scientific **data**

OPEN

DATA DESCRIPTOR

A Semantic Knowledge Graph of European Mountain Value Chains

Valentina Bartalesi[1], Gianpaolo Coro [1 ✉], Emanuele Lenzi[1], Nicolò Pratelli[1], Pasquale Pagano[1], Michele Moretti[2] & Gianluca Brunori[2]

The United Nations forecast a significant shift in global population distribution by 2050, with rural populations projected to decline. This decline will particularly challenge mountain areas' cultural heritage, well-being, and economic sustainability. Understanding the economic, environmental, and societal effects of rural population decline is particularly important in Europe, where mountainous regions are vital for supplying goods. The present paper describes a geospatially explicit semantic knowledge graph containing information on 454 European mountain value chains. It is the first large-size, structured collection of information on mountain value chains. Our graph, structured through ontology-based semantic modelling, offers representations of the value chains in the form of narratives. The graph was constructed semi-automatically from unstructured data provided by mountain-area expert scholars. It is accessible through a public repository and explorable through interactive Story Maps and a semantic Web service. Through semantic queries, we demonstrate that the graph allows for exploring territorial complexities and discovering new knowledge on mountain areas' environmental, societal, territory, and economic aspects that could help stem depopulation.

Background & Summary

The 2018 update of the World Urbanization Prospects, released by the United Nations Department of Economic and Social Affairs, projects a significant shift in the global population. Currently, 47% of the World population is rural, but this is expected to decline to 30% by 2050[1]. This transition raises concerns for traditional and cultural heritage, urban well-being, and ecological sustainability. Massive rural-to-urban migration will increase city populations, pollution, and energy consumption. The United Nations (UN) predicts that by 2050, 7 billion people will live in cities[2], creating unsustainable conditions for food, health, and energy security. Challenges include excessive public administration burdens, cost of living mismatched with salaries, surges in pollution and greenhouse gas emissions, and increased healthcare expenditures. Additionally, cities will depend more on distant resources, leading to negative environmental impacts. In Europe, 36% of the territory is mountainous, critical for public and private-goods supply. Therefore, understanding factors that can mitigate rural and mountain area depopulation is crucial for sustainability. This aligns with the UN's Sustainable Development Goal 11[2], emphasising strategies to preserve rural economies and services, including economic diversification and tourism enhancements. Recent international initiatives promote capacity-building and participatory processes involving stakeholders and policymakers to create resilient and sustainable mountain areas in response to climate change[3].

However, such endeavours demand substantial volumes of data to generate meaningful insights. Specifically, data related to environment, geography, demographics, and economics are essential for comprehending how regions and mountain-based value chains react and adapt to climate change. Such information is key to gaining new knowledge on the dynamics of mountain value chains. For example, it allows identifying the value chains sharing the same environmental characteristics (e.g. rivers, lakes, vineyards, chestnut trees) and issues (e.g. depopulation, emigration, climate change problems) or similar products across territories (e.g. cow or sheep milk cheese).

The present paper describes an extensive collection of data on 454 value chains from 23 European mountain areas belonging to 16 countries, representing as far as possible the diversity of European mountain areas (Supplementary Table 1). Although these data do not cover the entire spectrum of European mountain value chains, they aim to be as representative and reliable as possible by embedding information from local experts with comprehensive views. These experts overall selected the 454 value chains as those with the highest importance in the 23

[1]Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" del Consiglio Nazionale delle Ricerche (ISTI-CNR), Pisa, 56124, Italy. [2]Dipartimento di Scienze Agrarie, Alimentari e Agro-ambientali, dell'Università di Pisa, Pisa, 56124, Italy. ✉e-mail: gianpaolo.coro@isti.cnr.it

areas from a socio-ecological (innovation, stage of development, size, governance system, and environmental status and protection) perspective. When they had sufficient information, the experts also extended the data beyond their monitored territories to connect and compare the value chains with similar or related territories and other value chains not initially involved in the project (e.g., the Scandinavian Mountains, the Massif Central, and the Pyrenees).

Data representation and publication principles. Our data collection is organised as a *semantic knowledge graph*, i.e., a structured representation of knowledge, where knowledge entities and their relations are modelled in a graph format[4,5]. The structure of our graph is based on an ontology. An ontology (*computational ontology*, formally) is a model of the knowledge structure of a system in terms of the significant classes (e.g., event, person, location, object) and relations emerging from the observation of the system itself[5]. It is an abstract, simplified view of the system structured in a machine-readable format. Building an ontology requires identifying the relevant concepts and relations (symbolised by unary and binary predicates) of a domain of interest, and organising them in a hierarchical structure. Summarising, a semantic knowledge graph based on an ontology has nodes corresponding to the classes of the ontology and edges corresponding to the relations. A narrative is an example of a system describable as a knowledge graph modelled on an ontology.

Our collection is a semantic knowledge graph of narratives, where each narrative is a sub-graph explaining one among the 454 value chains. Each value chain narrative is a semantic network of narrative events related to each other through plot-dependent semantic relations. The overall graph is described under the Web Ontology Language (OWL) and complies with the Narrative Ontology[7] (NOnt), which provides a structure to represent the knowledge of a narrative formally. NOnt is used in several cultural heritage projects (e.g., CRAEFT[8], Mingei[9], and IMAGO[10]). It reuses classes and properties (and complements) the CIDOC CRM ISO standard[11] and other standard ontologies (FRBRoo[12], OWL Time[13], and GeoSPARQL[14]). NOnt reuses most concepts from the CIDOC CRM ontology - among which is the concept of event - because this is used by many European cultural and museum institutions for data representation. NOnt adds new concepts, such as the role of a narrative actor, the fabula and plot, and new properties, such as the causal dependency between two events and the geospatial belonging of an event to a country or a territory. By reusing concepts from other consolidated ontologies, NOnt enhances interoperability with other vocabularies and semantic knowledge bases and allows for building more extensive knowledge networks in agreement with the Linked Open Data paradigm[15]. From a conservation perspective, our analysed value chains are part of the European cultural heritage and suitably described as narratives because they relate to European territories' history, artisanal knowledge, and environmental characteristics.

We used narratives for data description also because they are central to human activities across cultural, scientific, and social domains and can establish shared meanings among diverse domain-specific communities[16]. Psychological theories assert that humans comprehend reality by organising events into narratives[17,18]. Stories can depict characters' intentions, emotions, and aspirations through the attributes of objects and events[19] and can describe overall territorial aspects beyond the analytical data[20,21].

Our value chain narratives include comprehensive information on the selected European product value chains, e.g., about the production of European cheese, beer, milk, sheep and dairy farming, flour, herbs, oil, wine, tourism, carpentry, food and drink, nuts, and others. Overall, they cover economic assets, biodiversity, and ecosystem service descriptions (e.g., food and water resources and touristic features). Therefore, our representation also includes geographical information such as maps, pinned locations, and polygonal areas. A map is a valuable support to represent the spatiotemporal structure of a territory story and the relationships between places[21]. For such reason, we also represented the value chain spatiotemporal narratives as *Story Maps*, i.e., online interactive maps enriched with textual/audio events and digital material narrating overall territorial complexity. Story Maps allow exploring and navigating a narrative through many digital devices (e.g., PCs, tablets, smartphones, interactive displays) and can be built through collaborative tools[20]. They are valuable to represent the life, emotions, reality, fiction, legends, and expectations associated with the described territory beyond a mere map representation[20–24], and fill the perceptual gap between a *territory-as-a-whole* and its map[25].

Our principal target data users and stakeholders are policymakers at all spatial scales, from local to European. The knowledge contained in the data is valuable to designing local, regional, national and/or European policies, strategies, and actions to promote the development of mountain areas, starting from the value chains that populate these areas. In fact, the stakeholders can use this knowledge to understand the prevailing economic sector (primary, secondary, or tertiary) of their respective regions. They can also infer information at a finer spatial scale, such as the resources (natural, cultural, and others) on which the productive fabrics depend. Based on this information, they can design place-based and data-driven policies supporting the socio-economic development of marginalised mountain areas. Other stakeholders of our data are citizens who wish to have an overview of their regions' territorial and economic assets and the related peculiarities, competitions, and challenges in Europe.

Our semantic knowledge graph is available as a collection on a Figshare repository and through a public semantic database instance, and is interactively explorable through online Story Maps (*Data Records* and *Usage Notes*). To our knowledge, this is the first extensive and structured collection of information on European mountain value chains.

The Figshare collection tries to meet the FAIR (Findable, Accessible, Interoperable, Reusable) principles as far as possible. Figshare indeed fosters the alignment of the hosted data collections to data FAIRness[26]. The data have a unique and persistent Digital Object Identifier (DOI) assigned (Findable-F1 property). The collection's metadata comply with the DataCite metadata interconnection schema, and we fulfilled all essential and mandatory fields (Findable-F2). The metadata schema contains a dedicated "doi" field (Findable-F3) and is indexed in both the Figshare search engine (without authentication required) and the major search engines (Findable-F4). Moreover, we added textual metadata for each data folder and a data dictionary to improve data interpretability. The data and metadata are accessible for download, without authentication required. They comply with the "Attribution 4.0 International" licence (*CC BY 4.0*), i.e., they can be copied, redistributed, transformed, and reused even for

commercial purposes. Access is also guaranteed through several open authentication protocols (Accessible-A1), and the collection's metadata and DOI will be preserved for the repository's lifetime (Accessible-A2). The metadata are accessible through the Figshare APIs and are exportable (through the collection's page) to several standards (Interoperable-I1). They conform to controlled vocabularies of research categorisation and open licences (Interoperable-I2). The data vocabulary contains a controlled list of concepts belonging to ontological standards (Interoperable-I3). Finally, the metadata description, the open licence, the availability of the input and output data (complemented by provenance description through the present paper), and the use of a semantic knowledge graph for data representation strongly support our collection's reusability (Reusable-R1 and R2).

Paper and project background.   In the present paper, we describe how we built our knowledge graph for 454 European value chains. The primary source data were unstructured textual documents provided by territory experts working in the MOuntain Valorisation through INterconnectedness and Green growth (MOVING) European project[3]. MOVING was an H2020 project (September 2020 - August 2024) involving 23 organizations and companies that monitor, support, and conduct value chains in mountain areas. The primary project target was to collect updated and comparable knowledge on mountainous territories, with the conjecture that this would lead to a deeper understanding of the context, trends, and potential evolution of mountain communities, territories, and businesses. Moreover, this understanding would help design new policies for conservation and evolution. As a main strategy, the project proposed a bottom-up participatory process involving value chain actors, stakeholders, and policymakers to co-design European policy frameworks for resilience and sustainability. The heterogeneous MOVING community of practice monitored 454 value chains. In the first two project years (2020-2021), the territory experts studied and collected local knowledge about geography, traditions, and societal and economic aspects. Each expert independently compiled information on his/her monitored value chains. The provided information was complete from the point of view of the MOVING project scope. The experts used a socio-ecological system (SES) approach to understand the value chain contributions to the mountain areas' resilience and sustainable development. Within the SES framework, they related the value chain processes and activities to the local natural resources, particularly those affected by climate change and major socioeconomic and demographic trends (e.g., out-migration, livelihoods, and basic-service provisioning). They prioritised land-use and land-use change indicators because most value chains were agri-food and forestry-based, heavily relying on land resources. However, they also included other regional assets when particularly relevant for the region (e.g. hydropower in Scotland, Portugal and Romania; tourism in Italy, Portugal, Spain, Scandinavian countries, Serbia, North Macedonia, Romania, and Bulgaria). The SES approach was also justified by the MOVING project's focus on understanding the balance between economically valuable activities and environmental protection. Finding the right balance between these contrasting stressors will likely be more difficult in the near future due to the increasing number of European natural protected areas[27]. The possibility of analysing the vulnerabilities of mountainous value chains' environments, actors, resources, governance, and processes altogether was critical in this context, and could also support decision authorities in the design of multi-actor (public and private) institutional arrangements and multi-level (local, regional, national, and European) policies.

While this approach generated valuable and new knowledge, a side effect was the non-homogeneity of the collected information, e.g., administrative codes and statistical data were sometimes missing, and the investigated territory and value chain data often did not focus on the same assets across the studies. The need for managing this heterogeneous-knowledge scenario was the primary motivation for our study. After approval by the MOVING scientific community, we automatically transformed the unstructured, expert-provided data into a semantic knowledge graph. Here, we also demonstrate - through queries in the *SPARQL Protocol and RDF Query Language* (SPARQL) - that this representation allows for extracting new valuable knowledge for societal, economic, and environmental monitoring and studies.

Methods

The present paper outlines a semi-automated workflow developed to convert unstructured data about European value chains (**VCs**) into a semantic knowledge graph, as depicted in Fig. 1 and elaborated in the current section.

Our input was a set of textual documents, each detailing practical aspects of European VCs, including economic, meteorological, climatic, ecological, cultural, and societal aspects, along with specifications about their geographical regions and nations.

During the *data preparation* phase, these documents were processed to create a preliminary semi-structured form of the VC narratives, organized in tables with rows corresponding to narrative events. Then, a *data augmentation* phase regarded the extraction of information, for each event, about the mentioned places, locations, organizations, and keywords, and the enrichment of the data with geospatial references. This enriched and structured narrative representation was then converted into a semantic knowledge graph using the OWL format (*Knowledge graph creation and publication*).

This OWL-formatted knowledge graph was subsequently published in an openly accessible online semantic triple store and visually represented through 454 Story Maps. The OWL file, being the main output of this research, is available for other researchers for import into their semantic triple stores (*Usage Notes*). It allows them to explore the rich information about European value chains that the graph encapsulates.

Data preparation.   Our data collection originated from textual documents on VCs written by territory experts (researchers, members of local authorities, non-governmental organisations, producers' and processors' cooperatives, Local Action Groups, extension services, and others) within the MOVING European project[3]. Each involved country had from 1 to 51 documents associated (Table 1). The textual documents coarsely followed one textual-data collection schema designed by researchers at the University of Pisa (UniPi), who were involved in the MOVING project. As a preliminary validation, the UniPi researchers checked each expert's document for
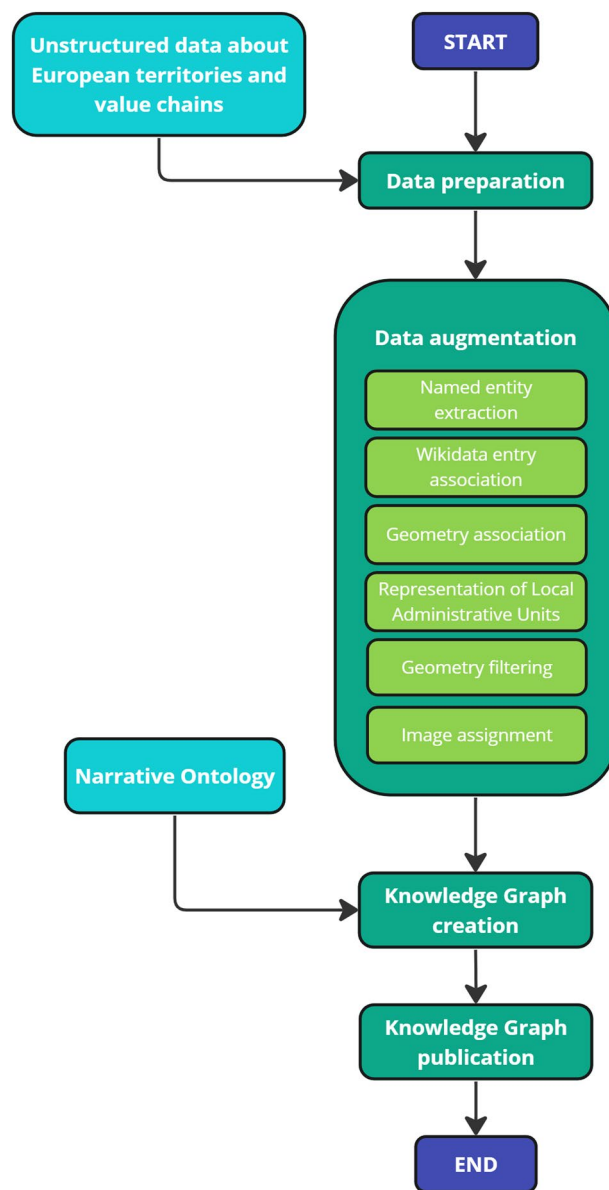
Fig. 1 Conceptual flowchart of our data preparation, augmentation, validation, and publication workflow.

inconsistencies in geographical locations, primary resources, and socioeconomic assets of the reference area and value chain. In the case of inconsistencies identified, they sent the document back to the expert(s) for adjustments, and repeated the checks on the updated document afterwards.

As an additional pre-processing step, we organised the information in the VC documents through an MS Excel table. This table contained one row for each VC and the columns corresponded to different VC aspects (Table 2). Some columns contained numeric values (e.g., for incomes and tourism). Other columns contained descriptions in natural language (e.g., the landscape description) or categorised information (e.g., Local Administrative Units). The table was very sparse since information on several columns was often unavailable. This table aimed to provide a first overview of the commonalities and heterogeneity between the VCs across European countries and regions. This file was the only manually prepared dataset of our workflow and the basis of the narrative building and augmentation part described in the next section. The MOVING project experts were also asked to check whether the MS Excel table correctly reported and represented the information they had provided.

As a further pre-processing step, we processed the MS Excel table to produce new tables in Comma Separated Value (CSV) file-format, one for each VC. Each CSV table was a rough, structured version of a VC (*VC table*). Our Figshare repository contains these files for consultation (*Data Records*). Each VC table contained 11 rows corresponding to the key events of a VC narrative (Table 3-right-hand column). Each row corresponded to one narrative *event*, with a *title* and a *description* column associated. To build the VC tables from the MS Excel table, we implemented a JAVA process that automatically mapped the column contents of one row of the MS Excel table onto the *description* column of one VC table. Table 3 reports this mapping. For one-to-one mappings, we

Country	Number of Value Chains
Austria	21
Bulgaria	20
Czech Republic	20
France	31
Greece	21
Germany	1
Hungary	23
Italy	46
North Macedonia	23
Portugal	21
Romania	20
Scotland	25
Serbia	20
Scandinavia	20
Slovakia	20
Slovenia	1
Spain	50
Switzerland	51
Turkey	20

Table 1. Distribution of the number of value chains over the countries involved in the MOVING European project.

Natural description	Quantitative description	Valorisation of VC
Member state responsible for the region	Local Administrative Unit (LAU)	Name of the VC
Description of the mountain landscape	Size of the area	Description of the VC
Presence of Protected Areas	Min and max altimetry of LAU	Challenges of the VC
Other interesting information	Total population	Type of VC
Reference mountain chain	Population density	Type of VC innovation
Reference mountain landscape	Population change rate in the last 10 years	Description of the local assets
	Regional population density	Issues of interest
	Average per capita income (P/R)	Reason of selection
	Gross Value Added (GVA) (P/R)	
	Primary sector share of GVA/year (P/R)	
	Secondary sector share of GVA/year (P/R)	
	Tertiary sector share of GVA/year (P/R)	
	Primary sector share of employment/year (P/R)	
	Secondary sector share of employment/year (P/R)	
	Tertiary sector share of employment/year (P/R)	
	Bed places in tourist accommodations/year (LAU)	
	Regional share of Bed places/year	
	Number of agricultural holdings	

Table 2. Information associated with each sub-region and value chain (VC), organised into three categories. We use the notation *P* to indicate information available at a province scale and *R* for regional scale.

directly reported the source-column's text content. When multiple columns corresponded to the same VC event, instead, we appended the column contents through text-harmonisation rules for the conjunctions.

This mapping process produced 454 VC tables, which were the input to the subsequent augmentation phase.

Data augmentation. In the present section, we describe all data augmentation steps in our workflow and eventually report the corresponding algorithm pseudo-codes.

Named entity extraction. Our workflow used a named entity extraction module we implemented in JAVA. This module characterised each event in the VC narrative with abstract or physical objects mentioned in the event description texts (*named entities*). The module used the NLPHub service[28], a cloud computing service that coordinates and consolidates the results of various state-of-the-art text-mining processes integrated within the D4Science e-Infrastructure[29–31]. In our workflow, we set the NLPHub to identify entities of types *location*, *person*, and *organisation*, plus the *keywords* of the text. Keywords were individual words or compound terms particularly

MS Excel column	Corresponding VC table row
Name of the VC	Introduction
Description of the VC	
Member state responsible for the region	
Issues of interest	Interest on the Value Chain
Reason of selection	
Description of the mountain landscape	Mountain landscape and reference chains
Reference mountain chain	
Description of the local assets	Key local assets
Local Administrative Unit (LAU)	
Challenges of the VC	Challenges
Type of VC	Innovation
Type of VC innovation	
Size of the area	Geography and population
Min and max altimetry of LAU	
Total population	
Population density	
Population change rate in the last 10 years	
Regional population density	
Average per capita income (P/R)	Income and gross value added
Gross Value Added (GVA) (P/R)	
Primary sector share of GVA/year (P/R)	
Secondary sector share of GVA/year (P/R)	
Tertiary sector share of GVA/year (P/R)	
Bed places in tourist accommodations/year (LAU)	Tourism
Regional share of Bed places/year	
Number of agricultural holdings	
Presence of Protected Areas	
Primary sector share of employment/year (P/R)	Employment
Secondary sector share of employment/year (P/R)	
Tertiary sector share of employment/year (P/R)	
Other interesting information	Concluding remarks

Table 3. Mapping between the columns reported in our workflow-input MS Excel table and the columns of a value chain (VC) narrative table.

meaningful within their respective contexts. The NLPHub exploited the D4Science cloud computing platform (named DataMiner[31]) to efficiently manage the processing of ∼5000 event texts in our dataset via distributed and concurrent cloud processing. The named entity extraction module augmented each VC table with one additional column (*named entities*) reporting a comma-separated list of named entities (and keywords) associated with each event.

Wikidata entry association. We used the named entities extracted by the previous step as the input of queries to the Wikidata semantic service's SPARQL endpoint[32]. A JAVA process executed these SPARQL queries to Wikidata to check if each narrative-event entity could correspond to a Wikidata entry. One special rule was adopted for *location*-type entities. By convention, Wikidata describes *location*-type entries with the first letter capitalised. Our process used this convention to check for the existence of Wikidata entries associated with *location*-type named entities.

In the case of a correspondence found, the process extracted the entry's Wikidata's Internationalized Resource Identifier (IRI). The IRI is part of the information the Wikidata SPARQL response returns for an entry. It persists also after entry-content update. For instance, the "Alps" entity had the following Wikidata entry IRI associated: https://www.wikidata.org/wiki/Q1286 which corresponded to the Q1286 identifier.

As an additional step, our process checked the consistency of the Wikidata entry retrieved. In particular, it explored the entry-associated *Wikipedia* pages. For a Wikidata entry to be valid, its associated Wikipedia pages should not correspond to (i) a disambiguation page, (ii) a page with a title not matching the named entity, or (iii) a page referring to a different named entity type. For example, the Wikipedia page associated with a *location*-type named entity had to correspond to a location. This check distinguished cases like *Tours* (the French city) from *tours* (journeys in the area). These rules overall improved the precision of the association between a Wikidata entry and a named entity, i.e., a validated Wikidata entry likely had the same meaning as the named entity.

At the end of the Wikidata entry retrieval and consistency check, our workflow added one column (named *IRIs*) to every VC table. This column contained, for each event, the valid IRIs of the event's entities. Entities without a valid IRI associated were discarded because they brought the risk of introducing false topics in the narratives.

Geometry association. As an additional data augmentation step, a Python process added a new column (named *geometry*) to each VC table containing spatial representations for the *location*-type entities. The process checked each valid *location*-type entity for having a corresponding coordinate pair in the associated Wikidata entry. In particular, it retrieved the Wikidata "coordinate location" property (P625) content as a reference longitude-latitude coordinate pair. Moreover, the process also checked if a polygon was possibly associated with the entity. To this aim, it used an instance of the open-access QLever endpoint of the University of Freiburg[33] to retrieve a possible polygon representation from the OpenStreetMap subgraph included in this large knowledge graph. QLever is a SPARQL engine capable of efficiently indexing and querying large knowledge graphs (even with over 100 billion triples) such as Wikidata, Wikimedia Commons, OpenStreetMap, UniProt, PubChem, and DBLP[34]. The University of Freiburg populated a large knowledge graph with these sources. Our process reported all geometries found on the QLever service as Well-Known Text (WKT) formatted strings[35]. The first VC event (*Introduction*), was always assigned the country's polygon and centroid. Our process added the found entities' geometries to the *geometry* column of their associated events. It reported both the polygon and point representations when existing. All geometries reported by our workflow used the EPSG:4326 geodetic coordinate system for World (equirectangular projection).

Representation of Local Administrative Units. The expert-provided data also included the indications of the 2-level Local Administrative Units[36] (LAUs) of the municipalities covered by each VC (Table 2). A VC could span more than one municipality and often had several LAUs associated. Eurostat, the statistical office of the European Union, has been producing regional statistics for these areas since 2003[37]. Different LAUs can form one "Nomenclature of Territorial Unit for Statistics" (NUTS), for which Eurostat produces additional statistics. These statistics help assess trends for local community typologies (rural, suburban, and urban), urbanisation degree (city, town and suburb, rural), functional urban areas (cities and their surrounding commuting zones), and coastal areas.

Our workflow included a Python process to retrieve a geometry representation of the VC-associated LAUs (as WKT strings). The process searched for a polygonal representation of each LAU code in two structured files published by Eurostat in their Geographic Information System of the Commission (GISCO)0[38]. GISCO is an EU-funded geographic information system that includes data on administrative boundaries and thematic information (e.g., population data) at the levels of European nations and regions. The first GISCO file our process used was a GeoJSON file[39] containing all WKT polygon representations of the Eurostat-monitored LAUs. However, the experts often reported NUTS codes instead of LAU codes. Therefore, if a polygon representation could not be found for one LAU code, our process searched for the same code in a second GISCO GeoJSON file containing NUTS polygon representations[40]. Since different countries could use the same LAU and NUTS codes for different territories, our process used the VC's belonging country code (e.g., IT, ES, UK) for disambiguation.

Our process found correspondences for all LAU and NUTS codes associated with our VCs (1224 total). It augmented each VC table's geometry column with LAU (or NUTS) geometries repeated for each event. It represented all geometries with equirectangular projection, also used in GISCO.

Geometry filtering. The geometries associated with the VC narrative events were checked for "geographical consistency" with the narrative itself. A story map set in the Austrian Alps that mentioned a cow breed also found in America might lead to the inclusion of United States regions' entities (and thus geometries) in the story. From a narrative point of view, associating a point too distant from the VC territory would be dispersing and produce jittery paths on the map that could confuse the reader. Therefore, we decided to avoid shifts from one continent to another or between far locations in our narratives while keeping a geographically coherent focus.

A dedicated JAVA process estimated a bi-variate log-normal distribution on the longitude-latitude pairs of each narrative. It included the LAU/NUTS centroids among the pairs. The process computed the upper and lower 95% log-normal confidence limits on the coordinates and considered the coordinates outside these boundaries as outliers. Consequently, if most coordinates pertained to a specific region, the calculated boundaries naturally surrounded that region. Otherwise, the boundaries likely encompassed all coordinates if these were uniformly distributed worldwide (e.g., in a global-scale narrative). We demonstrated the validity of a bi-variate log-normal distribution to estimate the primary geographical focus of a narrative in a previous work[20]. Each event in our narrative underwent outlier removal using this log-normal filter. By construction, at least the LAU/NUTS geometries remained associated with an event after the filtering. All geometries associated to an event are reported on a map during the event visualisation in a Story Map.

Image assignment. As a final data augmentation step, our workflow assigned images to the 11 events of each VC narrative through a dedicated Python process. The image associated with the first event (*Introduction*) was always the geographical map of the VC-associated country. This map was retrieved from Wikidata through a SPARQL query on the "locator map image" property (P242). Quantitative events such as "Income and gross value added" and "Employment" were not associated with images because their images should necessarily be conceptual. However, we verified that the MOVING community did not perceive such conceptual images as meaningful. For the remaining events, we used images the MOVING project members willingly provided for each country (without copyright violation). Six images per country were averagely available, which we enriched with additional region-specific

images from Wikimedia Commons[41] referring to the VC territories. Our Python process randomly sampled images from the VC's country-associated image set (without repetitions) and assigned them to the narrative events while prioritising the expert-provided images. For example, the narrative "Chestnut flour coming from the rediscovered chestnut cultivation activities in the area" was enriched with seven images of Tuscany by the MOVING members and two images of the Apuan Alps (where this chestnut flour is produced) from Wikimedia Commons.

Algorithms. In the present section, we report the algorithms of the data augmentation processes described so far. The data augmentation algorithm for named-entity extraction and geometry association can be summarised as follows:

Algorithm 1 Data augmentation algorithm - named entity extraction.

---

for each event description and title

> invoke the NLPHub to extract all *location*, *person*, *organisation* named entities, plus the *keywords*
>
> for each extracted named entity
>
>> check validity (i.e., no ambiguity exists) with respect to Wikidata/Wikipedia entries
>>
>> if the entity is of *location* type, try to retrieve the associated latitude and longitude coordinates from the Wikidata entry and polygonal geometry using the external QLever SPARQL server of University of Freiburg
>
> collect the list of entities and coordinates associated to the event.

The algorithm translating LAU/NUTS codes into WKT strings is the following:

Algorithm 2 Data augmentation algorithm - LAU/NUTS geometry extraction.

---

for each event:

> extract the country code from the member state description
>
> extract the LAU code
>
> check if the code-country pair exists in the GeoJSON LAU GISCO file
>
> if a match was not found → check if the code-country pair exists in the GeoJSON NUTS GISCO file
>
> if a match was found:
>
>> extract the WKT-polygon representation and save it to the geometry column (with a LAU/NUTS label)

The geometry filtering algorithm can be summarised as follows:

Algorithm 3 Data augmentation algorithm - geometry filtering.

---

identify coordinate outliers through bi-variate 95% log-normal confidence intervals

for each event

> remove all geometries including the outlier coordinates
>
> if the event has geospatial entities associated → assign the geometries of all associated entities
>
> else → assign the coordinates of the LAUs/NUTS.

The image assignment algorithms can be summarised as follows:

Algorithm 4 Data augmentation algorithm - image assignment.

---

select the image pool associated with the narrative

for each event

> sample from the expert-provided image set (without repetition)
>
> if no new image was sampled, sample an image from the Wikimedia Commons image set (without repetition)
>
> assign the image to the event.

Knowledge graph creation and publication. Our workflow used an additional Python process to translate all augmented VC tables into a semantic knowledge graph. As a first step, this process translated the VC tables into JSON files that followed a schema we designed and optimised in a previous work[20,42]. This JSON representation structurally describes the event sequence and the associated entities, images, geometries, and Wikidata IRIs. Our process also stored each JSON file in a PostgreSQL-JSON database for quick retrieval and use for narrative visualisation (*Usage Notes*).

As a second step, the process translated each JSON file into a Web Ontology Language (OWL) graph file and assembled all graphs into one overall OWL graph file[43]. To this aim, it invoked a JAVA-based semantic triplifier software we implemented for this specific sub-task. The VC-individual and the overall OWL graphs complied with the OWL 2 Description Logic[44], which assured the decidability of the language. They adhered to the Narrative Ontology model version 2.0[7,45], extended with the GeoSPARQL ontology[14], a standard of the Open Geospatial Consortium that handles geospatially explicit data in ontologies. We published the entire VC-narrative OWL graph (and sub-graphs) in the Figshare repository attached to the present paper (*Data Records*) to openly allow users to import them in a semantic triple store and query, explore, and infer knowledge on the 454 European VCs represented. This file was the main output of our workflow.

We also published the knowledge graph on a public-access Apache Jena GeoSPARQL Fuseki[46] semantic triple store to allow users to execute semantic and geospatial queries to our knowledge graph openly (*Usage Notes*). Fuseki is a SPARQL service that can internally store Resource Description Framework (RDF) data representing semantic triples consisting of a subject (node), a predicate (relation) and an object (node) of the knowledge graph. This service allows retrieving the stored RDF triples through a SPARQL/GeoSPARQL endpoint.

It is important to stress that the main target of the knowledge graph was to enhance the communication about the value chains to a broad, heterogeneous audience. Our target stakeholders were value chain and territory experts, citizens, and local and national administrations. These stakeholders need an overall understanding of the value chains, their role in characterising the territory, and the criticalities and strengths of the territory. Entities such as locations, persons, organisations, and keywords - enriched with images and geometries - matched their interests. Consequently, we did not include statistics and numeric indicators among the entities because data analytics was not a target of the narratives. Moreover, the unavailability of statistical data for several value chains would have created knowledge gaps across the narratives. Therefore, we reported statistical data, when available, in an understandable and readable textual format in the event text while leaving the possibility to conduct data analytics on the tabular-format files available in the Figshare repository.

As an additional note, we clarify that semantic knowledge graphs were a more suitable choice for data representation than tabular data models. Tabular data models, such as those used in relational databases, satisfy a predefined schema. Coercing our data to rows and columns was unsuitable for quickly capturing the complex relationships between value chain entities. Moreover, tabular models hardly manage inconsistent naming conventions, formats, and ambiguous identifiers like those in our data. Although foreign keys allow for modelling rich, interconnected data, they introduce complexity in the database schema, making knowledge extraction more challenging. Moreover, as the volume of data grows, managing and querying large relational tables can become inefficient and require dedicated distributed systems. In a scenario like ours, where data were many, heterogeneous, and dynamic, we could not assume that a traditional relational schema was efficient and effective. Instead, we used Linked Data and Semantic Web technologies because they offered more flexibility in quickly extending, modifying, and interconnecting a knowledge base of diverse and heterogeneous data. Moreover, semantic graphs could intuitively represent rich and complex relationships between the data while capturing real-world facts. They also enacted interoperability through the reuse of shared vocabularies and IRIs from other semantic knowledge bases, allowing the creation of interconnected, consistent data networks. Finally, as semantic technologies are Web-native, they quickly allowed for accessing and querying data through standard Web protocols.

Data Records

We made the data available on a public-access Figshare repository[47] (version 3, currently). One dataset is available for downloading the overall OWL knowledge graph, which allows other users to reproduce the entire knowledge base in another triple store (*Usage Notes*). This knowledge graph contains 503,963 triples. The data in the graph are also available in CSV and GeoPackage formats for easy import, manipulation, and inspection in multiple software solutions. Another dataset presents a folder hierarchy containing sub-graphs focusing on one VC at a time. The folder organisation is optimised for regional ecological, socioeconomic, and agricultural modelling experts. The files are organised into subfolders, each corresponding to a country. The name of each file reports the title of the corresponding value chain. Each file is in the OWL format (e.g. wood_charcoal_from_Gran_Canaria_island.owl). The complete file collection contains 454 files that can be imported independently of each other. The Figshare repository also contains all links to the JAVA and Python software used in our workflow.

Additionally, the repository contains a direct link to our public-access Apache Jena GeoSPARQL Fuseki instance hosting the entire VC knowledge graph. This instance allows the execution of SPARQL and GeoSPARQL queries (*Usage Notes*). The Figshare repository also contains the MS Excel file that was the input of our workflow. It allows for comparing our workflow's original and final products and repeating the technical validation. Finally, the repository contains all VC tables in CSV format resulting from the data preparation phase. The authors received authorisation by the MOVING project community to publish this material.

Technical Validation

Formal consistency of the knowledge graph. We used a semantic reasoner to validate the logic consistency of our entire OWL graph. A semantic reasoner is software designed to infer logical consequences from a set of asserted facts or axioms. We used the Openllet open-source semantic reasoner[48–50] to (i) check the consistency of our OWL graph (i.e., to guarantee that it did not imply contradictions), (ii) check that the class hierarchy

respected the one of the Narrative Ontology, (iii) test geometry consistency (polygon closures and correct WKT formatting), (iv) test the possibility to execute complex SPARQL and GeoSPARQL queries.

Openllet assessed the consistency of our knowledge graph on all the checks reported above. The reasoner confirmed that the subclass relations and the complete hierarchy between the classes fully respected the ones of the Narrative Ontology. The class hierarchy allowed the correct extraction of all subclasses of a class. Finally, all geometries were assessed as consistent. GeoSPARQL queries allowed the execution of spatial reasoning and all algebraic operations between sample-selected polygonal geometries from the VCs.

Additionally, we executed automatic checks to ensure that no event in the OWL graph was empty or contained meaningless or misreported content (e.g., "N/A", "Empty", "Unavailable", etc.). The checks also verified that every LAU had an associated WKT polygon from the GISCO database, and we manually verified that the correspondences were correct. An expert from the MOVING project also conducted a sample check to verify that the mapping between the pre-processed MS Excel table columns and the VC-narrative events (Table 3) produced meaningful and human-readable descriptions.

Performance of the named entity extraction-filtering process. We evaluated the performance of the combined process of named entity extraction plus Wikidata entry association (filtering). This process influences the results of the queries to the knowledge graph. An incomplete set of entities would indeed limit the information associated with an event. Moreover, as highlighted in the next section, it would also limit the discovery of interconnections between the events. When querying for events with several associated entities, a target event would only be selected if all entities were retrieved correctly.

To measure the quality of the named entity extraction-filtering process, we evaluated its information-extraction performance on manually annotated story events. To this aim, we selected a statistically meaningful set of events to annotate through the *sample size determination* formula with *finite population correction*, i.e.

$$n_0 = \frac{Z^2 \cdot p \cdot (1 - p)}{MOE^2}$$

$$n = \left(\frac{n_0 \cdot N}{N + n_0 - 1} \right)$$

where n is the target sample size adjusted for a finite population; n_0 is the initial sample size assuming an infinite population; Z is the Z-score corresponding to the desired confidence level (Z-score $=1.96$ for the 95% confidence level we used); p is the prior assessment (0.5 for uninformative conditions); MOE is the margin of error on the true error (5%), and N is the total number of events (population size).

This formula estimated that 330 events (corresponding to 30 stories) could be sufficient for performance assessment. Consequently, we randomly selected 30 stories from our collection. In this sub-collection, we identified the automatically extracted-filtered entities correctly associated with key event-related concepts (*true positives*, TP). Then, we identified those unrelated to key concepts (*false positives*, FP). Finally, we annotated additional entities (with valid Wikidata pages associated) from the events' texts that the extraction-filtering process missed (*false negatives*, FN). Based on these annotations, we calculated the following standard performance measurements:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{(Precision + Recall)}$$

The evaluation results are reported in Table 4. The high Precision (0.99) suggests that most extracted entities were correctly associable with key concepts expressed in the events. Maximising the reliability of the extracted entities was indeed the principal target of our entity extraction-filtering process. Instead, the lower Recall (0.93) suggests that the extracted entity set could be incomplete and could negatively impact multi-entity querying. However, the F1 measure (0.96) was one of a good information extraction system. Therefore, although the extraction-filtering process could be improved (which will be part of our future work), its performance was sufficiently high to support our knowledge graph reasonably.

Query-based validation. We verified that our knowledge graph could contribute to discovering new knowledge from the data, which was its principal aim. In particular, we collected the types of interrogations the MOVING-project scientists or stakeholders considered valuable, i.e., hard to identify without a semantic knowledge representation. These interrogations were collected (i) during plenary meetings with the MOVING community of practice, (ii) after identifying the principal study targets of the rural-area experts involved in the

Precision	0.99
Recall	0.93
F1	0.96
Total entities analysed	∼1700
Total events analysed	330
Total stories analysed	30

Table 4. Precision, Recall, and F1 measurements of our named entity extraction-filtering process.

project (typically value chains within their territories), and (iii) by reading project deliverables. For example, the experts' targets were the VCs sharing common environmental characteristics (e.g. rivers, lakes, vineyards, and chestnut trees), issues (e.g. depopulation, pollution, and deforestation), and similar products (e.g. cow/sheep milk and cheese). Discovering this knowledge from the data holds significant value for mountain ecosystems, as it aids in planning sustainable environmental management strategies[21]. Additionally, this knowledge is valuable in supporting the long-term ecological sustainability of urban areas and comprehending and mitigating the decline of fundamental services in mountain areas brought about by the ongoing depopulation trends[2,51,52]. We demonstrated that our knowledge graph could contribute to these directions.

We focussed on ten types of knowledge-extraction targets, corresponding to ten SPARQL/GeoSPARQL queries regarding different and complementary aspects of European mountain products and their related spatial distributions. In particular, we extracted the VCs with the following characteristics:

1. related to vineyard products (Q1)
2. possibly affected by deforestation (Q2)
3. involving cheese with Protected Designation of Origin (PDO) certification (Q3)
4. producing cheese made with cow and/or goat milk (Q4)
5. using sheep to produce cheese (Q5)
6. using sheep to produce wool (Q6)
7. operating in the Alps (Q7)
8. operating around Aosta city (Italy) (Q8)
9. operating in Scotland (Q9)
10. operating around long Italian rivers (>100) (Q10)

The information extracted by these queries overall covered the interests of the MOVING community experts. It would have been hard, indeed, to extract the same information through the usual data representation and technology adopted by this scientific community. Based on the query results, we calculated Precision, Recall, and F1. We showed that high Precision was achieved for most cases, i.e., even when the information retrieved was incomplete (mostly due to misdetection by the named entity extraction processes) the results were reliable. The performance measurements (Table 5) demonstrate the overall high quality of our knowledge graph and the general effectiveness of the queries. In the following, we report the details of the queries and the corresponding results.

Q1 - Value chains related to vineyard products
In the following, we report the SPARQL query corresponding to Q1.

SPARQL Query 1

```
PREFIX narra:<https://dlnarratives.eu/ontology#>
PREFIX ecrm:<http://erlangen-crm.org/current/>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>

SELECT DISTINCT?title?country
WHERE {
  ?event1 narra:partOfNarrative?narrative.
  ?narrative rdfs:label?title.
  ?narrative narra:isAboutCountry? countryIRI.
  ?countryIRI rdfs:label?country.
  {
    ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q22715>.
  }
UNION
  {
    ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q282>.
  }
UNION
  {
    ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q10978>.
```

VC number	Precision	Recall	F1
Q1	0.93	0.90	0.91
Q2	1	1	1
Q3	1	0.26	0.41
Q4	0.84	0.55	0.66
Q5	0.69	0.72	0.70
Q6	0.69	0.75	0.72
Q7	1	1	1
Q8	1	1	1
Q9	1	1	1
Q10	1	1	1

Table 5. Precision, Recall, and F1 measurements of our queries.

```
    }
    UNION
    {
        ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q22715>.
        ?event2 narra:partOfNarrative?narrative.
        ?event2 narra:hasEntity <https://dlnarratives.eu/resource/Q282>.
        FILTER (?event1!=?event2)
    }
    UNION
    {
        ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q22715>.
        ?event2 narra:partOfNarrative?narrative.
        ?event2 narra:hasEntity <https://dlnarratives.eu/resource/Q10978>.
        FILTER (?event1!=?event2)
    }
    UNION
    {
        ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q22715>.
        ?event2 narra:partOfNarrative?narrative.
        ?event2 narra:hasEntity <https://dlnarratives.eu/resource/Q282>.
        ?event3 narra:partOfNarrative?narrative.
        ?event3 narra:hasEntity <https://dlnarratives.eu/resource/Q10978>.
        FILTER (?event1!=?event2 & &?event1!=?event3 & &?event2!=?event3)
    }
}
ORDER BY lcase(?country)
```

The query retrieves distinct titles of narratives along with their associated countries. Several ontology prefixes are specified at the beginning of the query to shorten the corresponding IRIs, which are used in the subsequent parts of the query. The *SELECT* statement specifies the variables ("?title" and "?country") whose values the query will output. The *WHERE* clause contains the conditions that need to be satisfied for each result. It involves several semantic-triple patterns connected by the "." operator:

- The first triple pattern ("?event1 narra:partOfNarrative?narrative") connects an event to a narrative;
- The second triple pattern ("?narrative rdfs:label?title") retrieves the label (title) of the narrative.
- The third triple pattern ("?narrative narra:isAboutCountry?countryIRI") connects the narrative to its related country.
- The fourth triple pattern ("?countryIRI rdfs:label?country") retrieves the label (name) of the country.
- The subsequent *UNION* blocks combine pattern alternatives, each representing a condition under which events are selected. The blocks retrieve events associated with at least one entity among "vineyard" (id. Q22715), "wine" (id. Q282), and "grape" (id. Q10978). These entities were chosen with the help of an expert. They are the entities most related to vineyards in our knowledge graph. The expert was aided by an entity search tool included in our visualisation facilities (*Usage Notes*).
- The sets of narrative events containing the entities above are labelled "event1", "event2", and "event3", respectively. Filters are applied (e.g., "FILTER (?event1!=?event2)") to ensure that the entities can singularly appear in different events.

Finally, the *ORDER BY* clause sorts the results alphabetically by the lowercase label of the country. In the case of multiple sub-graphs imported instead of the overall graph, the query should be changed by adding a "FROM <urn:x-arq:UnionGraph>" clause, before the WHERE clause, to specify that the query should be conducted on the union of the sub-graphs.

Title	Country
Wines of Kitzek im Sausal, South Styria	Austria
Rosehip wine	Czech Republic
Clairette de Die - Wine production targeted towards export	France
Organic still wine production	France
Savoi - the Reconquest of the winemakers	France
Wines of Winningen, Mosel Valley	Germany
Activities connected to the production of wine in the PDO Peza zone, at eastern Crete	Greece
Production of traditional distilled spirit drinks from grape pomance 40%-65% ABV	Greece
Gastro village Köveskál Eat the View	Hungary
Nivegy-valley small family wineries	Hungary
Organic farming and wine production in Vadna	Hungary
'Wine and Tourism 2' (Liguria-5Terre)	Italy
Alto Trentino, organic wine production for premium wines	Italy
Beekeeping and the Amatrice honey certified	Italy
Central Appenninian (Abruzzo)	Italy
Colli Piacentini	Italy
Grapevine on terraces- Valtellina	Italy
Growing of organic chestnuts and processing of chestnuts (e.g. flour) by a historical family farm	Italy
Island Wine from Lipari, Aeolian Islands	Italy
Volcanic Wine (Sicily-Etna)	Italy
Wine and Tourism 1	Italy
Wine - Vranec	North Macedonia
DO Douro Wine - Douro Superior - Vila Nova de Foz CÃ´a	Portugal
Wine from Serra da Estrela, one of the subregions of Dão Demarcated Region	Portugal
Vipava valley orange wines	Slovenia
Spanish Vigneronsfrom Huesca Pirineos	Spain
Traditional local sweets and pastries	Spain
Wines of the Contraviesa	Spain
Merlot from Alps - Ticino	Switzerland
Fresh Fruit (Grapes)	Turkey

Table 6. Results of Query 1: Value chains related to vineyard products included in our Knowledge Graph.

In summary, this query retrieves the titles of the narratives and their associated countries, comprising events related to the "vineyard", "wine", and "grape" entities. The query produced the output reported in Table 6. To verify the correctness of the retrieved information, we manually checked, with the help of a MOVING expert, the VCs (among the 454) that contained information on vineyard products. Precision and Recall (0.93 and 0.90, respectively), were reasonably high, and F1 (0.91) indicated good retrieval performance. The main reason for Recall not reaching one was the presence of faults (false negatives) by the named entity extraction processes in detecting vineyard-related entities in the event texts. Precision was, instead, negatively affected by citations of vineyard-related products in VCs that did not focus on vineyard products (false positives).

Q2 - *Value chains possibly affected by deforestation*
In the following, we report the SPARQL query corresponding to Q2.

SPARQL Query 2

```
PREFIX narra:<https://dlnarratives.eu/ontology#>
PREFIX ecrm:<http://erlangen-crm.org/current/>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
SELECT DISTINCT?title?country
WHERE {?event narra:partOfNarrative?narrative.
?narrative narra:isAboutCountry? countryIRI.
    ?countryIRI rdfs:label?country.
    ?narrative rdfs:label?title.
    ?event narra:hasEntity <https://dlnarratives.eu/resource/Q169940>.
    }
order by lcase(?country)
```

Similarly to Q1, this query retrieves the titles and associated countries of the narratives mentioning deforestation. The notable difference compared to Q1 is in the *WHERE* clause, which retrieves the events ("?event

Title	Country
Firewood for households	North Macedonia
Forestry seed and seedlings	North Macedonia
Pine cones - Dry pine cones with purpose of decoration	North Macedonia
Round wood for processing	North Macedonia
ECO-FERTILISER (WOOL PELLETS)	Romania
Berries cultivated in Kopaonik Mountain/ National park, Dinaric Alps region; other mountain areas	Serbia
Wild mushrooms collected in the mountainous area of Kopaonik mountain	Serbia

Table 7. Results of Query 2: Value chains possibly affected by deforestation included in our Knowledge Graph.

narra:hasEntity <https://dlnarratives.eu/resource/Q169940>”) having “deforestation” (id. Q169940) among the associated entities.

This query produced the result reported in Table 7. Expert verification assessed that it retrieved the complete and correct set (Precision and Recall equal to 1) of all VCs affected by deforestation. Therefore, this query shows the value of our knowledge graph for discovering critical threats to the VCs and their territories.

Q3 - *Value chains involving cheese with Protected Designation of Origin certification*
In the following, we report the SPARQL query corresponding to Q3.

SPARQL Query 3

```
PREFIX narra: <https://dlnarratives.eu/ontology#>
PREFIX ecrm: <http://erlangen-crm.org/current/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT DISTINCT?title?country
WHERE {
  ?event1 narra:partOfNarrative?narrative.
  ?narrative rdfs:label?title.
  ?narrative narra:isAboutCountry?countryIRI.
  ?countryIRI rdfs:label?country.
    {
    ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q10943>.
    ?event2 narra:partOfNarrative?narrative.
    ?event2 narra:hasEntity <https://dlnarratives.eu/resource/Q13439060>.
  }
  UNION
  {
    ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q10943>.
    ?event2 narra:partOfNarrative?narrative.
    ?event2 narra:hasEntity <https://dlnarratives.eu/resource/Q13439060>.
    FILTER (?event1!=?event2)
  }
}
ORDER BY lcase(?country)
```

The query structure is similar to the one of Q1, with the difference that the entities “cheese” (id. Q10943) and “Protected designation of origin” (PDO) (id. Q13439060) are used to detect events, and consequently, the target VCs.

The query produced the results reported in Table 8. This case is peculiar because it demonstrates the potential bottleneck of the performance of the named entity extraction processes. Although the query did not produce false positives (i.e., Precision was 1), there were many false negatives due to frequently missed recognition of PDO mentions in the event texts (Recall was 0.26). One reason is that long and articulated entities like “Protected designation of origin” are often subject to misspelling, abbreviation, and native-language reporting (e.g., DOP, in Italian), which prevent algorithms from identifying them. Therefore, Q3 showed a potential limitation of our knowledge graph when searching articulated entities. However, the same complexity of these entities ensured that the results were correct when the entities were identified.

Q4 - *Value chains producing cheese made with cow and/or goat milk*
In the following, we report the SPARQL query corresponding to Q4.

SPARQL Query 4

```
PREFIX narra: <https://dlnarratives.eu/ontology#>
```

Title	Country
This cooperative produces traditional cheeses, mainly for export, using only local sheep's milk	Italy
Traditional sheep's milk cheese	Portugal
Traditional sheep's milk cottage cheese	Portugal
PDO cheese made from raw and whole cow's milk.	Switzerland

Table 8. Results of Query 3: Value chains involving cheese with Protected Designation of Origin certification included in our Knowledge Graph.

```
PREFIX ecrm: <http://erlangen-crm.org/current/>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>

SELECT DISTINCT?title?country
WHERE {
  ?event1 narra:partOfNarrative?narrative.
  ?narrative rdfs:label?title.
  ?narrative narra:isAboutCountry?countryIRI.
  ?countryIRI rdfs:label?country.
  {
    ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q10943>.
    ?event2 narra:partOfNarrative?narrative.
    ?event2 narra:hasEntity <https://dlnarratives.eu/resource/Q2934>.
}
UNION
{
    ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q830>.
    ?event2 narra:partOfNarrative?narrative.
    ?event2 narra:hasEntity <https://dlnarratives.eu/resource/Q2934>.
}
UNION
{
    ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q11748378>.
    ?event2 narra:partOfNarrative?narrative.
    ?event2 narra:hasEntity <https://dlnarratives.eu/resource/Q2934>.
}
UNION
{
    ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q10943>.
    ?event2 narra:partOfNarrative?narrative.
    ?event2 narra:hasEntity <https://dlnarratives.eu/resource/Q11748378>.
}
UNION
{
    ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q10943>.
    ?event2 narra:partOfNarrative?narrative.
    ?event2 narra:hasEntity <https://dlnarratives.eu/resource/Q830>.
}
UNION
{
    ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q10943>.
    ?event2 narra:partOfNarrative?narrative.
    ?event2 narra:hasEntity <https://dlnarratives.eu/resource/Q11748378>.
    FILTER (?event1!=?event2)
}
UNION
{
    ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q10943>.
    ?event2 narra:partOfNarrative?narrative.
    ?event2 narra:hasEntity <https://dlnarratives.eu/resource/Q2934>.
    FILTER (?event1!=?event2)
}
UNION
{
    ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q10943>.
    ?event2 narra:partOfNarrative?narrative.
    ?event2 narra:hasEntity <https://dlnarratives.eu/resource/Q830>.
    FILTER (?event1!=?event2)
```

Title	Country
Almenland mining tunnel cheese (Almenland Stollenkäse)	Austria
Goats - dairy products	Czech Republic
Sheep - dairy products	Czech Republic
Picodon AOP	France
Epirus Feta, a PDO cheese made of sheep and goat milk in the Epirus region, Ammotopos Arta	Greece
Production of Graviera Kritis cheese in the area Amari, on the western slopes of Psiloritis mountain	Greece
Nivegy-valley artisan chese makers	Hungary
This farm produces high quality certified milk and cheese in a mountain nature park	Italy
Organic Farming	Norway
PGI (Protected Geographical Indication) Goat cheese / Undredal Brown Cheese	Norway
Pultost cow cheese	Norway
'BRÂNZĂ DE BURDUF'	Romania
'BRÂNZĂ HOREZU'	Romania
'MOZZARELLA'	Romania
Dried saussage made of localy produced sheep, goat, beef or donkey meat mixture, PDO application	Serbia
White cheese in brine, produced of mixed milk - cow and sheep milk	Serbia
Suckling goat from Malaga	Spain
Cow's milk that has not been pasteurised, sterilised or thermised and is sold directly	Switzerland
Goat's milk cheese made by hand directly on the mountain pasture	Switzerland
GruyÃ¨re is a typical Swiss cheese, produced in the mountain pastures in summer	Switzerland
Jogitz yogurt made from Goat's milk (several flavours)	Switzerland
PDO cheese made from raw and whole cow's milk.	Switzerland
alpine pasture cheese with organic sheep's milk	Switzerland
A traditional local cheese from goat and sheep milk	Turkey
Special kind of tulum cheese from goat and sheep milk	Turkey
Mountain region hard white cows milk cheese with PGI from Caerphillly LAU1 region within Wales	United Kingdom

Table 9. Results of Query 4: Value chains producing cheese made with cow and/or goat milk included in our Knowledge Graph.

```
    }
    UNION
    {
        ?event1 narra:hasEntity <https://dlnarratives.eu/resource/Q10943>.
        ?event2 narra:partOfNarrative ?narrative.
        ?event2 narra:hasEntity <https://dlnarratives.eu/resource/Q11748378>.
        ?event3 narra:partOfNarrative ?narrative.
        ?event3 narra:hasEntity <https://dlnarratives.eu/resource/Q2934>.
        ?event4 narra:partOfNarrative ?narrative.
        ?event4 narra:hasEntity <https://dlnarratives.eu/resource/Q830>.
        FILTER (?event1!=?event2 && ?event1!=?event3 && ?event2!=?event3
        && ?event1!=?event4 && ?event2!=?event4 && ?event3!=?event4)
    }
}
ORDER BY lcase(?country)
```

This query operates a search for narratives bound on four entities: "cheese" (id. Q10943), "cow" (id. Q11748378), "goat" (id. Q2934), and "cattle" (id. Q830). The query structure is similar to Q1.

The query produced the results reported in Table 9. The results were still affected by the named entity extraction bottleneck because the query's success depended on the correct identification of all four terms in a narrative. Compared to Q3, the present query tested the retrieval of multiple, simpler terms. Precision (0.84) and Recall (0.55) were indeed higher than the one of the articulated-entity search of Q3 (0.26).

Q5-Q6 - Value chains using sheep to produce cheese vs wool
In the following, we report the SPARQL queries corresponding to Q5 and Q6.

SPARQL Query 5 - VCs using sheep to produce cheese

```
PREFIX narra: <https://dlnarratives.eu/ontology#>
PREFIX ecrm: <http://erlangen-crm.org/current/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

```
SELECT DISTINCT?title?country
WHERE {
  ?event1 narra:partOfNarrative?narrative.
  ?narrative rdfs:label?title.
  ?narrative narra: isAboutCountry?countryIRI.
  ?countryIRI rdfs:label?country.
    {
      ?event1 narra:hasEntity<https://dlnarratives.eu/resource/Q7368>.
      ?event2 narra:partOfNarrative?narrative.
      ?event2 narra:hasEntity<https://dlnarratives.eu/resource/Q10943>.
  }
  UNION
  {
      ?event1 narra:hasEntity<https://dlnarratives.eu/resource/Q7368>.
      ?event2 narra:partOfNarrative?narrative.
      ?event2 narra:hasEntity<https://dlnarratives.eu/resource/Q10943>.
      FILTER (?event1!=?event2)
  }
}
ORDER BY lcase(?country)
```

SPARQL Query 6 - VCs using sheep to produce wool

```
PREFIX narra: <https://dlnarratives.eu/ontology#>
PREFIX ecrm: <http://erlangen-crm.org/current/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT DISTINCT?title?country
WHERE {
  ?event1 narra:partOfNarrative?narrative.
  ?narrative rdfs:label?title.
   ?narrative narra:isAboutCountry?countryIRI.
  ?countryIRI rdfs:label?country.
  {?event1 narra:hasEntity<https://dlnarratives.eu/resource/Q7368>.
    ?event2 narra:partOfNarrative?narrative.
    ?event2 narra:hasEntity<https://dlnarratives.eu/resource/Q42329>.
  }
  UNION
  {
      ?event1 narra:hasEntity<https://dlnarratives.eu/resource/Q7368>.
      ?event2 narra:partOfNarrative?narrative.
      ?event2 narra:hasEntity<https://dlnarratives.eu/resource/Q42329>.
      FILTER (?event1!=?event2)
  }
}
ORDER BY lcase(?country)
```

These queries have the same structure as Q1. They share one entity, "sheep" (id. Q7368), with two different usages (corresponding to different entities used in the *WHERE* clause), i.e., "cheese" (id. Q10943) in Q5 and "wool" (id. Q42329) in Q6.

The results are reported in Tables 10 and 11. The performance measurements between the two queries were very similar. The false positives, which affected Precision, were due to mentions of sheep in other VCs that did not regard usage for cheese or wool production. Notably, although the two queries retrieved mostly different VCs, the fraction of correct narratives retrieved (Precision) was 0.69 in each case. Moreover, Recall values (0.72 and 0.75, respectively) were similar, due to similar fractions of undetected mentions (false negatives) of cheese and wool by the named entity extraction processes. Overall, a ∼0.70 F1 for both queries indicated an overall moderate-high reliability of the results.

Q7 - Value chains operating in the Alps

In the following, we report the GeoSPARQL query corresponding to Q7.

GeoSPARQL Query 7

```
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#>
```

Title	Country
Almenland mining tunnel cheese (Almenland Stollenkäse)	Austria
'FOOD FROM THE MOUNTAIN' FARMERS' ASSOCIATION	Bulgaria
Epirus Feta, a PDO cheese made of sheep and goat milk in the Epirus region, Ammotopos Arta	Greece
Production of Graviera Kritis cheese in the area Amari, on the western slopes of Psiloritis mountain	Greece
This cooperative produces traditional cheeses, mainly for export, using only local sheep's milk	Italy
Cheese (North Macedonia)	North Macedonia
Lamb meat from an autochthonous sheep breed	Portugal
Products made from Burel, a specific wool fabric	Portugal
Traditional sheep's milk cheese	Portugal
Traditional sheep's milk cottage cheese	Portugal
Wool Products from Bordaleira Serra da Estrela autochthonous sheep breed	Portugal
'BRÂNZĂ, DE BURDUF'	Romania
'MOZZARELLA'	Romania
'TELEMEA'	Romania
Dried saussage made of localy produced sheep, goat, beef or donkey meat mixture, PDO application	Serbia
Fresh lamb meat, PDO (whole animals), produced from Sjenica sheep - local autochthonous sheep breed	Serbia
Pirot hard cheese from sheep and cow milk	Serbia
White cheese in brine, produced of mixed milk - cow and sheep milk	Serbia
The labelled traditional cheese products from sheep or from a mixture of sheep and cow milk	Slovakia
Lamb Meat from Zuheros	Spain
A traditional local cheese from goat and sheep milk	Turkey
Erzincan Tulum Cheese	Turkey
Special kind of tulum cheese from goat and sheep milk	Turkey

Table 10. Results of Query 5: Value chains using sheep to produce cheese.

Title	Country
Osttiroler Mountain Lamb (Osttiroler Berglamm)	Austria
Sheep wool	Czech Republic
Lamb meat from an autochthonous sheep breed	Portugal
Products made from Burel, a specific wool fabric	Portugal
Traditional sheep's milk cheese	Portugal
Traditional sheep's milk cottage cheese	Portugal
Wool Products from Bordaleira Serra da Estrela autochthonous sheep breed	Portugal
ECO-FERTILISER (WOOL PELLETS)	Romania
Wool handmade carpets with traditional motives	Serbia
Sheep's wool processing helps maintain traditional sheep farming in Slovakia	Slovakia
Grazalema wool blankets, textile craftsmanship	Spain
Promotion of the production, harvesting and processing of wool from the Swiss Jura region by an association	Switzerland
Mountain wool for Harris tweed - Western Isles (LAU1)	United Kingdom

Table 11. Results of Query 6: Value chains using sheep to produce wool.

```
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?wktLau
WHERE {
    ?narra narra:isAboutCountry?country;
            narra:isAboutLAU?lau;
            rdfs:label?nlabel.
    ?country rdfs:label?clabel.
    ?lau geo:hasGeometry?glau.
    ?glau geo:asWKT?wktLau.
    {
     SELECT?wkt WHERE {
        SERVICE {
        <https://qlever.cs.uni-freiburg.de/api/osm-planet>{
```

```
                    ?osm_id osm2rdfkey:wikidata wd:Q1286;
                              geo:hasGeometry?geometry.
                    ?geometry geo:asWKT?wkt.
                }
            }
        }
    FILTER(geof:sfIntersects(?wktLau,?wkt)).
}
```

The query retrieves the VC narrative titles, countries, and LAU polygons that overlap a polygon defining the Alps region. A value chain's LAUs define the main areas where the VC operates (i.e., produces and sells products). The query internally calls the QLever endpoint provided by the University of Freiburg (Section *Geometry association*), and in particular, the Open Street Map ("oms") subgraph, to define the Alps polygonal region. The *SELECT* statement specifies the variables "nlabel" (narrative title), "clabel" (country name) and "wktLau" (LAU geometry in WKT format) that will be the output of the query. The *WHERE* clause contains the conditions that should be satisfied by each result. Differently from the previous queries, the following patterns are included:

- The triple pattern "?narrative narra:isAboutLAU?lau" connects a narrative to the corresponding LAU;
- the triple pattern "?lau geo:hasGeometry?glau" retrieves the geometry of the LAU;
- the triple pattern "?glau geo:asWKT?wktLau" retrieves the WKT description of the LAU geometry;

A nested *SELECT* clause retrieves the WKT description ("?wkt") under the following *WHERE* conditions:

- The SERVICE keyword is used to invoke the external QLever endpoint ("https://qlever.cs.uni-freiburg.de/api/osm-planet");
- The triple pattern "?osm_id osm2rdfkey:wikidata wd:Q1286" retrieves the instance corresponding to the QLever entity "Alps" (wd:Q1286);
- The triple pattern "?osm_id geo: geometry?geometry" retrieves the geometry-object of "Alps";
- The triple pattern "?geometry geo:asWKT?wkt" retrieves the WKT format of the "Alps" geometry.

A final *FILTER* clause operates the intersection between the LAU and the "Alps" geometries and retrieves all LAU geometries intersecting "Alps". The set of LAUs returned by this query can be imported into a Geographic information system (GIS) visualiser and overlapped with the reference region (Fig. 2).

The expert's evaluation highlighted that the LAUs this query retrieved were correct and complete (Precision and Recall were 1). Therefore, the query was valuable in retrieving region-specific VCs.

Q8 - Value chains operating around Aosta city (Italy)
In the following, we report the GeoSPARQL query corresponding to Q8.

GeoSPARQL Query 8

```
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#>
SELECT?nlabel?clabel?wktLau
WHERE {
        {
    ?narra narra:isAboutCountry?country;
            narra:isAboutLAU?lau;
            rdfs:label?nlabel.
    ?country rdfs:label?clabel.
    ?lau geo:hasGeometry?glau.
    ?glau geo:asWKT?wktLau.
    }
    FILTER(geof:sfIntersects(
        ?wktLau,
        geof:buffer(
            "POINT(7.3196649 45.7370885)"^^geo:wktLiteral,
        0.3, uom:degree))).
}
```

This query extracts the VC titles, countries and LAU geometries of the value chains operating within a maximum distance of 23 km from Aosta. The query structure is similar to the one of Q7, with the difference that it does not use an external endpoint to retrieve the reference geometry. Instead, the FILTER clause operates an intersection between all VCs' LAU geometries and a circular buffer of 0.3 degrees (~23 km) around the Aosta longitude-latitude coordinates.
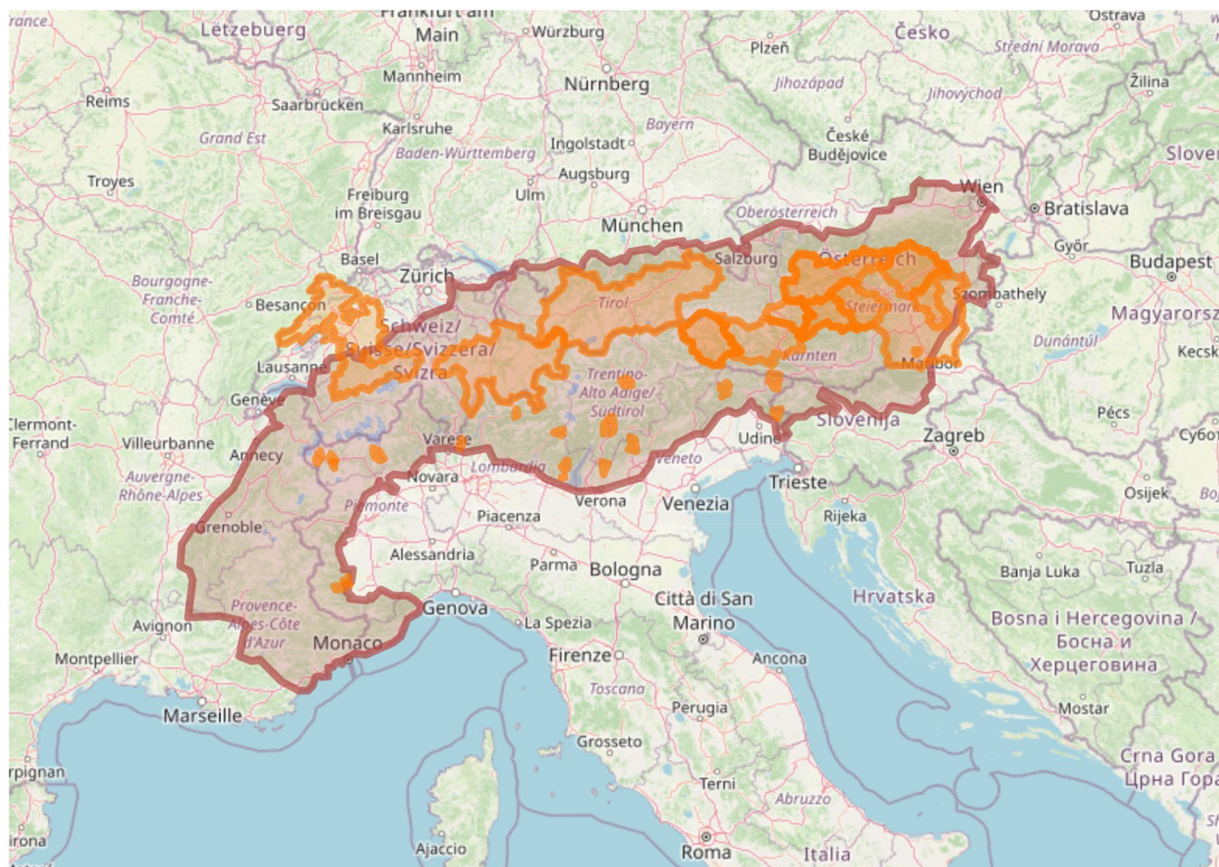
Fig. 2 Comparison between the Alps regions (as a red polygon) and the Local Administrative Units (orange polygons) of the value chains operating in this region. An interactive map is also reported in the Figshare repository associated with the present article.

The query produced the results visualised in Fig. 3. As in the case of Q7, the expert's evaluation highlighted that the LAUs retrieved by this query were correct and complete (Precision and Recall were 1). Therefore, the query was valuable in retrieving city-specific VCs.

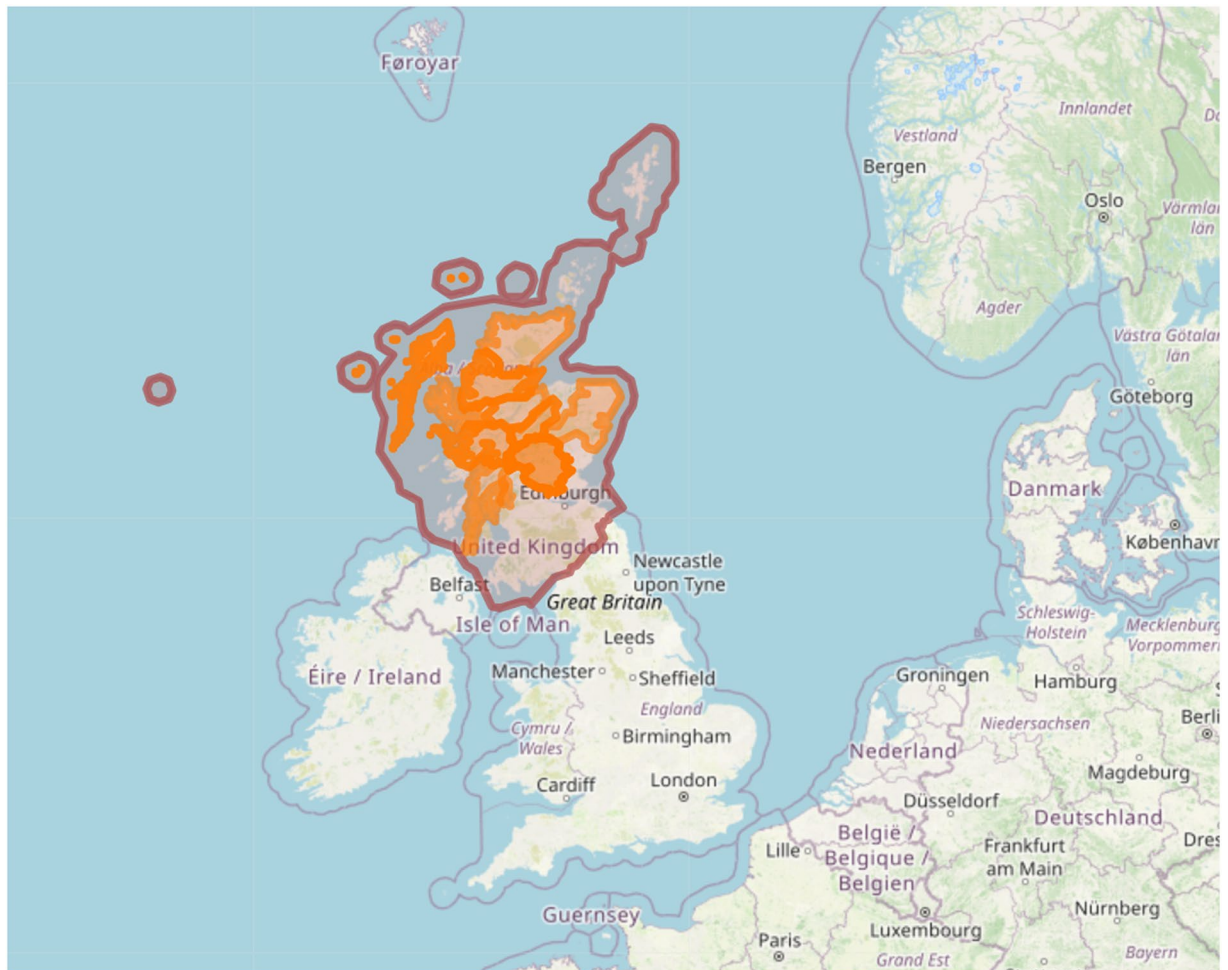### Q9 - Value chains operating in Scotland
In the following, we report the GeoSPARQL query corresponding to Q9.

GeoSPARQL Query 9

```
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 osmrel: <https://www.openstreetmap.org/relation/>
PREFIX schema: <http://schema.org/>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX osm2rdfkey: <https://osm2rdf.cs.uni-freiburg.de/rdf/key#>
SELECT?nlabel?clabel?wktLau
WHERE
    {
        ?narra narra:isAboutCountry?country;
                narra:isAboutLAU?lau;
                rdfs:label?nlabel.
        ?country rdfs:label?clabel.
        ?lau geo:hasGeometry?glau.
        ?glau geo:asWKT?wktLau.
        {SELECT?wkt WHERE {
```
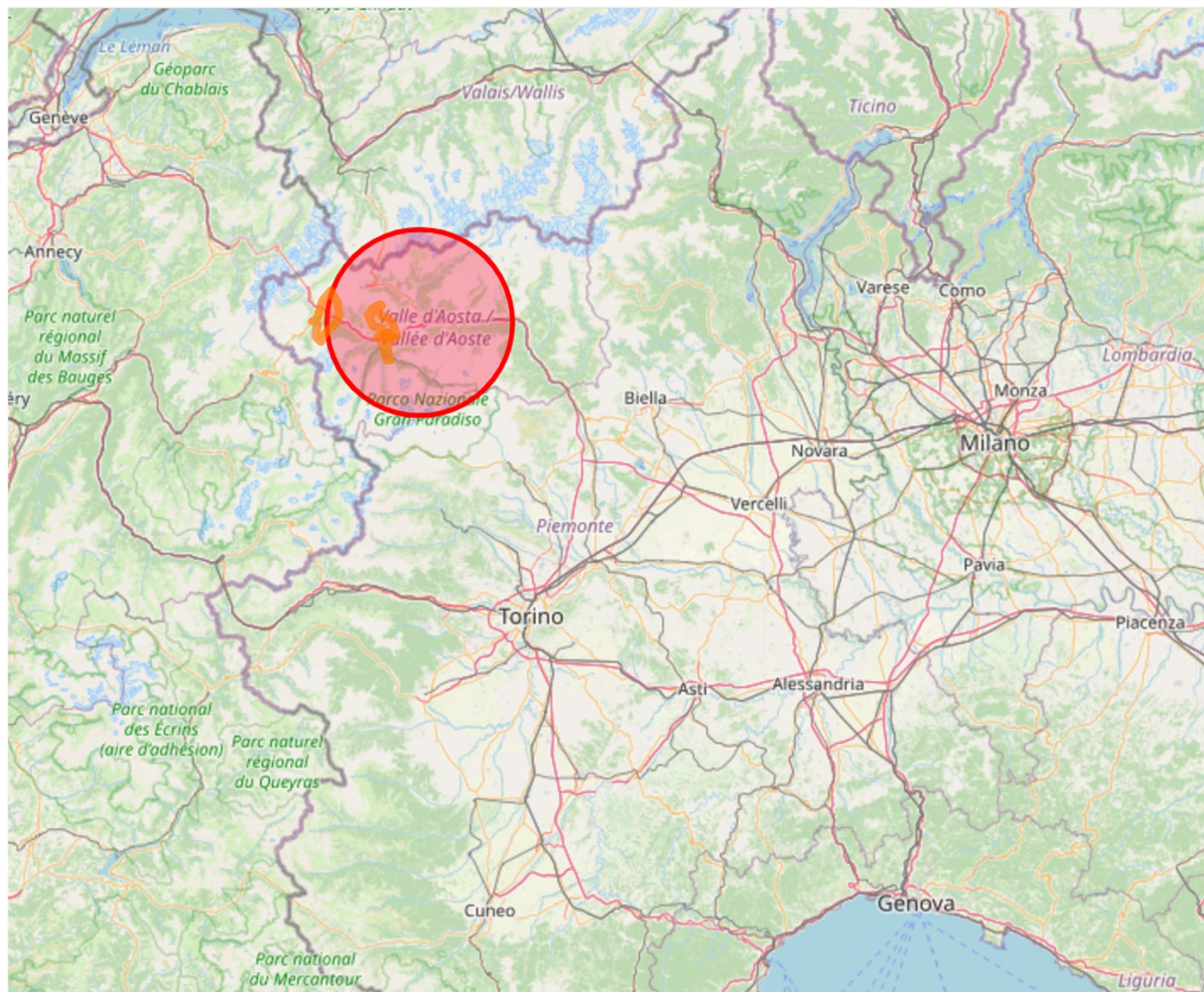
Fig. 3 Highlight of the Local Administrative Units (orange) of the value chains operating in a circle denoting a 0.3-degree area around Aosta, Italy (red). An interactive map is also reported in the Figshare repository associated with the present article.

```
SERVICE
<https://qlever.cs.uni-freiburg.de/api/osm-planet{
    ?osm_id osm2rdfkey:wikidata wd:Q22;
         a osm:relation;
         geo:hasGeometry?geometry.
       ?geometry geo:asWKT?wkt.
    }
  }
}
FILTER(geof:sfWithin(?wktLau,?wkt)).
}
```

This query extracts the VC titles, countries, and LAU geometries of the value chains operating in Scotland. The query structure is still similar to that of Q7. It uses the same external QLever Open Street Map endpoint to retrieve the geometry of Scotland boundaries. The FILTER clause operates the intersection between Scotland and the VCs' LAU geometries.

The query produced the results reported in Fig. 4. The expert's evaluation highlighted that the LAUs this query retrieved were correct and complete (Precision and Recall were 1). Therefore, the query was valuable in retrieving country-specific VCs.

Fig. 4 Highlight of the Local Administrative Units (orange) of the value chains operating in a polygon defining Scotland national boundaries (red). An interactive map is also reported in the Figshare repository associated with the present article.

Q10 - Value chains operating around long Italian rivers
In the following, we report the GeoSPARQL query corresponding to Q10.

GeoSPARQL Query 10

```
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
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 osm2rdfkey: <https://osm2rdf.cs.uni-freiburg.de/rdf/key#>

SELECT?nlabel?clabel?wktLau
WHERE {
    ?narra narra:isAboutCountry?country;
            narra:isAboutLAU?lau;
            rdfs:label?nlabel.
    ?country rdfs:label?clabel.
    ?lau geo:hasGeometry?glau.
    ?glau geo:asWKT?wktLau.
    {
```

```
SELECT?river_osm?river_wd?river_name?length?wktWHERE {
    SERVICE <https://qlever.cs.uni-freiburg.de/api/osm-planet>{
        ?river_osm a osm:relation;
                    osmkey:waterway?waterway;
                    geo:hasGeometry?geometry;
                    osmkey:name?river_name;
                    osm2rdfkey:wikidata?river_wd.
        ?geometry geo:asWKT?wkt.
    SERVICE <https://qlever.cs.uni-freiburg.de/api/wikidata>{
        ?river_wd wdt:P31/wdt:P279* wd:Q4022;
                    wdt:P17 wd:Q38;
                    wdt:P2043?length.
        FILTER (?length >100)
    }
    }
}ORDER BY DESC(?length)
}
FILTER(geof:sfIntersects(?wktLau,?wkt)).
}
```

This query retrieves all VCs operating close to an Italian river longer than 100 km. The query structure is very similar to that of Q7 and uses the same external endpoint. The main differences are the following:

- the nested *SELECT* clause operates on two different QLever-instance subgraphs: Open Street Map and Wikidata. The query retrieves the river geometries from the first. Then it uses the second to retrieve the list of "rivers" (id. Q4022) present in "Italy" (id. Q38) whose "length" (id. P2043) exceeds 100 km ("FILTER (?length > 100)");
- the final *FILTER* clause operates the intersection between the Italian rivers and the VCs' LAU geometries.

The query produced the results reported in Fig. 5. All VCs retrieved were correct and complete (Precision and Recall were 1). Therefore, the query was valuable in retrieving river-related VCs and, by extension, could be used to extract water-basin-related VCs.

Usage Notes

Our open-access Figshare repository[47] (version 3, currently) contains the entire OWL file and 454 OWL files corresponding to all VC tables. To perform non-explicit geospatial queries, a user can download and import the whole OWL graph (or a subset of the 454 OWL files to focus on specific regions or value chains) in an Apache Jena Fuseki triple store instance on a local machine[46]. After downloading and installing Fuseki[53], users should access the server interface through a Web browser and navigate to the "manage" interface section. Then, by clicking on the "new dataset" tab, they should create a new dataset, specifying a name and type (i.e., In-memory or Persistent). Next, they should select the just-created dataset and upload the entire OWL file (or a subset of the 454 OWL files) by clicking the "add data" button. Users can verify if the dataset was successfully populated by executing queries on the Fuseki SPARQL query interface. To perform geospatially explicit queries, a user should download and install Apache Jena GeoSPARQL Fuseki[54], a Fuseki version enhanced with GeoSPARQL features. Currently, this version does not have a Web interface to facilitate data import. Therefore, users should import data programmatically through the GeoSPARQL module embedded with this service[54]. The entire geospatialised narrative events' data (text, entities, geometries) in the knowledge graph are also available in CSV and GeoPackage formats, which can be imported, visualised, and manipulated through GIS software (e.g., QGIS[55]).

Our Figshare repository also contains a link to an overall visualisation facility for the VC narratives in the form of interactive and navigable Story Maps[56]. This facility allows our final users and stakeholders to easily explore the value chain locations, entities, events, and images. An overall map shows the distribution of the 454 VC narratives. After clicking on one reference pin, the user is redirected to the Story Map of the corresponding VC narrative. The Story Map lets the user go through the story events while panning, zooming, and inspecting the map locations at the right-hand side of the visualisation panel. The Story Maps are also available as an offline-visualisable HTML page collection in the Figshare repository.

Each Story Map allows users to access a "Search" Web interface (through a button at the top-left of the introductory page) that offers a visual search tool that executes semantic queries behind the scenes. This functionality interrogates the entire knowledge graph residing on a public-access Apache Jena GeoSPARQL Fuseki instance we currently offer and maintain[57]. The "Search" interface allows users to augment the knowledge reported in one Story Map through the knowledge in all other narratives. The "Search" functionality uses predefined SPARQL queries to extract:

1. All stories in which an entity appears;
2. All events across all narratives in which an entity appears;
3. The number of occurrences of one entity across all narratives;
4. The entities that co-occur with one entity across all events of all narratives.

Fig. 5 Highlight of the Local Administrative Units (orange) of the value chains intersecting Italian rivers longer than 100 km (red). An interactive map is also reported in the Figshare repository associated with the present article.

The purpose of this feature is to allow all users (also without skills in formal semantics) to explore narrative interconnections.

Code availability

Our public-access Figshare repository[47] contains all the JAVA and Python programs we used to execute the described workflow, along with the data inputs and outputs. It allows other scientists to conduct technical validation. The repository also contains an interactive, offline-visualisable HTML version of the result tables and maps.

Our code and data also embed information from the following external sources: Wikidata[58], OpenStreetMap[59] through the open-access QLever endpoint of the University of Freiburg[33], and geographic area definitions from Eurostat-GISCO[38–40].

References

1. UNPD (United Nations, D. o. E. & Social Affairs, P. D. World urbanization prospects: The 2018 revision (st/esa/ser.a/420). https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf (2018).
2. United Nations. Sustainable Development Goal 11, "Make cities and human settlements inclusive, safe, resilient and sustainable". on-line https://sdgs.un.org/goals/goal11 Accessed 4 January 2023 (2015).
3. H. Moving (mountain valorisation through interconnectedness and green growth https://www.moving-h2020.eu/ (2020).
4. Lehmann, F. Semantic networks. *Computers & Mathematics with Applications* **23**, 1–50 (1992).
5. Hogan, A. *et al*. Knowledge graphs. *ACM Computing Surveys (Csur)* **54**, 1–37 (2021).
6. Guarino, N., Oberle, D. & Staab, S. What is an ontology? *Handbook on ontologies* 1–17 (2009).
7. Meghini, C., Bartalesi, V. & Metilli, D. Representing narratives in digital libraries: The narrative ontology. *Semantic Web* **12**, 241–264 (2021).

8. CRAEFT European Project. Craft Understanding, Education, Training, and Preservation for Posterity and Prosperity (CRAEFT) Project Web site. on-line Accessed 12 July 2024 https://www.craeft.eu/ (2024).
9. Mingei European Project. Mingei - Representation and Preservation of Heritage Crafts Project Web site. on-line Accessed 12 July https://www.mingei-project.eu/ (2024).
10. IMAGO Italian PRIN Project. Index Medii Aevi Geographiae Operum (IMAGO) Project Web site. on-line https://imagoarchive.it Accessed 12 July (2024).
11. Doerr, M. The cidoc conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI magazine* **24**, 75–75 (2003).
12. Bekiari, C. *et al*. Definition of FRBRoo: A conceptual model for bibliographic information in object-oriented formalism. International Federation of Library Associations and Institutions (IFLA) repository https://repository.ifla.org/handle/123456789/659 (2017).
13. Pan, F. & Hobbs, J. R. Time ontology in owl. *W3C working draft, W3C* **1**, 1 (2006).
14. Battle, R. & Kolas, D. Geosparql, enabling a geospatial semantic web. *Semantic Web Journal* **3**, 355–370 (2011).
15. Thanos, C., Meghini, C., Bartalesi, V. & Coro, G. An exploratory approach to data driven knowledge creation. *Journal of Big Data* **10**, 1–15 (2023).
16. McInerny, G. J. *et al*. Information visualisation for science and policy: engaging users and avoiding bias. *Trends in ecology & evolution* **29**, 148–157 (2014).
17. Bruner, J. The narrative construction of reality. *Critical inquiry* **18**, 1–21 (1991).
18. Taylor, C. *Sources of the self: The making of the modern identity* (Harvard University Press, 1992).
19. Delafield-Butt, J. T. & Trevarthen, C. The ontogenesis of narrative: from moving to meaning. *Frontiers in psychology* **6**, 1157 (2015).
20. Bartalesi, V., Coro, G., Lenzi, E., Pagano, P. & Pratelli, N. From unstructured texts to semantic story maps. *International Journal of Digital Earth* **16**, 234–250 (2023).
21. Bartalesi, V. *et al*. Using semantic story maps to describe a territory beyond its map. *Semantic web (Online)* 1–18, https://doi.org/10.3233/SW-233485 (2023).
22. Caquard, S. & Cartwright, W. Narrative cartography: From mapping stories to the narrative of maps and mapping. *The Cartographic Journal* **51**, 101–106, https://doi.org/10.1179/0008704114Z.000000000130 (2014).
23. Peterle, G. Carto-fiction: narrativising maps through creative writing. *Social & Cultural Geography* **20**, 1070–1093, https://doi.org/10.1080/14649365.2018.1428820 (2019).
24. Bartalesi, V., Metilli, D., Pratelli, N. & Pontari, P. Towards a knowledge base of medieval and renaissance geographical latin works: The imago ontology. *Digital Scholarship in the Humanities* https://doi.org/10.1093/llc/fqab060 (2021).
25. Korzybski, A. A non-aristotelian system and its necessity for rigour in mathematics and physics. In *Science and sanity: an introduction to non-Aristotelian systems and general semantics* (Lancaster, 1933).
26. Figshare. How Figshare aligns with the FAIR principles. *Figshare Web site* = https://help.figshare.com/article/how-figshare-aligns-with-the-fair-principles (2024).
27. European Environment Agency. Terrestrial protected areas in Europe. https://www.eea.europa.eu/en/analysis/indicators/terrestrial-protected-areas-in-europe (2023).
28. Coro, G., Panichi, G., Pagano, P. & Perrone, E. Nlphub: An e-infrastructure-based text mining hub. *Concurrency and Computation: Practice and Experience* **33**, e5986 (2021).
29. Assante, M. *et al*. Enacting open science by d4science. *Future Generation Computer Systems* **101**, 555–563 (2019).
30. Coro, G., Candela, L., Pagano, P., Italiano, A. & Liccardo, L. Parallelizing the execution of native data mining algorithms for computational biology. *Concurrency and Computation: Practice and Experience* **27**, 4630–4644 (2015).
31. Coro, G., Panichi, G., Scarponi, P. & Pagano, P. Cloud computing in a distributed e-infrastructure using the web processing service standard. *Concurrency and Computation: Practice and Experience* **29**, e4219 (2017).
32. Wikidata. SPARQL entity retrieval specifications and examples. https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples (2024).
33. University of Freiburg. QLever instance. https://qlever.cs.uni-freiburg.de/ (2024).
34. Bast, H. & Buchhold, B. Qlever: A query engine for efficient sparql + text search. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 647–656 (2017).
35. Open Geospatial Consortium. Well-known text representation of coordinate reference systems. https://www.ogc.org/standard/wkt-crs/ (2024).
36. Eurostat. Local Administrative Units. https://ec.europa.eu/eurostat/web/nuts/local-administrative-units (2024).
37. Brandmueller, T., Schäfer, G., Ekkehard, P., Müller, O. & Angelova-Tosheva, V. Territorial indicators for policy purposes: Nuts regions and beyond. *Regional Statistics* **7**, 78–89 (2017).
38. Eurostat. GISCO - the Geographic Information System of the COmmission. https://ec.europa.eu/eurostat/web/gisco (2023).
39. Eurostat. LAU descriptions in GeoJSON format. https://ec.europa.eu/eurostat/web/gisco/geodata/statistical-units/local-administrative-units (2024).
40. Eurostat. NUTS descriptions in GeoJSON format. https://ec.europa.eu/eurostat/web/gisco/geodata/statistical-units/territorial-units-statistics (2024).
41. Commons, W. Wikimedia commons. *Retrieved June* **2** (2012).
42. Bartalesi, Valentina. JSON schema of the internal SMBVT data representation. https://dlnarratives.eu/schema%20JSON.json (2022).
43. W3C Consortium. RDF 1.1: On Semantics of RDF Datasets. https://www.w3.org/TR/rdf11-datasets/#bib-RDF11-MT (2024).
44. Ciccarese, P. & Peroni, S. The collections ontology: creating and handling collections in owl 2 dl frameworks. *Semantic Web* **5**, 515–529 (2014).
45. Meghini, C., Bartalesi, V., Metilli, D., Lenzi, E. & Pratelli, N. Narrative ontology. https://dlnarratives.eu/ontology/ (2024).
46. Jena, A. Apache jena fuseki. *The Apache Software Foundation* **18** (2014).
47. Bartalesi, V. *et al*. Figshare collection: A Knowledge Graph of European Mountain Territory and Value Chain data. *FigShare* https://doi.org/10.6084/m9.figshare.c.7098079 (2024).
48. Openllet. Openllet: An Open Source OWL DL reasoner for Java. https://github.com/Galigator/openllet (2023).
49. DuCharme, B. *Learning SPARQL: querying and updating with SPARQL 1.1* ("O'Reilly Media, Inc.", 2013).
50. Lam, A. N., Elvesæter, B. & Martin-Recuerda, F. A performance evaluation of owl 2 dl reasoners using ore 2015 and very large bio ontologies. In *In Proceedings of DMKG2023: 1st International Workshop on Data Management for Knowledge Graphs, May 28, 2023, Hersonissos, Greece* (2023).
51. United Nations. 68% of the world population projected to live in urban areas by 2050, says UN. on-line. https://www.un.org/en/desa/68-world-population-projected-live-urban-areas-2050-says-un (2018).
52. Dax, T. & Copus, A. European rural demographic strategies: Foreshadowing post-lisbon rural development policy? *World* **3**, 938–956 (2022).
53. Foundation, T. A. S. Apache jena fuseki. https://jena.apache.org/documentation/fuseki2/ (2024).
54. The Apache Software Foundation. GeoSPARQL Fuseki. https://jena.apache.org/documentation/geosparql/geosparql-fuseki.html (2024).
55. QGIS. Software download. https://qgis.org/download/ (2024).

56. Bartalesi, V. *et al*. Figshare collection: Visualisation of the MOVING 454 Story Maps. *FigShare* https://figshare.com/articles/online_resource/Moving_454_Storymaps/25334272?backTo=/collections/_/7098079 (2024).
57. Bartalesi, V., Lenzi, E. & Pratelli, N. Figshare collection: Fuseki instance for knowledge graph querying. *FigShare* https://tool.dlnarratives.eu/Moving_454_Storymaps/geosparql.html (2024).
58. Wikidata. Wikidata Web site. https://www.wikidata.org/wiki/Wikidata:Main_Page (2024).
59. OpenStreetMap. OpenStretMap data (Protocolbuffer Binary Format) in RDF representation enhanced by GeoSPARQL triples. https://planet.openstreetmap.org/pbf/planet-240701.osm.pbf (2024).

Acknowledgements

Author contributions

V.B. was one of the main developer of the Narrative Ontology, she orchestrated the experiment and conducted the validation; G.C. designed and developed the workflow for data augmentation and co-orchestrated the experiment; M.M. designed and developed the input MS Excel table collecting all information from the MOVING partners; N.P. designed and developed LAU/NUTS conversion through GISCO; E.L. designed and developed the Story Map visualisation and prepared the OWL graph instance on the Apache Jena Fuseki service; P.P. and G.B. supervised and supported the experiment through the MOVING project funding. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-024-03760-9.

Correspondence and requests for materials should be addressed to G.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.