/crs/OGC/1.3/CRS84> POINT (10.401 43.715)^^geosparql:wktLiteral.

The classes and properties reported and described above are shown in Figure 1. Nont+S is freely available for download at the following link: https://dlnarratives.eu/ontology/narrative_ontology.owl.

[3] https://www.iso.org/standard/76496.html

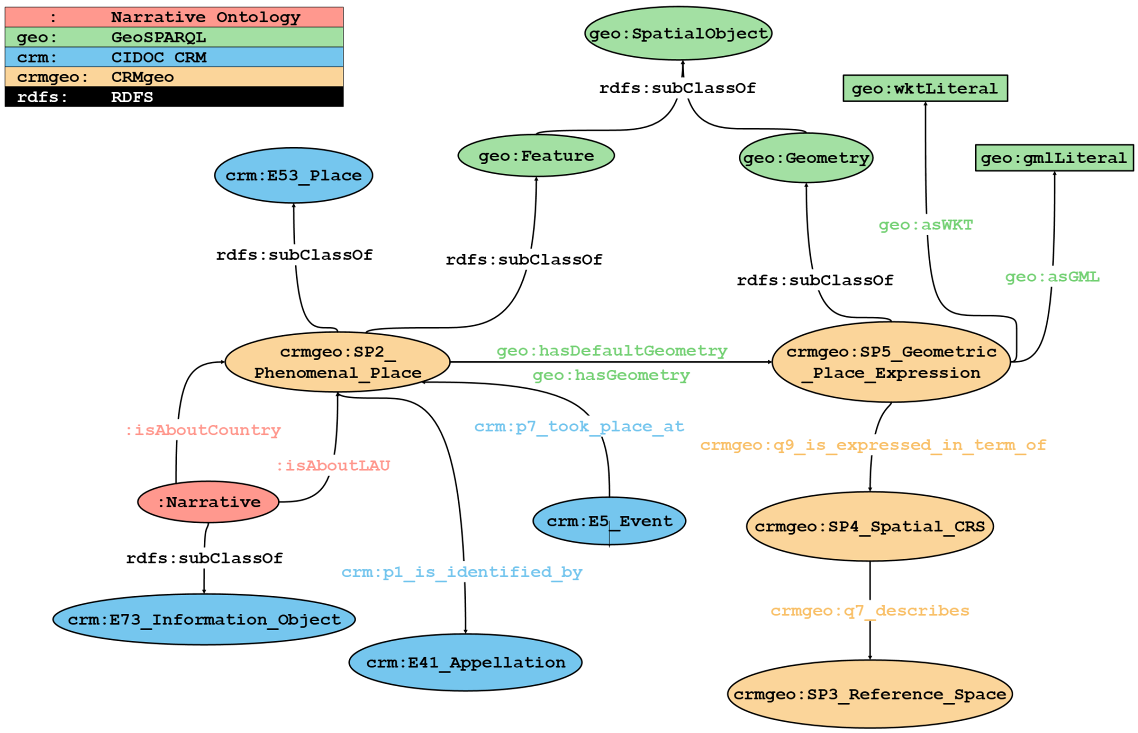

Fig. 1. The classes and properties of the NOnt+Space model.

4 ASSESSMENT OF NONT+SPACE USING TERRITORIAL DATA

As the partner expert in knowledge representation within the H2020 MOVING project, we applied NOnt+S to formally represent the collected data. MOVING aims to build capacities and co-develop relevant European policy frameworks for new or upgraded value chains that contribute to the resilience and sustainability of mountain areas under climate change. The project involves value chain actors, stakeholders and policy-makers in a bottom-up participatory process.

The original MOVING dataset is organized as an MS Excel file containing unstructured data. Specifically, the Excel file includes various types of data (e.g., income and gross value added, geography and population, employment, tourism, and geospatial information), which are not uniformly structured within the file but scattered as raw text across different Excel columns. No temporal information is included. We created a workflow that, in a semi-automatic manner, creates a KG compliant with the Narrative ontology starting from this Excel. Here, we report a brief description of the three main modules of the workflow, since describing them in detail is out of the scope of this paper. However, for an in-depth description of the entire workflow and data, please refer to [3]. The first module of the workflow preprocesses the Excel data to create a new textual document (in CSV format) for each MOVING region that corresponds to a story. In each CSV file, each row describes a single-story event and includes (i) a title, (ii) a description, (iii) an optional representative image, (iv) optional hyperlinks to online multimedia material, and (v) the event type. The sequence of rows in the CSV file corresponds to the sequence of events in the story for each region. The second module of the workflow employs natural language processing algorithms [15] to identify and retrieve key terms (such as persons, locations, organizations, and keywords) reported in the CSV files. Then, the module links these extracted terms to Wikidata entries, and for location entries, the corresponding geographic coordinates are extracted. Finally, the module adds this information to the events of each story in the CSV files. The third module of the workflow saves the sequence of events, along with all associated Wikidata entries and coordinates in a JavaScript Object Notation (JSON) file [6], according to the NOnt+S. Then, a Python script stores this document in a PostgreSQL-JSON database. Finally, the script calls a JAVA-triplifier software that converts the JSON document into a Web Ontology Language (OWL)[4] graph and stores it in an Apache Jena Fuseki triple store [26]. The workflow is designed to be application-independent and consists entirely of open-source components.

Once the OWL KG was created, we enriched it by also adding the polygons associated with the places cited in the source data. Indeed, the workflow retrieves only the corresponding coordinates of points from Wikidata for each location extracted from the source MS Excel file. To retrieve the polygons of the Wikidata location entries, we used the instance of the Qlever endpoint provided by the University of Freiburg[5], which allows querying the OpenStreetMap graph retrieving the results in the WKT format. Overall, QLever is an open-source general-purpose SPARQL engine with partial support for OGC GeoSPARQL 1.1 [7]. Furthermore, the source MS Excel file also reports the codes indicating the Local Administrative Units (LAUs) 2, whic omprise municipalities[6] as defined by the Nomenclature of territorial units of the statistical office of the European Union Eurostat. Currently, the LAU codes reported in Wikidata are only a subset (about 20000) of those defined by Eurostat (about 99000). For this reason, we decided to develop an algorithm[7] for mapping the LAU codes of the CSV files to the data provided by Eurostat. In particular, we mapped the codes reported in the MS Excel file with a GeoJSON file provided by Eurostat's Geographic Information System of the Commission (GISCO) [23]. This GeoJSON file reports the polygons in WKT format of all Eurostat-monitored LAU codes [19].

Figure 2 presents the population of the ontology depicted in Figure 1, using individuals. As a running example, a narrative resulting from query 1 (Q1), which is reported later in this section, has been selected. This example illustrates the representation of geospatial knowledge in the narrative about Betizu, an endangered breed of cattle from the Basque Country. This narrative is located in Spain, which is uniquely identified by the Wikidata IRI http://wikidata.org/entity/Q29, and, in particular, in the LAU named Goizueta, represented by the IRI https://dlnarratives.eu/place/373. Although this LAU has witnessed six events, only three are reported in the figure for the sake of brevity.

The LAU has a geometry indicated by the IRI https://dlnarratives.eu/geometry/373, expressed using the standard World Geodetic System (WGS) 84 (denoted by the IRI http://www.opengis.net/def/crs/OGC/1.3/CRS84), which describes a Euclidean space (http://wikidata.org/entity/Q17295). The geometry is serialized in WKT format, with the polygon only partially reported for brevity.

The resulting enriched KG contains ~200k (198676) triples that describe 172 narratives about as many different value chains, and it is available on GitHub[8]. The value chains we selected are from these countries: Italy, Spain, the UK and Switzerland.

The first outcome of our research is that NOnt+S is able to support the representation of all the selected MOVING value chains' data as narratives, including the associated geospatial information, i.e. points and polygons in WKT and GML formats.

[4]https://www.w3.org/OWL/
[5]https://qlever.cs.uni-freiburg.de/
[6]https://ec.europa.eu/eurostat/web/nuts/local-administrative-units
[7]https://github.com/AIMH-DHgroup/LAU_extraction
[8]https://github.com/AIMH-DHgroup/172_storymaps

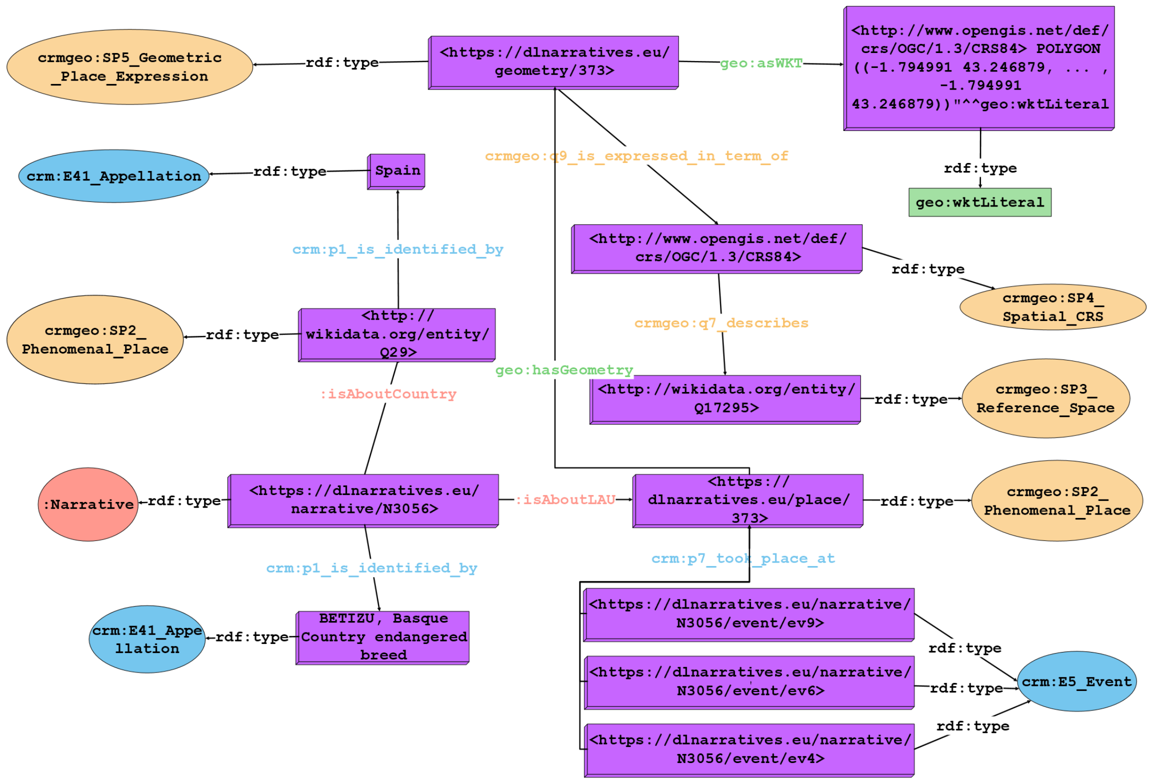

Fig. 2. A running example of the ontology population using individuals from the narrative titled "Betizu - Basque Country endangered breed", which is one of the narratives resulting from Query 1 (Q1).

To evaluate the representational power of NOnt+S and to assess the knowledge that can be inferred from the unstructured data collected in the MOVING project, we implemented several GeoSPARQL queries and used a semantic reasoner, i.e., Openllet [44]. GeoSPARQL provides spatial query capabilities such as finding spatial relationships, distances, and other geometric properties. Openllet is used to confirm that the class hierarchy aligned with the NOnt+S and to verify the consistency of the OWL graph (i.e., ensuring it did not produce contradictions). In particular, we performed the following queries (their code is reported in Annex A):

- Q1: The narratives whose location is in an area that overlaps with the Pyrenees polygon.
- Q2: The narratives whose location is in an area that overlaps with a 0.5-degree buffer (approximately 55.5 km) from the city of Madrid (Spain).
- Q3: The narratives whose location is in an area within the polygon representing Galles.
- Q4: The narratives whose location is in an area that overlaps with a 0.5-degree buffer (approximately 55 km) from Italian lakes.
- Q5: The Italian regions in which value chains producing cheese are located.

All these queries allow inferring new knowledge since the resulting information is not present in the source dataset. The results demonstrate how NOnt+S allows GeoSPRQL querying for different aspects of territories (i.e., inclusion, proximity, overlapping), making it possible to have a deeper view of the stories and their territories from a geospatial point of view. For example, retrieving the narratives sharing the same environmental characteristics (e.g. rivers, lakes, vineyards, chestnut trees) or insisting on territories near big cities or on a particular mountain chain. Overall, discovering new knowledge from data is particularly useful for mountain ecosystems to design sustainable environmental management pathways and contribute to long-term cities' ecological sustainability.

Figure 3 shows in a graphical form the results of query 1 (Q1). As shown in the figure, four narratives are located in the Pyrenees Mountains (Spain). These narratives are the following: Betizu - Basque Country endangered breed; Jacetania Bread, Indigenous pulses, Cereals and fruits of Ascara (Jaca), and Borda beer. Table 1 reports for each narrative, the title, the reference mountain region and a link to visualise the corresponding story map.

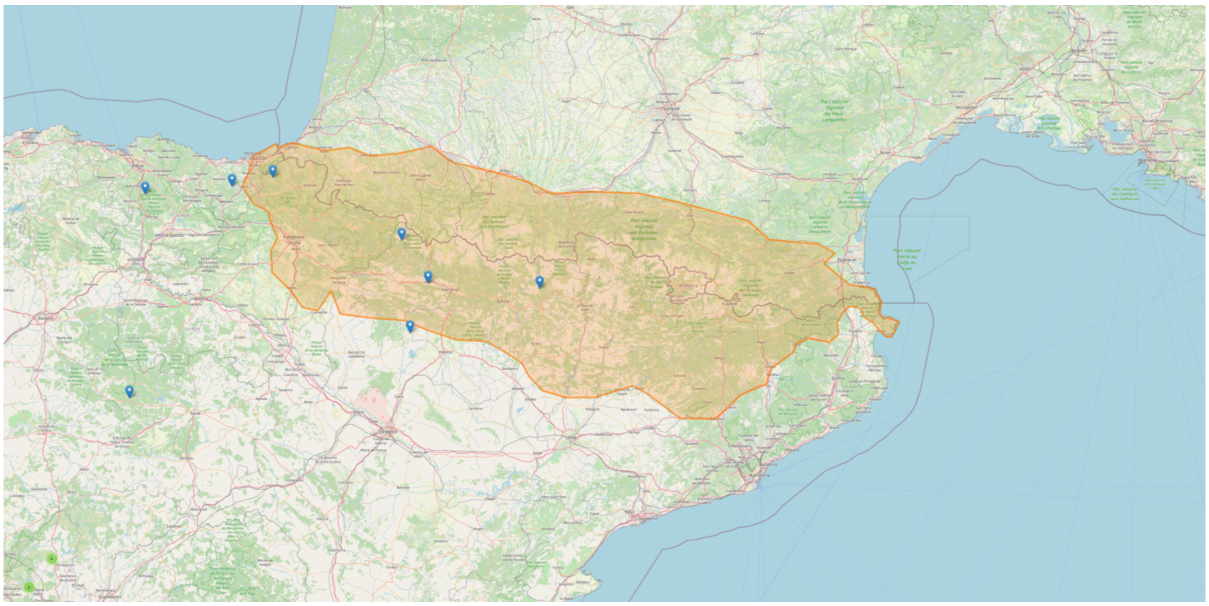

Fig. 3. A visual representation on a map of the Query 1 (Q1) results. Four narrative locations overlap the Pyrenees polygon.

Narrative Title	Reference region	Story Map
Betizu - Country endangered breed	Goizueta	https://acesse.dev/MOVINGnarrativeID56
Jacetania Breads	Ansó	https://encr.pw/MOVINGnarrativeID90
Indigenous pulse, cereals and fruits of Ascara	Jaca	https://encr.pw/MOVINGnarrativeID65
Borda beer	Sabiñanigo	https://l1nq.com/MOVINGnarrativeID79

Table 1. For each narrative regarding the value chains located in the Pyrenees mountains, the title, the reference mountain region and a link to visualise the corresponding story map are reported.

5 CONCLUSIONS

In this study, we propose a formal model to represent geospatial knowledge in narratives. Our approach leverages the advantages of knowledge graphs for structuring and interconnecting data narrated as stories, integrating the geospatial component. Furthermore, we exploit standard ontologies like CRMgeo and GeoSPARQL to ensure a consistent and interoperable formal representation of geospatial knowledge, enabling cross-disciplinary analysis and overcoming geospatial data silos.

We introduced the NOnt+Space ontology (NOnt+S), an extension of the CRM-based Narrative Ontology (NOnt) designed specifically for representing geospatial knowledge. This extension aligns with the GeoSPARQL standard, facilitating geospatial queries on knowledge graphs containing geographic data. Our assessment of NOnt+S, using data from the MOVING European project, demonstrated its effectiveness in representing geospatial knowledge within narratives and its capacity to infer new knowledge through GeoSPARQL queries and semantic reasoning.

The main challenge we faced in this research was to maximize the interoperability of the NOnt+S while minimizing the creation of custom classes and relationships and simultaneously, capture the diverse aspects of geospatial knowledge. For example, our model is able to represent points as well as polygons and territories at varying resolutions (from countries to Local Administrative Units). It also handles geospatial data in different formats such as WKT or GML. Most importantly, we integrated all these aspects of geospatial representation into a model that formalises narratives and associates each event with its corresponding geospatial characteristics.

Our work contributes to ongoing efforts in the Semantic Web community to integrate and represent geospatial information within narratives, promoting collaboration and interoperability across diverse scientific domains. Moving forward, systematic representation of geospatial knowledge in narratives will play a crucial role in uncovering new insights and deepening our understanding of the connections between territories and human experiences.

ACKNOWLEDGMENTS

The data used to validate the ontology were collected within the European Union's Horizon 2020 research and innovation programme under the MOVING project (grant agreement no 862739).

REFERENCES

[1] Dirk Ahlers. 2013. Assessment of the accuracy of GeoNames gazetteer data. In *Proceedings of the 7th workshop on geographic information retrieval*. 74–81.

[2] Elton Barker, Rainer Simon, Leif Isaksen, and Pau de Soto Cañamares. 2016. The Pleiades gazetteer and the Pelagios project. (2016).

[3] Valentina Bartalesi, Gianpaolo Coro, Emanuele Lenzi, Pasquale Pagano, and Nicolò Pratelli. 2023. From unstructured texts to semantic story maps. *International Journal of Digital Earth* 16, 1 (2023), 234–250.

[4] Valentina Bartalesi and Carlo Meghini. 2016. Formal Components of Narratives. In *Italian Research Conference on Digital Libraries*. Springer, Cham, 112–121. DOI: https://doi.org/10.1007/978-3-319-56300-8_11.

[5] Valentina Bartalesi, Carlo Meghini, and Daniele Metilli. 2016. Steps towards a formal ontology of narratives based on narratology. In *7th Workshop on Computational Models of Narrative (CMN 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

[6] Lindsay Bassett. 2015. *Introduction to JavaScript object notation: a to-the-point guide to JSON.* " O'Reilly Media, Inc.".

[7] Hannah Bast, Patrick Brosi, Johannes Kalmbach, and Axel Lehmann. 2021. An efficient RDF converter and SPARQL endpoint for the complete OpenStreetMap data. In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*. 536–539.

[8] Robert Battle and Dave Kolas. 2011. Geosparql: enabling a geospatial semantic web. *Semantic Web Journal* 3, 4 (2011), 355–370.

[9] David Beckett, Tim Berners-Lee, Eric Prud'hommeaux, and Gavin Carothers. 2014. RDF 1.1 Turtle. *World Wide Web Consortium* (2014), 18–31.

[10] Jerome Bruner. 1991. The narrative construction of reality. *Critical inquiry* 18, 1 (1991), 1–21.

[11] David S Burggraf. 2006. Geography markup language. *Data Science Journal* 5 (2006), 178–204.

[12] Sébastien Caquard and William Cartwright. 2014. Narrative cartography: From mapping stories to the narrative of maps and mapping. *The Cartographic Journal* 51, 2 (2014), 101–106.

[13] Open Geospatial Consortium. 2012. Geography Markup Language. https://www.ogc.org/standard/gml/

[14] Open Geospatial Consortium. 2019. Geographic information — Well-known text representation of coordinate reference systems. https://docs.ogc.org/is/18-010r7/18-010r7.html

[15] Gianpaolo Coro, Giancarlo Panichi, Pasquale Pagano, and Erico Perrone. 2021. NLPHub: An e-Infrastructure-based text mining hub. *Concurrency and Computation: Practice and Experience* 33, 5 (2021), e5986.

[16] Martin Doerr, Stefan Gradmann, Steffen Hennicke, Antoine Isaac, Carlo Meghini, and Herbert van de Sompel. 2010. The Europeana Data Model (EDM). In *World Library and Information Congress: 76th IFLA general conference and assembly*. IFLA, 10–15.

[17] Martin Doerr, Christian-Emil Ore, and Stephen Stead. 2007. The CIDOC Conceptual Reference Model: a new standard for knowledge sharing. In *Tutorials, posters, panels and industrial contributions at the 26th International Conference on Conceptual Modeling*. Australian Computer Society, Darlinghurst, 51–56.

[18] Vincent Ducatteeuw. 2021. Developing an Urban Gazetteer: A Semantic Web Database for Humanities Data. In *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Geospatial Humanities (GeoHumanities '21)*. Association for Computing Machinery, New York, NY, USA, 36–39. https://doi.org/10.1145/3486187.3490204

[19] Eurostat. 2024. LAU descriptions in GeoJSON format. https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/lau

[20] Christian Freksa. 1991. Qualitative spatial reasoning. In *Cognitive and linguistic aspects of geographic space*. Springer, 361–372.

[21] Aldo Gangemi, Nicola Guarino, Claudio Masolo, Alessandro Oltramari, and Luc Schneider. 2002. Sweetening ontologies with DOLCE. *Knowledge engineering and knowledge management: Ontologies and the Semantic Web* (2002), 223–233. DOI: https://doi.org/10.1007/3-540-45810-7_18.

[22] Øyvind Eide Gerald Hiebel, Martin Doerr and Maria Theodoridou. 2013. CRMgeo. https://www.cidoc-crm.org/crmgeo/

[23] Eurostat GISCO. 2015. Geographical Information and Maps. https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/countrie

[24] Tom Heath and Christian Bizer. 2011. *Linked data: Evolving the web into a global data space*. Vol. 1. Morgan & Claypool Publishers.

[25] Gerald Hiebel, Martin Doerr, and Øyvind Eide. 2017. CRMgeo: A spatiotemporal extension of CIDOC-CRM. *International Journal on Digital Libraries* 18 (2017), 271–279.

[26] Apache Jena. 2014. Apache jena fuseki. *The Apache Software Foundation* 18 (2014).

[27] Roman Kontchakov, Ian Pratt-Hartmann, and Michael Zakharyaschev. 2014. Spatial reasoning with RCC8 and connectedness constraints in Euclidean spaces. *Artificial Intelligence* 217 (2014), 43–75.

[28] Manolis Koubarakis and Kostis Kyzirakos. 2010. Modeling and Querying Metadata in the Semantic Sensor Web: The Model stRDF and the Query Language stSPARQL. In *The Semantic Web: Research and Applications*, Lora Aroyo, Grigoris Antoniou, Eero Hyvönen, Annette ten Teije, Heiner Stuckenschmidt, Liliana Cabral, and Tania Tudorache (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 425–439.

[29] Greg J McInerny, Min Chen, Robin Freeman, David Gavaghan, Miriah Meyer, Francis Rowland, David J Spiegelhalter, Moritz Stefaner, Geizi Tessarolo, and Joaquin Hortal. 2014. Information visualisation for science and policy: engaging users and avoiding bias. *Trends in ecology & evolution* 29, 3 (2014), 148–157.

[30] Carlo Meghini, Valentina Bartalesi, and Daniele Metilli. 2017. Using formal narratives in digital libraries. In *Italian Research Conference on Digital Libraries*. Springer, 83–94.

[31] Carlo Meghini, Valentina Bartalesi, and Daniele Metilli. 2021. Representing narratives in digital libraries: The narrative ontology. *Semantic Web* 12, 2 (2021), 241–264.

[32] Ravi Bhushan Mishra and Sandeep Kumar. 2011. Semantic web reasoners and languages. *Artificial Intelligence Review* 35 (2011), 339–368.

[33] Europeana Data Model. 2013. Definition of the Europeana Data Model v5.2.8. https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.8_102017.pdf

[34] MOVING. 2020. The MOVING project reference regionsBlue Cloud European project. https://www.moving-h2020.eu/reference-regions/.

[35] Gilles-Antoine Nys, Muriel Van Ruymbeke, and Roland Billen. 2018. Spatio-Temporal Reasoning in CIDOC CRM: An Hybrid Ontology with GeoSPARQL and OWL-Time. In *CEUR Workshop Proceedings*, Vol. 2230. RWTH Aachen University, Aachen, Germany.

[36] Jeff Z Pan. 2009. Resource description framework. In *Handbook on ontologies*. Springer, 71–90.

[37] Jorge Pérez, Marcelo Arenas, and Claudio Gutierrez. 2009. Semantics and complexity of SPARQL. *ACM Transactions on Database Systems (TODS)* 34, 3 (2009), 1–45.

[38] Europeana Data Model Primer. 2013. Europeana Data Model Primer. https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf

[39] Yves Raimond and Samer Abdallah. 2007. *The Event Ontology*. Technical Report. Queen Mary University of London. http://motools.sourceforge.net/event

[40] Babak Ranjgar, Abolghasem Sadeghi-Niaraki, Maryam Shakeri, and Soo-Mi Choi. 2022. An Ontological Data Model for Points of Interest (POI) in a Cultural Heritage Site. *Heritage Science* 10, 1 (March 2022), 13. https://doi.org/10.1186/s40494-021-00635-9

[41] Ansgar Scherp, Thomas Franz, Carsten Saathoff, and Steffen Staab. 2009. F: A model of events based on the foundational ontology Dolce+ DnS ultralight. In *Proceedings of the fifth International Conference on Knowledge Capture*. ACM, 137–144. DOI: https://doi.org/10.1145/1597735.1597760.

[42] Brian Schiff. 2012. The function of narrative: Toward a narrative psychology of meaning. *Narrative Matters* 2, 1 (2012), 33–47.

[43] Ryan Shaw, Raphael Troncy, and Lynda Hardman. 2009. LODE: Linking Open Descriptions of Events. In *The Semantic Web*. Springer, Berlin, Heidelberg, 153–167. DOI: https://doi.org/10.1007/978-3-642-10871-6_11.

[44] Gunjan Singh, Sumit Bhatia, and Raghava Mutharaju. 2020. OWL2Bench: a benchmark for OWL 2 reasoners. In *The Semantic Web–ISWC 2020: 19th International Semantic Web Conference, Athens, Greece, November 2–6, 2020, Proceedings, Part II 19*. Springer, 81–96.

[45] Christian Strobl. 2008. Dimensionally extended nine-intersection model (de-9im). (2008).

[46] Charles Taylor. 1992. *Sources of the self: The making of the modern identity.* Harvard University Press.
[47] Willem Robert Van Hage, Véronique Malaisé, Roxane Segers, Laura Hollink, and Guus Schreiber. 2011. Design and use of the Simple Event Model (SEM). *Web Semantics: Science, Services and Agents on the World Wide Web* 9, 2 (2011), 128–136. DOI: https://doi.org/10.1016/j.websem.2011.03.003.
[48] James V Wertsch and Henry L Roediger. 2008. Collective memory: Conceptual foundations and theoretical approaches. *Memory* 16, 3 (2008), 318–326.

A GEOSPARQL QUERIES

This appendix contains the code of the SPARQL queries that we have used to evaluate the representational power of the geospatial extension of NOnt, i.e., NOnt+S. These queries were run on the dataset available on GitHub at the following link: https://github.com/AIMH-DHgroup/172_storymaps.

Query 1: Narratives whose location is in an area that overlaps with the Pyrenees polygon.

```
PREFIX ecrm: <http://erlangen-crm.org/current/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX narra: <https://dlnarratives.eu/ontology#>
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX osm2rdfkey: <https://osm2rdf.cs.uni-freiburg.de/rdf/key#>
PREFIX crmgeo: <https://dlnarratives.eu/crmgeo/>

SELECT ?nlabel ?clabel
WHERE {
    ?narra narra:isAboutCountry ?country ;
           narra:isAboutLAU ?lau ;
           rdfs:label ?nlabel .
    ?country rdfs:label ?clabel .
    ?lau geo:hasGeometry ?glau .
    ?glau geo:asWKT ?pglau .
    FILTER(geof:sfWithin(?pglau,?wkt)).
    {
     SELECT ?wkt WHERE {
         SERVICE
         <https://qlever.cs.uni-freiburg.de/api/osm-planet> {
             ?osm_id osm2rdfkey:wikidata wd:Q12431 ;
                        geo:hasGeometry ?geometry .
                ?geometry geo:asWKT ?wkt .
                FILTER(REGEX(?wkt, "MULTIPOLYGON", "i" ))
         }
     } LIMIT 1
   }
}
```

Query 2: Narratives whose location is in an area that overlaps with a 5-degree buffer (approximately 55.5 km) from the city of Madrid (SP).

```
PREFIX uom: <http://www.opengis.net/def/uom/OGC/1.0/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX narra: <https://dlnarratives.eu/ontology#>
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX osm2rdfkey: <https://osm2rdf.cs.uni-freiburg.de/rdf/key#>

SELECT ?nlabel ?clabel
WHERE{
    ?narra narra:isAboutCountry ?country ;
           narra:isAboutLAU ?lau ;
            rdfs:label ?nlabel .
    ?country rdfs:label ?clabel .
    ?lau geo:hasGeometry ?glau .
    ?glau geo:asWKT ?pglau .
    FILTER(geof:sfWithin(
        ?pglau, geof:buffer(?wkt, 0.5, uom:degree))
    )
  {
    SELECT ?wkt WHERE {
     SERVICE
      <https://qlever.cs.uni-freiburg.de/api/osm-planet> {
        ?osm_id osm2rdfkey:wikidata wd:Q2807 ;
                geo:hasGeometry ?geometry .
        ?geometry geo:asWKT ?wkt .
        FILTER(REGEX(?wkt, "MULTIPOLYGON", "i" ))
        }
      }LIMIT 1
    }
}
```

Query 3: Narratives whose location is in an area within the polygon representing Galles

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX narra: <https://dlnarratives.eu/ontology#>
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX osm2rdfkey: <https://osm2rdf.cs.uni-freiburg.de/rdf/key#>

SELECT ?nlabel ?clabel
WHERE {
?narra narra:isAboutCountry ?country ;
       narra:isAboutLAU ?lau ;
       rdfs:label ?nlabel .
?country rdfs:label ?clabel .
?lau geo:hasGeometry ?glau .
?glau geo:asWKT ?pglau .
   FILTER(geof:sfWithin(?pglau,?wkt)).
{
     SELECT ?wkt WHERE {
          SERVICE
          <https://qlever.cs.uni-freiburg.de/api/osm-planet> {
              ?osm_id osm2rdfkey:wikidata wd:Q25 ;
                      geo:hasGeometry ?geometry .
                ?geometry geo:asWKT ?wkt .
              FILTER(REGEX(?wkt, "MULTIPOLYGON", "i" ))
          }
     }LIMIT 1
  }
}
```

Q4: Narratives whose location is in an area that overlaps with a 0.5-degree buffer (approximately 55 km) from Italian lakes

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX narra: <https://dlnarratives.eu/ontology#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX uom: <http://www.opengis.net/def/uom/OGC/1.0/>

SELECT DISTINCT ?nlabel ?clabel
WHERE {
    {
        SELECT ?coordinates WHERE {
            SERVICE
            <https://query.wikidata.org/sparql> {
                ?lake_wd wdt:P31/wdt:P279* wd:Q23397 ;
                        wdt:P17 wd:Q38 ;
                        wdt:P625 ?coordinates .
            }
        }
    }
    ?narra narra:isAboutCountry ?country ;
            narra:isAboutLAU ?lau ;
            rdfs:label ?nlabel .
    ?country rdfs:label ?clabel .
    ?lau geo:hasGeometry ?glau .
    ?glau geo:asWKT ?pglau .
    FILTER(geof:sfWithin(
        ?pglau, geof:buffer(?coordinates, 0.5, uom:degree))
    )
}
```

Q5: The Italian regions in which value chains producing cheese are located.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX narra: <https://dlnarratives.eu/ontology#>
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX osm2rdfkey: <https://osm2rdf.cs.uni-freiburg.de/rdf/key#>
PREFIX ecrm: <http://erlangen-crm.org/current/>
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>

SELECT DISTINCT ?name
WHERE {
  BIND(<https://dlnarratives.eu/resource/Q38> AS ?countryIRI)
    ?event narra:partOfNarrative?narrative .
    ?narrative narra:isAboutCountry ?countryIRI ;
           narra:isAboutLAU ?lau ;
       rdfs:label ?nlabel .
    ?lau geo:hasGeometry ?glau .
    ?glau geo:asWKT ?pglau .
    ?countryIRI rdfs:label ?country .
    ?narrative rdfs:label ?title .
    ?event narra:hasEntity <https://dlnarratives.eu/resource/Q10943> .

  FILTER(geof:sfIntersects(?pglau,?wkt)).
{
    SELECT DISTINCT ?name ?wkt  WHERE {
    SERVICE <https://qlever.cs.uni-freiburg.de/api/osm-planet>{
     ?region a osm:relation ;
        osm2rdfkey:wikidata?wd_id.
      ?region osmkey:name ?name .
      ?region geo:hasGeometry/geo:asWKT ?wkt .

        SERVICE <https://qlever.cs.uni-freiburg.de/api/wikidata>{
         ?wd_id wdt:P31/wdt:P279* wd:Q16110 ;
         wdt:P131 wd:Q38 . }
     }
    }
}
}
```