Linked Geospatial Data for Rural Territorial Sustainability: A Knowledge Graph of European Mountain Value Chains

Nicolò Pratelli
CNR-ISTI
Pisa, Italy
nicolo.pratelli@isti.cnr.it

Emanuele Lenzi
CNR-ISTI
DII-University of Pisa
Pisa, Italy
emanuele.lenzi@isti.cnr.it

Valentina Bartalesi
CNR-ISTI
Pisa, Italy
valentina.bartalesi@isti.cnr.it

***Abstract*—As global urbanisation trends continue, with projections indicating that 70% of the world population will reside in cities by 2050, there is growing concern over the sustainability of rural and mountain regions. These areas face increasing depopulation, threatening their socio-economic value chains (VCs). Addressing these challenges requires access to reliable and interoperable geospatial data. Knowing the exact location of mountain VCs can provide important insights to support territorial resilience. This paper investigates how Knowledge Representation and Semantic Web technologies can enhance the analysis of geographic data to support sustainable development in rural territories. As a case study, we use data from the H2020 MOVING project, encompassing 454 VCs across 16 European mountain regions. Our findings show that semantic technologies offer a valuable framework for integrating heterogeneous datasets, thereby improving decision-making and fostering resilience in rural areas.**

***Index Terms*—Knowledge Representation, Semantic Web, GeoSPARQL, CIDOC CRM, Wikidata, OpenStreetMap**

I. INTRODUCTION

As more people worldwide move from rural to urban areas, the global urban population is expected to reach approximately 70% by 2050. This significant change raises concerns about preserving cultural heritage, managing sustainable urban growth, and addressing environmental impacts. Increasing populations in cities will likely result in higher pollution levels and greater consumption of resources, creating challenges related to food supply, public health, and energy security. Additionally, cities will experience increased costs, administrative complexity, and a more substantial reliance on resources from distant areas. To mitigate the growth of the global urban population, it is essential to find effective ways to prevent rural and mountain regions — and preserve their value chains (VCs) — from experiencing depopulation. Such efforts align with global initiatives aimed at strengthening sustainable development and resilience in rural communities [1]. However, successfully addressing these issues requires access to significant amounts of reliable data. Geospatial data are especially significant in this context. Knowing the exact location of mountain VCs and their proximity or inclusion in natural parks, rivers, or lakes can provide important insights for local administrative organisations and stakeholders to evaluate the impact on the environment and to support the territory's resilience and its development. Often, these valuable geographic data are scattered across different repositories and are not easily combined or analysed. Knowledge representation and Semantic Web technologies can offer solutions by organising and linking diverse data according to common meanings and concepts. These technologies help make data more accessible, understandable, and useful for decision-making. This paper explores how this approach to data can effectively represent and manage geographic information to support solutions to urban-rural sustainability challenges. As a case study, we utilised data collected within the H2020 MOuntain Valorisation through INterconnectedness and Green growth (MOVING) project[1]. These data are organised in an MS Excel file and describe 16 European mountain regions and their 454 associated VCs.

The paper is structured as follows: in Section 2, we provide a state of the art of geospatial data representation. Section 3 describes the data collected within the MOVING project and introduces the ontology we used to represent these data formally. Section 4 describes the methodological steps to transform the source MOVING data into a knowledge Graph (KG). Section 5 reports several examples of GeoSPARQL queries that support and facilitate the discovery of new insights from the data collected in the KG and the evaluation of the query results. Finally, the conclusions are drawn in Section 6.

II. GEOSPATIAL DATA REPRESENTATION: STATE OF THE ART

The increasing demand for enriched representation and seamless querying of geospatial data has driven significant

This work was partially supported by ITSERR (Italian Strengthening of the ESFRI RI RESILIENCE), funded under the MUR National Recovery and Resilience Plan funded by the European Union-NextGenerationEU. The funders had no involvement in the study's design, data collection and analysis, publication decisions, or manuscript preparation.

[1]https://www.moving-h2020.eu/

advancements through the integration of Semantic Web technologies. The ability to semantically represent geographic and spatial information has expanded opportunities for data interoperability, automated reasoning, and enhanced knowledge extraction. Several research initiatives and European projects have laid the foundational groundwork in this domain, demonstrating diverse methodologies and application scenarios.

A. Semantic Enrichment of Geospatial Data

Semantic enrichment of geospatial data is crucial for improving interoperability and automated processing across various domains [2]–[4]. Especially in the context of crisis informatics [5], the semantification of geospatial information has been shown to integrate and manage heterogeneous, multi-source datasets effectively [3]. This approach leverages GeoSPARQL queries [6] and custom extensions for JSON [7], XML [8], and Comma-separated values (CSV) [9] data types, significantly enhancing data interoperability and interpretability. Notably, the approach extends the Apache Jena Semantic Web framework [10], yielding substantial performance improvements in geospatial querying, thus highlighting the importance of semantic enrichment in scenarios requiring rapid and accurate information synthesis. A related challenge, particularly in disaster management, is the heterogeneity of data sources and the lack of schema documentation. To address this, Prudhomme et al. [4] proposed a semantic interpretation process that combines ontology and schema matching with geocoding and natural language processing techniques. This methodology systematically transforms tabular geospatial data, often lacking explicit semantics, into structured Resource Description Framework (RDF) representations linked to established vocabularies such as GeoSPARQL.

B. European Research Initiatives on Geospatial Semantic Integration

Several European research projects have significantly contributed to integrating geospatial data into the Semantic Web. The GeoKnow EU project (2013–2015) [11] played a pivotal role in advancing geospatial semantic technologies by developing a comprehensive suite of tools under the GeoKnow Generator framework. This initiative streamlined workflows for integrating, semantically annotating, and manipulating heterogeneous geospatial datasets, facilitating more effective geographic information processing in the context of Linked Data [12].

The GeoKnow Generator Workbench [13] served as an enterprise-grade interface, incorporating essential functionalities such as data authorization and authentication, ensuring security and scalability. This comprehensive framework demonstrated the technical feasibility of geospatial linked data while significantly advancing Semantic Web applications requiring spatial intelligence.

Building on these foundations, Debruyne et al. [14] introduced a lightweight methodology to explore, enrich, and utilise geospatial data without requiring specialised GIS systems. Their approach involved transforming non-RDF data into RDF (uplifting), linking datasets, performing client-side geospatial processing, and converting enriched data back to conventional formats (downlifting). Successfully applied to Ireland's open data portal[2], this methodology enhanced accessibility and utility for non-specialist users without competing directly with dedicated GIS tools.

C. Advancements through AI and Linked Data Standards

The Horizon 2020 ExtremeEarth project (2019–2022) [15] marked a significant leap forward by combining deep learning and Semantic Web technologies to process and integrate vast volumes of Earth Observation (EO) data from the Copernicus program [16]. ExtremeEarth innovatively merged neural network-based analysis with semantic annotations, constructing sophisticated geospatial KGs that supported complex spatiotemporal queries. This integration reduced barriers for non-expert users, enabling actionable insights from EO data. The project set a new precedent for geospatial knowledge management, directly influencing European data infrastructures and demonstrating the transformative potential of semantic enrichment.

D. Semantic Web for Territorial Sustainability

Beyond technological advancements, several European initiatives have explored the role of semantically enriched geospatial data in promoting sustainability and resilience, particularly concerning mountain VCs. Projects such as MOVING (2020–2024) , TOP-Value (2017–2019), and AlpFoodway (2016–2019) have investigated integrated knowledge representation approaches to support territorial sustainability and cultural heritage preservation. While these initiatives address diverse objectives - from enhancing resilience in mountain agriculture to valorising food heritage - they collectively underscore the impact of semantic technologies in regional decision-making, fostering innovation, and enhancing community engagement.

III. Background

The MOVING project, funded under the European Union's Horizon 2020 programme (2020–2024), aims to enhance the resilience and sustainability of European mountain areas. It involves 23 partner organisations across Europe, including academic institutions, private companies, and territorial agencies engaged in monitoring and supporting VCs in mountain regions across 16 countries. MOVING focuses on strengthening the resilience of mountain territories in response to climate change and rural depopulation. As part of this initiative, project experts collected data on 454 mountain VCs. These data are organised in an MS Excel file[3], with each row representing a specific VC. The MS Excel file includes 31 columns per VC, containing a wide range of information that is not thematically structured (e.g., by environmental or economic categories). Overall, the data cover three main aspects: (i) qualitative descriptions of the territories' natural characteristics;

[2]https://data.gov.ie/

[3]https://github.com/prate91/Triplificatore/tree/main/excel

(ii) quantitative indicators related to geography, demographics, economics, and tourism; and (iii) specific attributes of the VC products, including the names of traditional products and details about the people or organisations involved in their production.

Despite the richness of the dataset, its current organisation in an MS Excel format presents several limitations. Tabular data offer limited support for modelling complex relationships between entities, such as VC actors or geographic regions. Furthermore, each VC is treated as an isolated entry, with no explicit links to others, thereby hindering correlation and comparative analysis across VCs. This siloed structure limits the identification of shared patterns, interdependencies, and common challenges among different mountain regions. In addition, the rigid schema of tabular data complicates integration with other datasets, particularly when data models are not aligned. To overcome these limitations, we applied a Knowledge Representation approach, transforming the tabular data into a knowledge graph (KG). A KG captures semantic relationships between entities - such as actors, products, and territorial characteristics - enabling the interconnection of VCs that share common features. This interconnected structure supports comparative analysis and the discovery of patterns across mountain regions. Moreover, by adopting a semantic model, the KG facilitates data interoperability and seamless integration with external datasets.

The KG is designed based on the NOnt+S ontology [17], developed by CNR-ISTI. NOnt+S extends the Narrative Ontology (NOnt) [18], enhancing its capacity to represent geospatial information. In the NOnt+S, a narrative is modelled as a semantic network of geospatial and temporal events interconnected through semantic relations. Each event involves multiple entities - such as people, places, organisations, works, and concepts - uniquely identified through Internationalized Resource Identifiers (IRIs). The NOnt+S builds upon the ISO standard CIDOC CRM [19] and its geospatial extension CRMgeo [20]. Specifically, for geospatial representation, the NOnt+S incorporates classes and properties from the CRMgeo vocabulary, ensuring compatibility with both CIDOC CRM and the Open Geospatial Consortium (OGC) standard GeoSPARQL [6].

IV. Methodology

This section reports the methodology we used to transform the data contained in the MOVING MS Excel file into a KG of VC narratives.

A. Data Pre-Processing

To automate the creation of a KG of VC narratives compliant with the NOnt+S ontology, we developed a Java algorithm[4]. This algorithm transforms the MS Excel file into a set of new CSV files, one for each VC. Each CSV file represents a preliminary structured version of a narrative, containing 11 rows - each corresponding to an event - and two columns for the event's title and textual description. The algorithm aggregates relevant columns from each Excel row to create comprehensive event descriptions that encapsulate the three VC aspects described in Section III. As a result, we obtained 454 CSV files, each representing a VC narrative[5].

B. Data Enrichment

This section describes the data augmentation procedures applied to the 454 CSV files.

1) Named Entity and Keyword Extraction: As the first step in the data enrichment process, we developed a Java-based entity extraction algorithm[6]. This algorithm was designed to identify named entities and contextually significant keywords within the textual descriptions of events in the CSV files. It integrates NLPHub [21], a cloud-based text mining service within the D4Science e-Infrastructure [22], which aggregates results from multiple state-of-the-art natural language processing tools.

To ensure compatibility with the NOnt+S ontology, NLPHub was configured to recognise named entities in the following categories: *Location*, *Person*, and *Organization*, along with keywords deemed semantically relevant in the given context. The output of this process is a set of 454 enriched CSV files, each containing a new column titled "entities", listing the identified named entities and keywords.

2) IRIs Association: To support the construction of a KG representing VC narratives, each extracted named entity and keyword was matched to its corresponding Wikidata IRI, as we rely on Wikidata as the reference knowledge base. This matching process involved querying the Wikidata SPARQL endpoint [23] using a Java software[7] we developed. This script performed queries to determine whether each named entity and keyword had a corresponding Wikidata entry.

Next, the Java software validated the returned IRIs by examining their linked Wikidata pages. In particular, an IRI was considered invalid under any of the following conditions: (i) it pointed to a Wikidata disambiguation page; (ii) the Wikidata entry's title did not match the entity detected by NLPHub; (iii) the Wikidata entry represented a different type of entity (e.g., distinguishing Austria the country (Q40) from Austria the typeface (Q783470)).

The output of this step was a new "IRI" column in each of the 454 CSV files, containing only the validated Wikidata IRIs for the identified named entities and keywords.

3) Geospatial Enrichment: To enrich the semantic representation of location-type entities, a Python software[8] was developed. This software retrieves geospatial coordinates from the corresponding Wikidata entries. Specifically, it extracts coordinate pairs under the Wikidata property P625 (coordinate location), recording them as longitude-latitude points.

[4]https://github.com/cybprojects65/MovingStoryBuilder

[5]https://github.com/cybprojects65/MovingStoryBuilder/blob/main/stories.zip

[6]https://github.com/cybprojects65/MovingStoryBuilder/blob/main/src/org/gcube/moving/nlphub/NLPHubCaller.java

[7]https://github.com/cybprojects65/MovingStoryBuilder/blob/main/src/org/gcube/moving/semantic/WikidataExplorer.java

[8]https://github.com/prate91/LAU_extraction

In addition to coordinate points, the software also attempts to retrieve polygon shapes. To achieve this goal, the Python software queries a large KG populated by the University of Freiburg with datasets such as Wikidata, DBpedia, and the RDF data from OpenStreetMap. This KG is queryable through an instance of the open-access QLever [24] endpoint made available by the University of Freiburg [25]. When available, polygons from OpenStreetMap were retrieved and formatted in Well-Known Text (WKT) format.

Furthermore, the original MS Excel file also reports the codes of the Local Administrative Units (LAUs) [26] of the municipalities relevant to the VCs, e.g. the LAU code 22205 for Trento, a municipality in Italy. To enrich the semantic representation of the LAUs, the Python software also extracts polygon geometries from one of Eurostat's GISCO [27] datasets. GISCO is a geographic information system funded by the European Union that provides geographic data for Nations and regions across Europe. The dataset we used is a GeoJSON file [28] containing WKT polygons for all LAUs in Europe.

All the extracted point and polygon geometries were stored in a new "geometry" column in each of the 454 CSV files.

C. Knowledge Graph Creation

This phase aims to transform the enriched 454 CSV files into an OWL 2 DL knowledge graph [29]. This transformation was carried out by a custom Java-based triplifier software[9] we developed, which models the knowledge according to the NOnt+S. Once created, the graph was stored on a public Apache Jena Fuseki server with GeoSPARQL support[10], which provides a SPARQL endpoint to access the data in the KG. The KG consists of a total of 503.963 triples. Each narrative in the global graph is modelled as a subgraph, representing one of the 454 VCs.

To ensure the logical consistency of the KG, we used a semantic reasoner - a tool designed to infer conclusions and detect contradictions based on a set of facts and rules. Specifically, we employed the Openllet [30] open-source reasoner to perform four key validation tasks: (i) Verifying that the graph is logically consistent (i.e., free from contradictions), (ii) confirming that the class hierarchy aligns with the structure defined by the NOnt+S, (iii) checking the integrity of the geometry data, including polygon closure and correct WKT formatting, and (iv) evaluating the ability to run complex SPARQL and GeoSPARQL queries.

Openllet successfully validated all tasks for our KG.

V. Inferring new Geospatial Knowledge for data discovering and analysis

The primary goal of our research is to verify that the KG we created facilitates and supports the discovery of new insights from the data, useful to support mountain territorial resilience and development. To this end, we gathered several queries that scientists and stakeholders involved in the MOVING project deemed valuable. This section presents a further selection of representative GeoSPARQL queries, extending the set of examples provided in [31]. The queries presented in this paper support and demonstrate the validity of the approach proposed in [31] through a broader evaluation conducted on geospatial data from territories different from those originally examined, thereby confirming its applicability across diverse territorial contexts and its potential to infer new knowledge. In particular, we present four GeoSPARQL queries to retrieve the VCs with the following geospatial characteristics:

1) located in the Carpathian Mountains (Q1)
2) located around Trento city (Italy) (Q2)
3) located in the Iberian Peninsula (Q3)
4) located around the main European rivers ($>$ 500km) (Q4)

The results of these queries can be used by experts to analyse the number and types of VCs operating in specific areas, allowing them to assess whether and how these VCs impact the exploitation of environmental resources, and potentially how to support the resilience and sustainable development of the VCs and their territories.

The code of the GeoSPARQL queries is reported in a GitHub repository, where it is also possible to automatically visualise the results. The repository is accessible through the following link: https://github.com/prate91/geosparql-ieee-ch.

1) Q1 - Value chains located in the Carpathian Mountains: Q1 retrieves the titles of VC narratives, their corresponding countries, and the LAU polygons that intersect with a polygon representing the Carpathian Mountains region. The results are shown in Figure 1. The LAU polygons indicate the specific territories where each VC is located. As illustrated in the figure, multiple VCs from different countries are identified and grouped based on the overlap of their LAU polygons with the Carpathian region. This geospatial overlap is not explicitly stated in the dataset and must be inferred through geospatial querying. The query accomplishes this by accessing the QLever endpoint hosted by the University of Freiburg, leveraging the OpenStreetMap subgraph to define the Carpathians as a polygonal geometry. It retrieves the title of each narrative, the corresponding country, and the WKT representation of the associated LAU geometry. A nested query accesses the external QLever endpoint to obtain the Carpathian Mountains' geometry using the corresponding Wikidata entity (Q1288). Finally, a spatial FILTER function identifies all LAU geometries that intersect with the Carpathian polygon. This approach enables the extraction of VCs operating within the Carpathian region based on geospatial inference.

2) Q2 - Value chains located around Trento city (Italy): Q2 extracts the titles, countries, and LAU geometries of the VCs operating within a maximum distance of 23 km from Trento, a city located in northern Italy. The results are displayed in Figure 2. The query structure is similar to that of Q1, but it does not use an external endpoint to retrieve the reference geometry. Instead, the FILTER clause performs an intersection between all the VC LAU geometries and a circular buffer of 0.3 degrees (40 km) around the Trento longitude-latitude coordinates.

[9]https://github.com/prate91/Triplificatore
[10]https://tool.dlnarratives.eu/Moving_454_Storymaps/geosparql.html

Fig. 1. Visualization of Q1 results

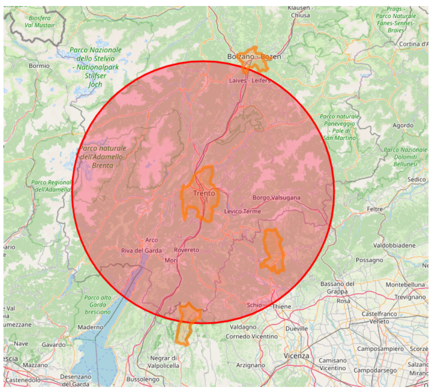

Fig. 2. Visualization of Q2 results

3) Q3 - Value chains lacated in the Iberian Peninsula: Q3 extracts the titles, countries, and LAU geometries of the VCs located in the Iberian Peninsula. The results are shown in Figure 3. The query uses the same external QLever OpenStreetMap endpoint as Query 1 to retrieve the geometry of the Iberian Peninsula's boundaries. The FILTER clause performs an intersection between the Iberian Peninsula polygon and the LAU geometries of the VCs.

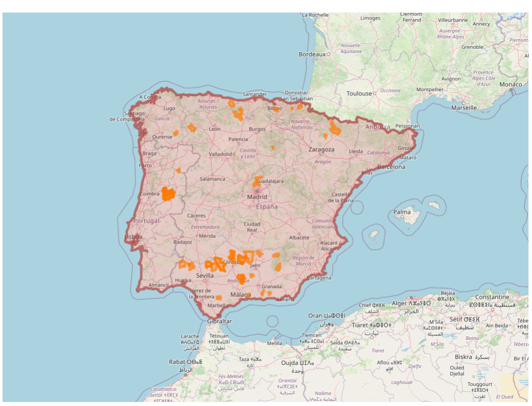

Fig. 3. Visualization of Q3 results

4) Q4 - Value chains located around the main European rivers: Q4 retrieves all VCs located in proximity to European rivers longer than 500 kilometres. The results are presented in Figure 4. While the overall structure of the query is similar to that of Query 1 and relies on the same external QLever endpoint hosted by the University of Freiburg, it introduces two key modifications. First, the nested SELECT clause operates across two QLever subgraphs: OpenStreetMap and Wikidata. River geometries are extracted from the OpenStreetMap subgraph, while the Wikidata subgraph is queried to identify entities classified as "rivers" (wd:Q4022), located in Europe (wd:Q46), and having a "length" property (P2043) exceeding 500 kilometres. This constraint is applied using a FILTER (?length $>$ 500) condition. Second, another FILTER clause determines which VC-associated LAU geometries intersect with the geometries of the selected European rivers.

Fig. 4. Visualization of Q4 results

A. Query Validation

To verify the correctness of the query results, we manually validated them with the support of a MOVING project expert. We calculated standard performance metrics: precision, which verifies that all retrieved instances are relevant, and recall, which ensures that all relevant instances are retrieved. The evaluation confirmed that the results were correct and complete, with precision and recall values equal to 1. This demonstrates that the queries are effective in retrieving (i) VCs located within a specific mountain range, (ii) VCs included in a circular buffer around a city, (iii) country-specific VCs, and (iv) VCs located in proximity to rivers.

VI. Conclusion

In this paper, we present a workflow for constructing a knowledge graph of narratives describing mountain value chains (VCs), based on data collected within the European MOVING project. The source dataset consists of information on 454 European mountain VCs, initially stored in an MS Excel file. The knowledge base is modelled using the NOnt+S, a CRM-based ontology designed for the formal representation of narratives composed of geospatial events. To build the graph, the original data were pre-processed to generate a separate CSV file for each VC, aggregating relevant cells from the MS Excel file to construct the events forming the narrative structure. These CSV files were further enriched by identifying named entities and keywords mentioned in the data, as well as the geographic coordinates (points and polygons) of referenced locations. For each identified named entity and keyword, we linked the corresponding IRI from Wikidata, which was selected as the reference knowledge base. The final output is an OWL 2 DL-compliant knowledge graph, which was stored in a Fuseki triplestore. We illustrated the potential of our approach by presenting four GeoSPARQL queries, developed in collaboration with MOVING project experts. These queries aim to retrieve insights aligned with experts' and stakeholders' needs for supporting territorial resilience and sustainable management. As demonstrated by the query results, the constructed KG provides a valuable decision-support tool for local administrative bodies and stakeholders, enabling a deeper understanding of the environmental impact of mountain VCs and facilitating enhanced analysis through geospatial data integration.

References

[1] United Nations. (2015) Sustainable Development Goal 11, Make cities and human settlements inclusive, safe, resilient and sustainable. Accessed: 2025-04-10. [Online]. Available: https://sdgs.un.org/goals/goal11

[2] T. Homburg, C. Prudhomme, F. Würriehausen, A. Karmacharya, F. Boochs, A. Roxin, and C. Cruz, "Interpreting Heterogeneous Geospatial Data Using Semantic Web Technologies," pp. 240–255. [Online]. Available: https://www.academia.edu/80545832/Interpreting_Heterogeneous_Geospatial_Data_Using_Semantic_Web_Technologies

[3] C. Stadler, S. Bin, L. Bühmann, N. Radtke, K. Junghanns, S. Gründer-Fahrer, and M. Martin, "Semantification of Geospatial Information for Enriched Knowledge Representation in Context of Crisis Informatics."

[4] C. Prudhomme, T. Homburg, J.-J. Ponciano, F. Boochs, C. Cruz, and A.-M. Roxin, "Interpretation and automatic integration of geospatial data into the Semantic Web," vol. 102, no. 2, pp. 365–391. [Online]. Available: https://doi.org/10.1007/s00607-019-00701-y

[5] L. Palen, S. Vieweg, J. Sutton, S. B. Liu, and A. Hughes, "Crisis informatics: Studying crisis in a networked world," in *Proceedings of the third international conference on E-Social Science*, 2007, pp. 7–9.

[6] R. Battle and D. Kolas, "Enabling the geospatial Semantic Web with Parliament and GeoSPARQL," *Semantic Web*, vol. 3, no. 4, pp. 355–370, Jan. 2012.

[7] C. J. Ihrig, "JavaScript Object Notation," in *Pro Node.Js for Developers*, C. J. Ihrig, Ed. Apress, pp. 263–270. [Online]. Available: https://doi.org/10.1007/978-1-4302-5861-2_17

[8] T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler, and F. Yergeau, "Extensible markup language (xml)," *World Wide Web Journal*, vol. 2, no. 4, pp. 27–66, 1997.

[9] Y. Shafranovich, "Common format and mime type for comma-separated values (csv) files," 2005.

[10] Apache. Apache Jena Framework. Accessed: 2025-04-10. [Online]. Available: https://jena.apache.org/

[11] J. Lehmann, S. Athanasiou, A. Both, L. Buehmann, A. Garcia-Rojas, G. Giannopoulos, D. Hladky, K. Hoeffner, J. Grange, A. Ngonga Ngomo *et al.*, "The geoknow handbook," Technical report, Tech. Rep., 2015.

[12] C. Bizer, T. Heath, K. Idehen, and T. Berners-Lee, "Linked data on the web (ldow2008)," in *Proceedings of the 17th international conference on World Wide Web*, 2008, pp. 1265–1266.

[13] A. Both, M. Wauer, A. G. Rojas, D. Hladky, and J. Lehmann, "The GeoKnow Generator Workbench - an Integrated Tool Supporting the Linked Data Lifecycle for Enterprise Usage," in *Joint Proceedings of the Posters and Demos Track of 11th International Conference on Semantic Systems - SEMANTiCS2015 and 1st Workshop on Data Science: Methods, Technology and Applications (DSci15)*, ser. CEUR Workshop Proceedings, A. Filipowska, R. Verborgh, and A. Polleres, Eds., vol. 1481. CEUR, pp. 92–95. [Online]. Available: https://ceur-ws.org/Vol-1481/#paper28

[14] C. Debruyne, K. McGlinn, L. McNerney, and D. O'Sullivan, "A lightweight approach to explore, enrich and use data with a geospatial dimension with semantic web technologies," in *Proceedings of the Fourth International ACM Workshop on Managing and Mining Enriched Geo-Spatial Data*, ser. GeoRich '17. Association for Computing Machinery, pp. 1–6. [Online]. Available: https://dl.acm.org/doi/10.1145/3080546.3080548

[15] D. H. Hagos, T. Kakantousis, V. Vlassov, S. Sheikholeslami, T. Wang, J. Dowling, C. Paris, D. Marinelli, G. Weikmann, L. Bruzzone, S. Khaleghian, T. Kræmer, T. Eltoft, A. Marinoni, D.-A. Pantazi, G. Stamoulis, D. Bilidas, G. Papadakis, G. Mandilaras, M. Koubarakis, A. Troumpoukis, S. Konstantopoulos, M. Muerth, F. Appel, A. Fleming, and A. Cziferszky, "ExtremeEarth Meets Satellite Data From Space," vol. 14, pp. 9038–9063. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9523756

[16] S. Jutz and M. P. Milagro-Perez, "Copernicus: the european earth observation programme," *Revista de Teledetección*, no. 56, pp. V–XI, 2020.

[17] V. Bartalesi and N. Pratelli, "Representing geospatial knowledge in narratives," *J. Comput. Cult. Herit.*, vol. 18, no. 1, Feb. 2025. [Online]. Available: https://doi.org/10.1145/3703918

[18] C. Meghini, V. Bartalesi, and D. Metilli, "Representing narratives in digital libraries: The narrative ontology," *Semantic Web*, vol. 12, no. 2, pp. 241–264, 2021.

[19] M. Doerr, C.-E. Ore, and S. Stead, "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 - Volume 83*, ser. ER '07. AUS: Australian Computer Society, Inc., 2007, p. 51–56.

[20] G. Hiebel, M. Doerr, and Ø. Eide, "Crmgeo: A spatiotemporal extension of cidoc-crm," *International Journal on Digital Libraries*, vol. 18, pp. 271–279, 2017.

[21] G. Coro, G. Panichi, P. Pagano, and E. Perrone, "Nlphub: An e-infrastructure-based text mining hub," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 5, p. e5986, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.5986

[22] M. Assante, L. Candela, D. Castelli, R. Cirillo, G. Coro, L. Frosini, L. Lelii, F. Mangiacrapa, P. Pagano, G. Panichi, and F. Sinibaldi, "Enacting open science by d4science," *Future Generation Computer Systems*, vol. 101, pp. 555–563, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X1831464X

[23] Wikidata. (2024) SPARQL entity retrieval specifications and examples. Accessed: 2025-04-20. [Online]. Available: https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples

[24] H. Bast and B. Buchhold, "Qlever: A query engine for efficient sparql+text search," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, ser. CIKM '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 647–656. [Online]. Available: https://doi.org/10.1145/3132847.3132921

[25] University of Freiburg. (2024) QLever instance. Accessed: 2025-04-10. [Online]. Available: https://qlever.cs.uni-freiburg.de/

[26] Eurostat. (2024) Local Administrative Units. Accessed: 2025-04-10. [Online]. Available: https://ec.europa.eu/eurostat/web/nuts/local-administrative-units

[27] Eurostat. GISCO. (2023) the Geographic Information System of the Commission. Accessed: 2025-04-10. [Online]. Available: https://ec.europa.eu/eurostat/web/gisco

[28] Eurostat. (2024) LAU descriptions in GeoJSON format. Accessed: 2025-04-10. [Online]. Available: https://ec.europa.eu/eurostat/web/gisco/geodata/statistical-units/local-administrative-units

[29] W. W. W. Consortium *et al.*, "Owl 2 web ontology language document overview," 2012.

[30] Openllet. (2023) Openllet: An Open Source OWL DL reasoner for Java. Accessed: 2025-04-10. [Online]. Available: https://github.com/Galigator/openllet

[31] V. Bartalesi, G. Coro, E. Lenzi, N. Pratelli, P. Pagano, M. Moretti, and G. Brunori, "A semantic knowledge graph of european mountain value chains," *Scientific Data*, vol. 11, 2024.