Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

KGE 2024 - Trentino Tourist Facilities

Document Data:	Reference Persons:
February 1, 2025	Dongili Lorenzo, Grotto Giulia, Sacchet Stefano

Index:

Revision History:

Revision	Date	Author	Description of Changes
0.0	October 16^{th}, 2024	Grotto Giulia, Dongili Lorenzo, Sacchet Stefano	Document created
0.1	October 30^{th}, 2024	Grotto Giulia, Dongili Lorenzo, Sacchet Stefano	Phase 1 completed
0.2	November 13^{rd}, 2024	Grotto Giulia, Dongili Lorenzo, Sacchet Stefano	Phase 2 completed
0.3	November 25^{th}, 2024	Grotto Giulia, Dongili Lorenzo, Sacchet Stefano	Phase 3 completed
0.4	December 4^{th}, 2024	Grotto Giulia, Dongili Lorenzo, Sacchet Stefano	Phase 4 completed
0.5	February 1^{st}, 2025	Grotto Giulia, Dongili Lorenzo, Sacchet Stefano	Project completed

Introduction

Reusability is one of the main principles in the Knowledge Graph Engineering (KGE) process defined by iTelos. The KGE project documentation plays an important role to enhance the reusability of the resources handled and produced during the process. A clear description of the resources as well as of the process (and single activities) developed, provides a clear understanding of the project, thus serving such an information to external readers for the future exploitation of the project's outcomes.

The current document aims to provide a detailed report of the project developed following the iTelos methodology. The report is structured as follows:

- Section 2: Definition of the project's purpose and its domain of interest.
- Section 3: High level description of the project development, based on the Produce role's objectives.
- Sections 4, 5, 6, 7 and 8: The description of the iTelos process phases and their activities, divided by knowledge and data layer activities.
- Section 9: The description of the evaluation criteria and metrics applied to the project final outcome.
- Section 10: The description of the metadata produced for all (and all kind of) the resources handled and generated by the iTelos process, while executing the project.
- Section 11: Conclusions and open issues summary.

1 Purpose Definition

The iTelos methodology provides a structured approach aimed at reducing the effort in building Knowledge Graphs (KGs) for the purpose expressed by the final user. This section delves into the first phase of the methodology.

1.1 Informal Purpose

The goal of this project is to build a Knowledge Graph (KG) that stores, organizes, and provides easy access to information about tourist facilities in the Trentino region, with a specific focus on outdoor activities. The KG will serve as a centralized knowledge hub, connecting tourists, travel planners, and local businesses to the wealth of outdoor opportunities the region has to offer, such as hiking trails, walks, climbing spots, cycle routes and so on.

1.2 Domain Of Interest (DoI)

The Domain of Interest (DoI) for this project is the Trentino region in the year 2024, with a particular focus on outdoor activities. The Trentino region is renowned for its diverse landscapes, ranging from the Dolomite Mountains to lakes and lush valleys. The geographical scope of the project spans the entire region, encompassing its natural landscapes and alpine terrains, providing tourists with a rich selection of outdoor activities. Key features of domain include:

- *Mountain Adventure*: Activities include climbing, via ferrata, skiing, and trekking, offering experiences for all skill levels across Trentino alpine landscapes.
- *Sustainable Tourism and Ecotourism*: Eco-friendly options like wildlife observation, botanical walks, and guided tours focused on local flora and fauna designed to minimize environmental impact.

1.3 Scenarios Definition

In this section, a set of usage scenarios, describing the multiple aspects considered by the project purpose, are presented.

1. **Matteo Rossi** and his friends are planning a winter trip to Trentino. They want a ski resort with several black slopes and a snowpark for freestyle activities, as well as easy access to restaurants or cafes for relaxation after skiing. Since some friends aren't skiers, they're also looking for nearby winter activities, like hiking or sledding. Additionally, they hope to try traditional Trentino cuisine, ideally at or near the resort.

2. **Vincent Baguette** and his family are planning a summer trip in Trentino. They're looking for a scenic bike trail suitable for families, ideally near a lake where they can cool off with a swim afterward. Since they have young children, they need a safe route with rest areas along the way. They're also hoping to find an hotel where to stay for some nights in order to visit more than one landscape.
3. **Asja Bak** and her friends plan a trekking adventure in Trentino. They're looking for a scenic trail that leads to a mountain refuge where they can enjoy a hearty local meal. Since they prefer moderate difficulty, they want a route that's challenging but accessible, with view-points along the way for photo stops. A cozy refuge with traditional Trentino dishes at the end of the trek is a must for their experience.
4. **Karl Schneider**, an FBK employee, is planning a rock climbing trip in Trentino with his friends. They're looking for a climbing crag with routes suited to various skill levels so everyone in the group can participate. Ideally, they'd like a location that offers clear difficulty ratings and safe conditions, with nearby rest spots to take breaks and enjoy the scenery.
5. **Alex Megos** and his fiancee are planning a summer road trip to Trentino in their van. They seek challenging climbing routes from grade 7 and above, that offer thrilling experiences. Marco is focused on finding crags with medium-high difficulty levels. They are also interested in exploring local cuisine by visiting nearby restaurants. The family plans to use their knowledge graph to locate ideal climbing, and cultural experiences during their trip.

1.4 Personas

In this section a set of real users acting within the scenarios defined above are defined.

1. **Matteo Rossi** is a 24 years old guy studying Computer Science in Trento. He loves snow-boarding especially on long slopes and practicing at the snow park.
2. **Vincent Baguette** is a 47 years old with a family with kids respectively of 7 and 9 years old. who loves nature mountains and bike rides. He is always looking for beautiful trails and landscapes.
3. **Asja Bak** is a 21 years old Erasmus student from Poland studying in Trento. She loves the adrenaline and seeks out unique experiences.
4. **Karl Schneider** is a 27 years old guy from Austria, currently working in FBK. He would like to start climbing.
5. **Alex Megos** is a 30 years old man. He is a pro climber who loves nature and traveling with

his fiance around the world in search of beautiful climbing routes. He likes climbing and constantly challenging himself by attempting increasingly difficult routes.

1.5 Competency Questions (CQs)

In Table 1, a list of Competency Questions (CQs) is created, taking into account the personas defined in the scenarios.

Person	No.	Question
Matteo Rossi	1.1	Which skiing infrastructures are available in the Trentino Region?
Matteo Rossi	1.2	Where can they rent ski equipment?
Matteo Rossi	1.3	Which restaurant can they choose to go?
Matteo Rossi	1.5	Where can they find a hotel near the skiing resort?
Matteo Rossi	1.6	Which ski area has the longest extension?
Matteo Rossi	1.7	Which natural attraction is the closest to the skiing site that he choose?
Matteo Rossi	1.8	Which ski resort has black slopes?
Vincent Baguette	2.1	Are there family-friendly bike trails in Trentino?
Vincent Baguette	2.3	Where can families find scenic bike trails with restaurants nearby?
Vincent Baguette	2.5	Which bicycle stop is near to the restaurant they chose to eat?
Vincent Baguette	2.6	Where is an hotel in which the family can stay for some nights?
Asja Bak	3.1	What scenic trekking trails in Trentino lead to a mountain refuge?
Asja Bak	3.2	Which trekking routes in Trentino offer moderate difficulty and are suitable for hikers seeking a challenge?
Asja Bak	3.3	Which trekking trails end at a refuge offering traditional Trentino cuisine?
Karl Schneider	4.1	What rock climbing crags in Trentino offer routes suitable for various skill levels?
Karl Schneider	4.2	Where can climbers find routes with different difficulty levels within the same location?
Karl Schneider	4.3	Where can they rent climbing equipment?
Karl Schneider	4.4	Where can they eat or drink something?
Alex Megos	5.1	Which climbing crags in Trentino offer grades from 7a and above?
Alex Megos	5.2	Are there challenging climbing locations in Trentino with nearby designated camping spots for vans or tents?
Alex Megos	5.3	Which climbing areas have scenic viewpoints or cultural landmarks nearby for added exploration?
Alex Megos	5.4	What markets or restaurants near climbing locations in Trentino offer authentic local dishes?

Table 1: Competency Questions

1.6 Concepts identification

In this section, the key concepts representing entities and their associated properties within the Knowledge Graph are outlined. In Fig. 1 there is the Purpose Formalization sheet which details the formal objectives and scope of the knowledge graph.

Scenarios	Personas	Competency Questions	Entities	Properties	Focus
1	Matteo Rossi	1.1, 1.8	Ski Slope	SlopeID, OSMid, Resort, Difficulty	Contextual
1, 3	Matteo Rossi, Asja Bak	3.1, 3.2, 3.3	Hiking trail	HikingID, OSMid, Length, Elevation gain	Contextual
7	Alex Megos	7.1	Climbing route	ClimbRouteID, Crag, Country, Sector, tall_recommended_sum, grade_mean, cluster, rating_total	Contextual
2	Vincent Baguette	2.1, 2.2, 2.3	Cycle routes	CycleRouteID, OSMid, Length, Elevation gain, Surface, Difficulty	Contextual
*	*	*	Activity	ActivityID, Type, Name, Path	Common
*	*	*	Location	LocationID, OSMid, LocationName, Position	Common
1, 5	Matteo Rossi, Alex Megos	1.7, 5.3	View point	ViewPointID, Elevation	Core
2	Vincent Baguette	2.5	Hotel	HotelID, star_rating, e-mail, website, Altitude, ServiceType, Services, Price, nBeds	Core
1, 2, 4, 7	Matteo Rossi, Vincent Baguette, Karl Schneider, Alex Megos	2.5, 2.3, 4.4, 7.2, 1.1, 2.4, 4.3	Building	BuildingID, Type, Phone, Municipality, PostalCode, Address, HouseNumber	Common
*	*	*	Coordinate	Latitude, Longitude	Common
1, 2, 3, 4, 5	Matteo Rossi, Vincent Baguette, Asja Bak, Karl Schneider, Alex Megos	1.3, 2.5, 3.3, 4.4, 5.4	Reastaurant	RestaurantID, OSMid	Core
1, 2	Matteo Rossi, Vincent Baguette, Karl Schneider	1.2, 2.4, 4.3	Ski Rental	SkiRentID, OSMid	Core

Figure 1: Purpose Formalization

1.7 ER model definition

The KG purpose is defined as providing to tourists easy access to information about outdoor activities, such as the trails where to go hiking and if there is some rest area or rent shop nearby. So, the ER is thought reflecting a need for the user to find this information easily, centering the ER on two entities: *Activity* and *Location*, seen respectively as the input and the output of the user request. These two entities are colored as common entities cause they store main information for all the other entities. We set also another common entity: *Coordinate*, which will be an object property in the final KG and contain longitude and latitude.

Connected to *Activity*, there is a set of entities, defined as contextual, containing specific properties about climbing routes, cycleways, ski slopes and hiking trails.

Instead, connected directly to Location, there are some entities which are satisfying some second needs for the tourists, such as *View Point* defined as core, and the common entity *Building*, which contain information about restaurants, rent shops and hotels.

1.8 Report

In this chapter, we set up a solid base by defining our main purpose, creating realistic scenarios, and developing specific competency questions. This helps make sure the Knowledge Graph meets the needs of Trentino's tourism, with a focus on tourists and activity planners. The user personas add helpful context, guiding us to design the graph around practical needs, like finding

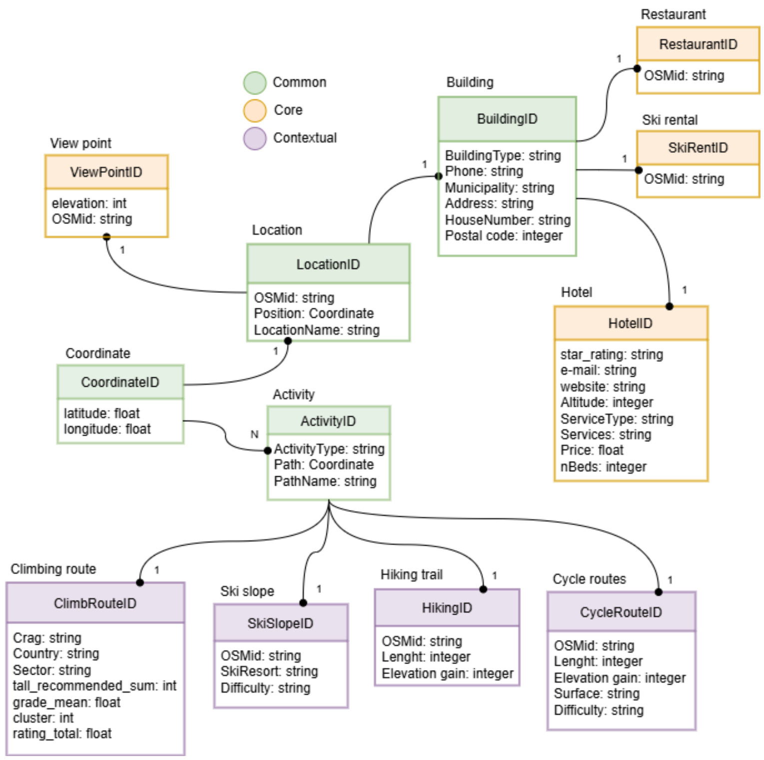

Figure 2: ER model

hiking trails, nearby dining options, and different outdoor activities.

A key strength is the user-centered design that emerges from these scenarios, which justify the inclusion of entities like activities and locations. This focus highlights the rich diversity of Trentino's tourism offerings, making the Knowledge Graph more relevant to its intended audience. Initially, we considered including data on transportation services, but we decided that this would shift the focus away from our main theme of outdoor activities and local experiences.

A potential challenge lies in our initial Entity-Relationship (ER) model, which is structured to capture key attributes but may face data availability issues. Since some attributes are yet to be confirmed in available datasets, this may affect our ability to answer all competency questions as thoroughly as planned. To address this, we plan to revisit and refine the ER model after the data-gathering phase, making adjustments based on the data we find. This iterative approach will help us create a more practical and user-centered Knowledge Graph.

2 Information Gathering

In this section, an overview of the primary input data sources available for the project is provided. This includes a list of resources that span across languages, schemas, and data values.

2.1 Knowledge Sources

In this phase, our goal is to gather high-quality, relevant information to support the construction of our Knowledge Graph, ensuring alignment with the purpose and scope defined earlier. This involves a meticulous selection of data sources and content that can accurately populate our KG with meaningful, structured information.

To address the project's knowledge resource needs, we used mainly OpenStreetMap, which provide a comprehensive set of predefined schemas for structured data on the internet. Schema.org was used only for one entity type, and together with OSM are collaborative, community-driven initiatives aimed at creating, maintaining, and promoting flexible schemas that can be applied across diverse domains. By aligning our vocabulary domain with these schemas, we aimed to enhance flexibility and improve the dataset's readability and reusability for future users.

All the query performed on OpenStreetMap data, were run via overpass-turbo.eu. Overpass Turbo is a web-based tool designed to help users interactively query and visualize data from OSM, the open-source global mapping platform. The primary purpose of Overpass Turbo is to allow users to create custom, location-based queries to extract specific types of geospatial data, such as locations of specific types of buildings, landmarks, roads, natural features, and more.

Due to the big amount of data gather from OSM, we inserted the OSMid in all the entity type containing data gained from it, to enhance the reusability.

These datasets adheres to the *OSM* schema, which provides a structured and standardized format for organizing restaurant data. This schema helps to ensure consistency in data presentation and compatibility, facilitating integration with other resources and applications.

2.2 Climbing Routes Resource

The **Climbing Resource** is a **data value dataset** which provides data on climbing routes, specifically focusing on various climbing crags within the Trentino region. This resource includes detailed metrics for each route, offering insights into climbing difficulty, popularity, and community preferences. The original dataset can be found on Kaggle, and while it already contained valuable information, it did not initially focus solely on the Trentino region nor did it include route coordinates. To refine the dataset, these limitations were addressed by filtering for Trentino routes and adding geographic coordinates using a Python script and the geopy library. It is important to note that no predefined schemas are available for this resource. Consequently, data organization and management rely on the structure and attribute definitions set within the dataset itself, without additional formal schema references.

Key attributes in the dataset are:

- **name_id** (unsigned int)
- **Type** (string): it contains ”*climbing*”, the type of activity
- **country** (string)
- **crag** (string): the specific climbing area or rock formation where the route is found, helping to identify the general location within a broader region.
- **sector** (string): a sub-area within the crag where the route is located, offering a more precise location.
- **name** (string): the official or commonly used name of the climbing route.
- **height_difficulty_score** (string): this metric reflects accessibility based on climber height, further details can be found at 2.
- **grade_mean** (float): this represents the median grade for each route, calculated after standardizing grades.
- **route_category** (string): further details can be found at 3.
- **rating_total** (float): calculated based on comment sentiment, rating, and user recommendations/
- **coordinates** ([float, float])

2.3 Hotel Resource

The **Hotel Resource** is a **data value dataset** that provides information on hotels, specifically focusing on those within the Trentino region. The original dataset can be found at DatiOpen.it. However, it initially contained additional information that was not relevant for this specific purpose. To refine the dataset, these limitations were addressed through a data-cleaning process.

Value	Description
recommend_high	Recommended for tall climbers
recommend_low	Recommended for short climbers

Table 2: Climbing resource, *height difficulty score* attribute.

Value	Description
soft_route	Soft routes, easier to climb
preferred_by_women	Routes preferred by women, indicating gender-influenced popularity
famous_route	Famous routes, known and recommended in the community
hard_route	Very hard routes, challenging even for experienced climbers
repeated_route	Highly repeated routes, popular and frequently climbed
chipped_route	Chipped routes with a softer grading
non_chipped_route	Traditional, unmodified routes that remain close to natural conditions
easy_on_sight	Routes that are easy to on-sight but not frequently repeated
less_repeated	Iconic routes, renowned yet less repeated and often less traditional

Table 3: Climbing resource, *route category* attribute.

The dataset was filtered to include only valuable information, and geographic coordinates were added using a Python script and the geopy library. Additionally, the dataset follows the schema outlined by Schema.org Hotel schema, which provides a structured and standardized way to organize hotel data. This schema enhances data consistency and ensures compatibility with commonly used data frameworks.

Key attributes in the dataset are:

- **Municipality** (string)
- **Star Rating** (string): from 1 to 5 starts.
- **Facility Name** (string)
- **Address** (string)
- **Postal Code** (unsigned int)
- **Phone** (string)
- **Email Address** (string)
- **Web Site** (string)

- **Service Type** (string): can be *full-board*, *bed-and-breakfast* or *room-only*.
- **Altitude** (unsigned int)
- **Number of Beds** (unsigned int)
- **Price** (float)
- **Services** (string): describes extra services the hotel offers.
- **Coordinates** (Coordinate)

2.4 Ski Slopes Resources

The **Ski Slopes Resource** is a **data value dataset** that provides information about Ski slopes and resorts, specifically focusing on those within the Trentino region. The final dataset is a mixture of the dataset that can be found at datiopen.it, which gave us information about the ski resort in Trentino where the slopes are located, and the result of a query performed on OpenStreetMap data that gives us insights about the coordinates of each slope, their name, and its difficulty.

To finalize the dataset with the information needed for our purpose, we used a script implementing Nominatim API (nominatim.org), which takes the name of the ski resort, gave us the coordinates of it. Then, with another script, we used the coordinates of the resorts and the first coordinates of each slope to couple them together when the distance between them is less than 15km. These two scripts were made in Python.

The query run on Overpass-Turbo is:

```
[out:json];
area[name="Trentino-Alto Adige/Südtirol"];
(
  way["piste:type"="downhill"](area);
);
(._;>;);
out body;
```

Key attributes in the dataset are:

- **id** (string)
- **osmId** (string): original id from OpenStreetMap to increase reusability.

- **Name** (string)
- **Difficulty** (string): can be *easy*, *intermediate*, and *advanced*
- **Ski Resort**
- **Coordinates**: the list containing the coordinates all along the slope

2.5 Hiking Trails Resources

The **Hiking Trails Resource** is a **data value dataset** that provides information about hiking trails, specifically focusing on those within the Trentino region.

The final dataset is composed from the output of the query performed on OSM, and the make up of some properties with a Python script. In this specific dataset, the total distance and the total elevation gain are calculated through the list of coordinates and the use of the math library in Python.

The query run on Overpass-Turbo is:

```
[out:json];
area[name="Trentino-Alto Adige/Südtirol"];
(
    way["highway"="path"]["foot"="yes"]
    (area);
    way["highway"="footway"]
    (area);
    way["highway"="hiking"]
    (area);
);
(._;>;);
out body;
```

Key attributes in the dataset are:

- **id** (string)
- **osmId** (string): original id from OpenStreetMap to increase reusability.
- **Name** (string)
- **Type** (string): it contains ”*hiking*”, the type of activity

- **Coordinates** (Coordinate): the list containing the coordinates all along the trail
- **total_distance_m** (float)
- **total_elevation_gain** (float)

2.6 Rent Shops Resources

The **Rent shops Resource** is a **data value dataset** that provides information about rent shops, specifically focusing on those within the Trentino region.

With this dataset we faced a problem of non-availability of data. In fact, from OSM there aren't bike rental shops, and only 6 ski rental shops. Moreover, these 6 shops doesn't have any information except from the name and the coordinates. So, we use Google Maps to retrieve them.

The query run on Overpass-Turbo is:

```
[out:json];
area[name="Trentino-Alto Adige/Südtirol"];
(
    node["shop"="ski_rental"]
    (area);
    way["shop"="ski_rental"]
    (area);
    relation["shop"="ski_rental"]
    (area);
);
out body;
```

Key attributes in the dataset are:

- **id** (string)
- **osmId** (string): original id from OpenStreetMap to increase reusability.
- **Name** (string)
- **Type** (string): it contains "*skiing*", the type of activity
- **addrCity** (string)
- **addrPostcode** (string)

- **addrStreet** (string)
- **phone** (string)
- **Coordinates** (Coordinate)

2.7 Restaurants Resources

The **Restaurant Resource** is a **data value dataset** that provides detailed information on restaurants, specifically focusing on establishments within the Trentino region.

The final dataset is the output of a filtering on the rows of the dataset that was gained from OSM. The filtering is made with a Python script with the aim of deleting those restaurants which doesn't have enough information to fit in our purpose.

The query run on Overpass-Turbo is:

```
[out:json];
area[name="Trentino-Alto Adige/Südtirol"];
(
    node["amenity"="restaurant"]
    (area);
    way["amenity"="restaurant"]
    (area);
    relation["amenity"="restaurant"]
    (area);
);
out body;
```

Key attributes in the dataset are:

- **id** (string)
- **osmId** (string): original id from OpenStreetMap to increase reusability.
- **addrCity** (string)
- **addrPostCode** (string)
- **addrStreet** (string)
- **addrHousenumber** (string)
- **phone** (string)

- **coordinates** (Coordinate)

2.8 Cycle routes Resources

The **Cycle Routes Resource** is a **data value dataset** that provides information about cycleways, specifically focusing on those within the Trentino region.

The final dataset is composed from the output of the query performed on OSM, and the make up of some properties with a Python script. In this specific dataset, the total distance and the total elevation gain are calculated through the list of coordinates and the use of the math library in Python.

The query run on Overpass-Turbo is:

```
[out:json];
area[name="Trentino-Alto Adige/Südtirol"];
(
    way["highway"="cycleway"]
    (area);
    way["highway"="path"]["bicycle"="yes"]
    (area);
    relation["type"="route"]["route"="bicycle"]
    (area);
);
(._;>;);
out body;
```

Key attributes in the dataset are:

- **id** (string)
- **osmId** (string): original id from OpenStreetMap to increase reusability.
- **Type** (string): it contains ”*biking*”, the type of activity
- **Name** (string)
- **Coordinates**: the list containing the coordinates all along the trail
- **Difficulty** (string): can be *easy*, *intermediate*, and *advanced*
- **Surface** (string): can be *asphalt*, *unpaved*, *wood*, *gravel*, *metal*, *concrete*

- **total_distance_m** (float)
- **total_elevation_gain** (float)

2.9 View Points Resources

The **View Point resources** is a **data value dataset** that provides information about view points, focusing on those within the Trentino region.

The final dataset is composed from the output of the query performed on OSM in addition with a property, *elevation*, gained with the use of a Python script and the open-elevation.com API, which calculates the elevation of a point based on the coordinates of it.

The query run on Overpass-Turbo is:

```
[out:json];
area[name="Trentino-Alto Adige/Südtirol"];
(
    node["tourism"="viewpoint"]
    (area);
);
out body;
```

Key attributes in the dataset are:

- **id** (string)
- **osmId** (string): original id from OpenStreetMap to increase reusability.
- **Name** (string)
- **Elevation** (integer)
- **coordinates** (Coordinate)

2.10 Location Resources

The **Location resources** is a **data value dataset** that provides information about all the locations present in our KG. This dataset is make up with properties from other entities, such as *View Point*, *Building*, and *Hotel*.

Key attributes in the dataset are:

- **id** (string)
- **osmId** (string): original id from OpenStreetMap to increase reusability.
- **Position** (Coordinate)
- **Name** (string) : name present in the dataset of each specific entity

2.11 Activity Resources

The **Activity resources** is a **data value dataset** that provides information about all the activities present in our KG. This dataset is make up with properties from other entities, such as *Hiking trail*, *Climbing route*, *Ski slope* and *Cycle path*.

Key attributes in the dataset are:

- **id** (string)
- **Type** (string): type of activity, taken from the dataset of each activity
- **Path** (Coordinate) : list of coordinates highlighting the paths and trails
- **Name** (string)

2.12 Building resources

The **Building resources** is a **data value dataset** that provides information about all the buildings-type locations present in our KG. This dataset is make up with properties from other entities, such as *Restaurant*, *Hotel*, and *Rent shop*.

Key attributes in the dataset are:

- **id** (string)
- **Type** (string): type of activity which will be imported from each building dataset
- **phone** (Coordinate) : coordinate of the single point where the building is located
- **Municipality** (string)
- **Address** (string)
- **Postal Code** (string)

2.13 Report

In this phase of the project, we faced several challenges with data availability, as anticipated in 1.8. Some of the data we included in our original ER model was unavailable, leading us to remove certain entities, specifically the *Boathouse* and *Bathing Establishment*.

Additionally, our initial ER model included a high level of detail, that we could not fully support due to data limitations. For example, rather than maintaining separate entities for *Climbing Route* and *Crag*, we consolidated them by keeping only the *Climbing Route* entity, which now has an attribute specifying the crag name. We applied a similar simplification to *Ski Resort* and *Slope* by merging these entities into a single entity..

These issues led us to narrow the scope of our Domain of Interest by excluding lakeside tourism and water-based activities. As a result, we could no longer address some of the initial scenarios we had planned to support.

Another issue we faced was the redundancy of certain attributes across multiple entities. For instance, the Name attribute was originally present in every entity. To reduce redundancy and enhance the structure, we used the IS-A relationship to inherit shared attributes among related entities. For instance, we defined Name within the *Activity* entity and removed it from *Climbing Route*, *Hiking Trail*, *Cycle Routes*, and *Ski Slope*. The same procedure has been applied to *Rent Shop*, *Restaurant*, and *Hotel*, which were composed of the same properties: Name, Phone, Municipality, Address and Postal Code that have been moved to a new entity *Building*.

This adjustment enabled a more streamlined and hierarchical organization of attributes, better reflecting the relationships and focus of each entity.

In the *Information Gathering* phase, we encountered significant challenges with data quality. The data we obtained was often incomplete, inconsistent, or formatted in ways that required extensive cleaning and formatting to make it suitable for our purpose.

Additionally, much of the data lacked specific attributes that were essential according to our initial ER model. To address this, we implemented Python scripts to fill in missing information where possible. These scripts allowed us to automate parts of the data completion process, helping to bridge the gaps in our dataset and align it more closely with our project's initial specifications.

For the Climbing Route Resources, we were not able to find a predefined schema either in schema.org or OSM, so we will define one in the next phases.

3 Language Definition

This section describes the language definition phase. This phase is important cause it ensures clarity and consistency by addressing ambiguity and diversity in natural and domain-specific languages. It enables accurate data annotation, minimizes misunderstandings, and supports seamless integration across systems. This fundamental step is crucial for effective communication and interoperability in complex, multilingual, or multi-domain projects.

The outcomes produced during this phase can be found in the Github repository as *LanguageResources* sheet.

3.1 Concept Identification

The team began by identifying relevant concepts, assigning each attribute an unique identifier (ConceptID) from UKC. This ensures re-usability and reduces ambiguity. Concepts were accompanied by a gloss (Gloss-en) to provide clarity. For example:

- ConceptID: UKC-32153, *Word: ”coordinate,” Gloss: ”a number that identifies a position relative to a reference point.”*
- ConceptID: UKC-50, *Word: ”location,” Gloss: ”a point or extent in space.”*

During the Language Definition activity, we aimed to align all concepts with those available on the Universal Knowledge Core (UKC). However, not all concepts were readily available or defined as we intended them. Consequently, we relied on formal descriptions from (Open Street Map) as an alternative source. For example, the word *ski rental* is not available in UKC; therefore, we formalized our concept using the definition found in OSM wiki for ski rental.

In cases where formal descriptions were unavailable for certain attributes, we created custom definitions. This approach ensured that all necessary concepts were represented, maintaining the comprehensiveness required. For instance, *hiking trail* is not available in UKC. While the closest term, *trail*, exists, its definition was too vague for our purpose. As a result, we decided to create a custom concept tailored to our specific needs.

During this phase, we did not add any object properties. The only relationships present in our conceptual framework are is-a and has-a, which are standard and already well defined. These relationships did not require any additional customization or formalization for our purposes.

3.2 Dataset Filtering

The Dataset Filtering activity is designed to align the data layer resources, previously collected and formalized, with the concepts identified and structured during the parallel Concept Identification activity. In our case, this filtering process was already conducted during the earlier phases of the methodology. As a result, the dataset provided at this stage was already pre-aligned with the concepts identified in the Concept Identification activity. Consequently, no additional filtering was required during this phase.

4 Knowledge Definition

The Knowledge Definition phase is a critical step within the iTelos methodology, aimed at formalizing and aligning the gained information to construct a Knowledge Graph. This phase consists in creating a unified representation of knowledge by defining teleontologies and aligning datasets. It consists of sub-activities including kTelos (teleology and teleontology definitions) and dataset cleaning and formatting. This report outlines the activities, decisions, and outcomes achieved in this phase.

We used Protégé, a widely recognized ontology editor, which facilitated the creation and management of the ontology by providing a comprehensive interface to define entity types, object properties, and data properties. Additionally, we used the given OpenStreetMap (OSM) ontology as a reference (OSM schema), putting an additional annotation in those eTypes we created which has been reused, to specify the synonym of the name present on OSM schema. This increased the reusability of existing knowledge structures, reducing redundancy and aligning with established standards. The result of this phase can be found inside our Github repository.

However, during the alignment of the OSM ontology with our schema, we identified a key limitation: the absence of the property *way* in the OSM ontology. This property, as defined in OpenStreetMap, allows the representation of a list of coordinates corresponding to a path, such as a cycleway. This property was critical for our project as it would enable precise modeling of entities involving paths (*ski trail*, *cycle route*, *hiking trail*). Furthermore, we discovered that not all the entity types implemented for our purpose had corresponding definitions in the given OSM ontology. For example, specific eTypes such as *ski rental*, *ski trail* and *climbing route* were absent. To address this gap, we documented the missing property and outlined its significance to ensure future updates to the ontology could accommodate such requirements.

The final solution for our schema will be rooted in the entity ***Location*** which is extended by all the other entities. In particular:

- ***Building***, containing *Hotel*, *Restaurant*, and *Ski Rental*
- ***Sport Location***, containing *Ski Trail*, *Cycleway*, *Climbing Route*, and *Hiking Trail*
- ***View***.

This solution allows us to maximize the reusability taking into consideration the structure of the data we collected in the Information Gathering phase (2).

4.1 Consequences for Competency Questions

The modifications and limitations identified in the Knowledge Definition phase affected our ability to answer some of the competency questions (CQs) that can be found at 1. Specifically:

- CQ 4.3: We cannot fully answer this question because no data on climbing rental shops was found during the Information Gathering phase.
- CQ 5.2: This question cannot be answered because no data on camping spots was available or included in our schema.
- CQ 3.3: We can partially answer this question as we have data on restaurants but lack specific information on refuges.

These gaps highlight the importance of refining the data acquisition process in future iterations to ensure a more complete coverage of all competency questions.

4.2 Corrections

Following the corrections identified in the Language Definition phase (3), we revised the naming of attributes to enhance clarity and precision. For certain attributes, we adopt more expressive names that better explained their purpose and role, ensuring they were intuitive and easy to understand. For others, we selected more specific terms that accurately reflected the detailed nature of the concept they represented.

5 Entity Definition

This section is dedicated to the description of the Entity Definition phase. Like in the previous section, it aims to describe the different sub-activities performed by all the team members, as well as the phase outcomes produced.

5.1 Overview and Objectives

The Entity Definition phase is the final step in the iTelos methodology, merging the knowledge and data layers into a unified Knowledge Graph. The input for this phase consists of cleaned and aligned data resources, as well as the teleontology created in previous stages. The objective is to address the heterogeneity of data values, ensuring that entities across datasets are uniquely identified, matched, and mapped. The output of this phase includes the finalized KG and a set of mapping models in RDF-Turtle format.

The Entity Definition phase is structured into the following sub-activities:

- **Entity Matching:** resolves discrepancies in the representation of real-world entities across different datasets, considering both schema and data-level heterogeneity.
- **Entity Identification:** ensures each entity is uniquely identified, using either existing identifiers or constructed identifying sets.
- **Entity Mapping:** combines the teleontology with corresponding data values in the datasets, generating the final KG.

5.2 Entity Matching

Entity matching addressed the challenge of reconciling multiple representations of the same real-world entity across datasets. In our case we used only 1 dataset for each eType, because they were sufficiently large to develop a KG respecting our purpose. The only exception is the ski slope dataset which is a mix-up of 2 different datasets carrying different informations for entities that in many cases corresponded, that are the ones we relied on.

In the specific case of restaurants, even if the data derived only from one dataset, there were many features with a high percentage of missing values. Therefore, we selected only those that were useful for our purpose, ensuring a compromise between the quantity and quality. This choice also led to some trade-offs: for example, we had to exclude the feature related to traditional Trentino cuisine, which was part of our competency questions.

5.3 Entity Identification

Once matching conflicts were resolved, the focus shifted to entity identification. This step is crucial to ensure that each entity in the dataset is uniquely and consistently represented within the Knowledge Graph. We create custom identifiers for each dataset by combining the resource name with its corresponding entity, for example *restaurant17*. Moreover, whenever the dataset is extracted from Open Street Map, we also maintain the OSM ID provided for each entity, as it serves as a reliable and standardized unique identifier within the OSM environment. Also for the climbing routes dataset, we maintain the id given by the author of the Kaggle dataset. This method ensures that every entity can be uniquely identified, enhancing reusability for future uses, and mitigating the challenges posed by incomplete or inconsistent data.

5.4 Entity Mapping

Entity mapping integrates the teleontology with the datasets by defining the relationships between entity types and their data values. This activity is implemented using the Karma tool, which facilitates the mapping process and ensures the correct representation of entities within the Knowledge Graph. The output includes an RDF-Turtle format for each eType, defining all mapping operations performed, and containing the data values from the *Information Gathering* phase mapped against the teleology output of the *Knowledge Definition* phase.

5.5 Phase Outcomes

The Entity Definition phase ends with the creation of a comprehensive and unified Knowledge Graph. This final output integrates knowledge and data layers, addressing schema, structure, and value-level heterogeneity to provide a reliable resource for answering competency questions and achieving the project's goals.

5.6 Decisions and Reflections

During this phase, the team made some decisions to ensure a robust and reusable KG.

The mapping between the eTypes and their data values has been made only in the leaf nodes of our graph, that are also the subclasses in the Knowledge Level (*View*, *Hotel*, *Restaurant*, *Ski Rental*, *Rock Climbing Route*, *Cycleway*, *Hiking Trail*, *Ski Slope*), because we don't have any request to satisfy within the CQs for which the superclasses can be useful. For example, one of our superclasses is *Building*, and we do not have CQs such as "Which are the building in the Trentino region?". Anyway, the data properties of these superclasses are important for our purpose and that's why they are inherited by the subclasses and mapped with specific data values.

- **Strengths:** The superclasses are integrated in our Teleology ensuring reusability for whom it may concern. This allows the integration of such data about superclasses in the future. Moreover, maintaining the OSMid where possible, ensures another level of reusability for the future.

- **Weaknesses:** Some datasets have many entities with general information, but not so much with specific ones, which could help us improve our final KG.

6 Evaluation

This section presents the evaluation conducted at the conclusion of the iTelos methodology, focusing on evaluating the final Knowledge Graph (KG) generated through the process. The evaluation aims to provide an analysis of the KG structure, quality, and effectiveness in meeting the project purpose. The evaluation is structured into several components to ensure a complete examination of the final outcome:

- **Knowledge Graph Statistics** 6.1: a detailed report on the structural characteristics of the KG, including the number of entity types, properties, and entities associated with each etype, as well as other relevant metrics.
- **Knowledge Layer Evaluation** 6.2: an analysis of the knowledge layer of the KG using defined evaluation metrics to assess its completeness, and alignment with the teleontology.
- **Data Layer Evaluation** 6.3: a review of the data layer, ensuring that the underlying data supports the teleontology effectively.
- **Query Execution** 6.4: an assessment of KG ability to fulfill the project purpose through the execution of competency queries. This includes testing the KG ability to provide accurate and relevant responses to the queries defined earlier in the process.

This evaluation, aims to validate the suitability, reusability, and overall quality of the final Knowledge Graph as the main deliverable of the iTelos methodology.

6.1 Knowledge Graph Statistics

This paragraph provides an overview of the structural characteristics of the final Knowledge Graph, detailing the number of entity types, properties, and entity distributions. The final Knowledge Graph consists of **11 etypes**, structured as follows: *Location, Building, Hotel, Restaurant, Ski Rental, Sport Location, Cycleway, Hiking Trail, Rock Climbing Route, Ski Trail, and View*.

In terms of properties, the Knowledge Graph includes:

- **Object Properties: 0**. During this *Knowledge Definition* phase, we did not introduce any object properties. The only relationships present in our conceptual framework are *is-a*, which are standard and already well defined. These relationships did not require additional customization or formalization for our purposes, and are exhaustive for our prospective of the KG.

- **Data properties: 32**. These properties capture various attributes of eTypes, ensuring the proper representation of relevant information within the Knowledge Graph.

6.2 Knowledge Layer Evaluation

The Knowledge Layer evaluation assesses how well the Teleontology meets two key objectives. The primary objective (Teleontology vs Competency Questions 6.2.1) evaluates its coverage of entities and properties extracted from Competency Questions (CQs). The secondary objective (Teleontology vs Reference Ontologies 6.2.2) measures how much the Teleontology aligns with the entity types and properties from reference ontologies.

6.2.1 Teleontology vs Competency Questions

Given a set of (CQ), the etype coverage (Cov_E) of Teleontology (T) is:

$$Cov_e(CQ_E) = \frac{|CQ_E \cap T_E|}{CQ_E} = \frac{8}{14} \approx 0.571 \quad (1)$$

Where:

- CQ_E is the number of etypes extracted from the CQs.
- T_E is the number of etypes of the Teleontology.
- We are missing etypes: *Cultural Landmark, Market, Camping Spot, Climbing Rental, Mountain Hat and Bicycle Stop*.

Given a set of (CQ), the property coverage (Cov_p) of the Teleontology (T) is:

$$COV_p(CQ_P) = \frac{|CQ_P \cap T_P|}{CQ_P} = \frac{24}{32} \approx 0.75 \quad (2)$$

Where:

- CQ_P is the number of properties extracted from the CQs.
- T_P is the number of properties of the Teleontology.

6.2.2 Teleontology vs Reference Ontologies

Given a set of (RO), the etype coverage (Cov_E) of the Teleontology (T) is:

$$Cov_E(RO_E) = \frac{|RO_E \cap T_E|}{RO_E} = \frac{5}{76} \approx 0.06 \tag{3}$$

Where:

- RO_E is the number of etypes extracted from the ROs.
- T_E is the number of etypes of the Teleontology.

The entity type coverage (Cov_E) of 0.06 reflects the large number of entity types in the reference ontology, which comes from OpenStreetMap. Given the broad scope of OSM, only a small subset was relevant to our project, and some of the eTypes derived from OSM are used for the visualization and a geographical content that we do not implement. This low coverage value highlights the extensive nature of OSM rather than a limitation of our teleontology, which was designed to focus on the most pertinent entity types.

Given a set of (RO), the property coverage (Cov_p) of the Teleontology (T) is:

$$Cov_p(RO_p) = \frac{|RO_p \cap T_p|}{RO_p} = \frac{3}{8} = 0.375 \tag{4}$$

Where:

- RO_p is the number of properties extracted from the ROs.
- T_p is the number of properties of the Teleontology.

6.3 Data Layer Evaluation

The evaluation of the Knowledge Graph (KG) data layer aims to assess its connectivity and *density*. This is examined across two dimensions: entity connectivity 6.3.1, which measures the degree to which entities are linked to each other, and property connectivity 6.3.2, which evaluates how closely entities are connected to their associated property values.

6.3.1 Entity Connectivity

$EC(X)$ Entity Connectivity for the EType X:

$$EC(x) = \frac{\sum_{Y=1}^{N}(X,Y)}{OP(x)} \tag{5}$$

	Location	Building	Hotel	Restaurant	Ski Rental	Sport Location	Cycleway	Hiking Trail	Rock Climbing Route	Ski Trail	View
Location	**16606**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Building	**0**	**15525**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Hotel	**0**	**0**	**16868**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Restaurant	**0**	**0**	**0**	**7424**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Ski Rental	**0**	**0**	**0**	**0**	**48**	**0**	**0**	**0**	**0**	**0**	**0**
Sport Location	**0**	**0**	**0**	**0**	**0**	**12969**	**0**	**0**	**0**	**0**	**0**
Cycleway	**0**	**0**	**0**	**0**	**0**	**0**	**5808**	**0**	**0**	**0**	**0**
Hiking Trail	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**6065**	**0**	**0**	**0**
Rock Climbing Route	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**10243**	**0**	**0**
Ski Trail	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**6595**	**0**
View	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1274**

Table 4: Connectivity Matrix.

$$EC(Location) = \frac{0}{0} = 0 \qquad EC(Building) = \frac{0}{0} = 0$$
$$EC(Hotel) = \frac{0}{0} = 0 \qquad EC(Restaurant) = \frac{0}{0} = 0$$
$$EC(SkiRental) = \frac{0}{0} = 0 \qquad EC(CycleWay) = \frac{0}{0} = 0$$
$$EC(HikingTrail) = \frac{0}{0} = 0 \qquad EC(RockClimbingRoute) = \frac{0}{0} = 0$$
$$EC(SkiTrail) = \frac{0}{0} = 0 \qquad EC(View) = \frac{0}{0} = 0$$
$$EC(SportLocation) = \frac{0}{0} = 0$$

Where:

- (X, Y) is a cell in the connectivity matrix 4.
- $OP(X)$ is the number of object properties of ETypes X.

$EC(KG)$ Entity Connectivity for the whole KG:

$$EC(KG) = \sum_{X=1}^{N} EC(X) = 0 \tag{6}$$

6.3.2 Property Connectivity

$PC(X)$ Property Connectivity for EType X:

$$PC(X) = \frac{(X, X)}{DP(X)} \tag{7}$$

$$PC(Location) = \frac{16606}{3} \approx 5535 \text{ (with respect to 5859 entities)}$$
$$PC(Building) = \frac{15525}{9} \approx 1725 \text{ (with respect to 1814 entities)}$$
$$PC(Hotel) = \frac{16868}{17} \approx 992 \text{ (with respect to 1000 entities)}$$
$$PC(Restaurant) = \frac{7424}{10} \approx 742 \text{ (with respect to 809 entities)}$$
$$PC(SkiRental) = \frac{48}{10} \approx 5 \text{ (with respect to 5 entities)}$$
$$PC(SportLocation) = \frac{12969}{4} \approx 3242 \text{ (with respect to 3732 entities)}$$
$$PC(Cycleway) = \frac{5808}{9} \approx 645 \text{ (with respect to 732 entities)}$$
$$PC(HikingTrail) = \frac{6065}{7} \approx 866 \text{ (with respect to 1000 entities)}$$
$$PC(RockClimbingRoute) = \frac{10243}{12} = 854 \text{ (with respect to 1000 entities)}$$
$$PC(SkiTrail) = \frac{6595}{7} \approx 942 \text{ (with respect to 1000 entities)}$$
$$PC(View) = \frac{1274}{5} \approx 255 \text{ (with respect to 313 entities)}$$

Where:

- (X, X) is a cell in the connectivity matrix 4.
- $DP(X)$ is the number of data properties of ETypes X.

$PC(KG)$ Property Connectivity for the whole KG:

$$PC(KG) = \sum_{X=1}^{N} PC(X) = 15803 \tag{8}$$

6.4 Query Execution

To assess the suitability of the final Knowledge Graph in achieving the project objectives, a set of competency queries was executed. These queries were designed to evaluate the Knowledge Graph ability to retrieve relevant information and verify its structural and semantic consistency. The following competency queries were tested, with each query having a corresponding link to the GitHub repository containing the complete query implementation:

- **1.2** Ski Rental Based On Ski Resort 6.4.1
- **1.8** Ski Resort With Black Slopes 6.4.2
- **2.1** Family Bike Path 6.4.3
- **2.6** Hotel Positions 6.4.4
- **4.1** Various Skill Level Sector 6.4.5
- **5.1** Crags Harder Than 7a 6.4.6
- **5.3** Climbing Area With View 6.4.7

6.4.1 Ski Rental Based On Ski Resort

Where can they rent ski equipment?

This query retrieves ski trails, their associated ski resorts and nearby ski rental shops based on a given search term. First, it filters ski trails whose name or associated ski resort contains the specified search term, ensuring case-insensitive matching. Then, it extracts the geographical coordinates of both ski trails and ski rentals, converting them into decimal values for comparison. Finally, it identifies ski rentals located within a defined proximity of a ski trail by applying a tolerance range to latitude and longitude values. The execution of this query allows for the identification of rental shops situated near ski trails or resorts.

6.4.2 Ski Resort With Black Slopes

Which ski resort has black slopes?

This query retrieves the names of ski resorts that have at least one ski trail classified as *advanced* in difficulty. It begins by identifying all entities of type *Ski_trail* and filtering them based on the difficulty_level property, selecting only those marked as *advanced*. The query then extracts the associated *ski_resort* for each of these trails, ensuring that only distinct results are returned. Finally, the results are sorted in alphabetical order to facilitate readability.

6.4.3 Family Bike Path

Are there family-friendly bike trails in Trentino?

This query retrieves distinct cycleway names from a Knowledge Graph where the entities are of type *Cycleway* and have a difficulty level set to *easy*. The query first identifies entities that are classified as *Cycleways*. Then, it filters those entities by the *difficulty_level* property, selecting only those marked as *easy*. Finally, the query extracts the name of each matching cycleway and returns a list of unique cycleway names.

6.4.4 Hotel Positions

Where is an hotel in which the family can stay for some nights?

This query retrieves information about hotels from the Knowledge Graph. It identifies entities of type *Hotel* and optionally extracts attributes such as the hotel *name*, *municipality*, *postal code*, *street*, and *house number*. The query uses the OPTIONAL keyword to ensure that even if some attributes are missing for certain hotels, the query will still return results for those hotels. The retrieved hotel details are ordered by the hotel name.

6.4.5 Various Skill Level Sector

What rock climbing crags in Trentino offer routes suitable for various skill levels?

This query calculates the range of difficulty levels for climbing routes within each sector. It first retrieves all climbing routes, identifying their sector and difficulty degree. The query then computes the difference between the maximum and minimum difficulty levels within each sector, using *MAX* and *MIN* functions. The results are grouped by sector and ordered by the calculated difficulty range in descending order. Finally, the query limits the output to the top 5 sectors with the highest difficulty ranges.

6.4.6 Crags Harder Than 7a

Which climbing crags in Trentino offer grades from 7a and above?

This query retrieves the names of crags that have climbing routes within a specific difficulty range. The query defines a preferred lower bound (49.0, equivalent to a French 7a difficulty) and an upper bound (100.0). It then retrieves climbing routes filtering those whose difficulty falls within the defined range. The difficulty values are converted from their original format (French difficulty) to integers using *xsd:decimal* for comparison. The query returns distinct crag names where the climbing routes meet the specified difficulty criteria.

6.4.7 Climbing Area With View

Which climbing areas have scenic viewpoints or cultural land-marks nearby for added exploration?

This query retrieves information on climbing routes and their corresponding views based on geographic proximity. It first identifies climbing routes, extracting the *crag*, *sector*, and *coordinates* of each route. The coordinates are parsed to separate the latitude and longitude values. The query then identifies view entities, retrieving their names and coordinates. The view coordinates are parsed similarly and the query results are filtered to ensure that the coordinates of the climbing routes and views are within a specified proximity, allowing for a tolerance of ±*0.008* in both latitude and longitude. The results are ordered by *crag*, *sector*, *climbing route name*, and *view name*.

7 Metadata Definition

This section provides the definitions of all metadata associated with the various resources produced throughout the project. The metadata described here refer to both the final outcomes and the intermediate results of each project phase, including language, schema, and standardized values of data sources. Defining these metadata is essential to facilitate the distribution and sharing of resources via data catalogs. Therefore, it is crucial to outline that these metadata will be published in the DataScientia catalog to ensure the accessibility and reusability of the resources.

The structure of this section is organized to describe the metadata related to all types of resources generated during the project as follows:

- **Project Metadata description** 7.1
- **People Metadata description** 7.2
- **Dataset Metadata description** 7.3

7.1 Project Metadata Description

This category includes metadata attributes that provide essential information about the identity, scope, and timeline of the project. The *ds:prjTitle* attribute encodes the name of the DataScientia project as a string in natural language. The *ds:prjURL* attribute represents the dereferenciable URL of the DataScientia project. The *ds:prjKeywords* attribute includes a set of natural language keywords that capture the main theme of the project, helping to classify it. The *ds:prjDescription* attribute provides a textual description of the project, summarizing its objectives and activities and the *ds:prjOsbervation* report some issues and characteristics of the project. The table can be found at 5, and the original spreadsheet can be found here.

Title	URL	Keywords	Type	Description	Start Date	End Date	Funding Agency	Input	Output	Coordinator	Observations
Trentino Tourist Facilities	...	Outdoor activities, nature, tourism	KG Construction	…	October 14^{th} 2024	January 31^{th} 2025	None	Datasets	KG	Fausto Giunchiglia	…

Table 5: Project Metadata table.

7.2 People Metadata Description

This category includes metadata attributes related to the individuals involved in the iTelos project, which is essential for tracking and recognizing their contributions throughout the project. The table can be found at 6, and the original spreadsheet can be found here.

Identifier	First Name	Last Name	Email	Nationality	Gender	Affiliation	Personal Web Page
247204	Lorenzo	Dongili	lorenzo.dongili@studenti.unitn.it	IT	M	University of Trento	None
248457	Giulia	Grotto	giulia.grotto@studenti.unitn.it	IT	F	University of Trento	None
247199	Stefano	Sacchet	stefano.sacchet@studenti.unitn.it	IT	M	University of Trento	None

Table 6: People Metadata table.

7.3 Data Resources Metadata Description

This category includes metadata attributes that describe dataset resources where we found the initial dataset from which we construct our KG. These attributes help define the dataset identity, accessibility, ownership, and technical specifications. The *ds:DatLicense* attribute specifies the dataset license, ensuring clarity on usage rights. The *ds:DatURL* provides a dereferenceable link to access the dataset. The *ds:DatPublisher*, *ds:DatCreator*, and *ds:DatOwner* attributes record information about the dataset publisher, creator, and owner, respectively. The *ds:DatLanguage* attribute indicates the natural language(s) used in the dataset. We only had two datasets in Italian, namely *hotels* and *ski slopes*, which were translated into English in the *Information Gathering* phase. The *ds:DatName* provides the dataset's name in natural language, while the *ds:DatPublicationTimestamp* records the date and time of its publication in the catalog with its respective version in *ds:DatVersion* attribute. The *ds:DatDescription* offers a textual explanation of the dataset content and for the OSM data extractions we made, the used queries. Finally, the *ds:DatFileFormat* attribute encodes the dataset format, ensuring clarity on its structure and usability. In our case, we found only csv files. The table can be found at **??**, and the original spreadsheet can be found here.

8 Open Issues

The project successfully respected the initial scheduling, completing the planned phases within the expected deadline. However, while significant progress was made, the final results did not fully satisfy the original purpose. Specifically, two aspects were not covered: climbing rental shops, mountain huts, and lake tourism. These elements were initially considered part of the project scope but were not included due to data availability constraints.

One of the key challenges encountered was ensuring the completeness and consistency of the data across all datasets. Some attributes were either missing, incomplete, or non-updated, requiring additional validation steps that could not be carried out within the available time. Moreover, significant effort was spent on refining the datasets by adding attributes such as coordinates, altitude, and other relevant information. While progress was made, further testing and optimization are required to enhance data retrieval and usability. Furthermore, challenges emerged in processing and standardizing the various datasets, particularly in harmonizing different structures. Some inconsistencies in language resources were addressed by translating datasets into English, but a more systematic approach is needed for future iterations.

Finally, it is also important to note that the project started with a very specific domain of interest, which was later refined based on the data we found. The adjustments made during the process allowed us to align the project with the available resources while respecting our initial purpose. While solutions were considered for several of these challenges, time constraints prevented their full implementation, leaving room for further improvements in future iterations.