



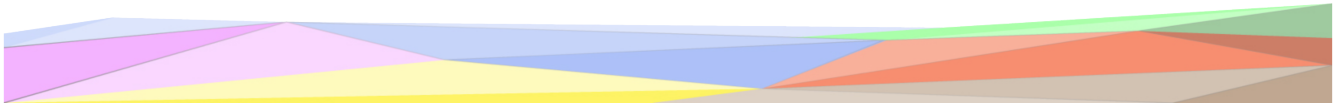
Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

# Trentino Territory & Tourism Facilities

| Document Data: | Reference Persons: |
|---|---|
| January 24, 2024 | Marina Bueno García, Lucía Trillo Carreras |





# Index:

# 1 Introduction

# 2 Purpose and Domain of Interest (DoI)

## 2.1 Project purpose

The purpose of this project is to provide in a single place all the information that is available on nature tourism in the Trentino territory. To provide that service we will build a knowledge graph (KG) that provides all the information about the natural tourism attractions such as lakes, natural parks, camp sites, waterfalls... and how to get there via bus/train/bike/taxi.

## 2.2 Project Domain of Interest

The domain of interest for this project is natural tourism in the 176 municipalities of the Trentino province 1. There is no temporal boundary as the region receives tourism throughout the cold (ski tourism) and hot seasons.

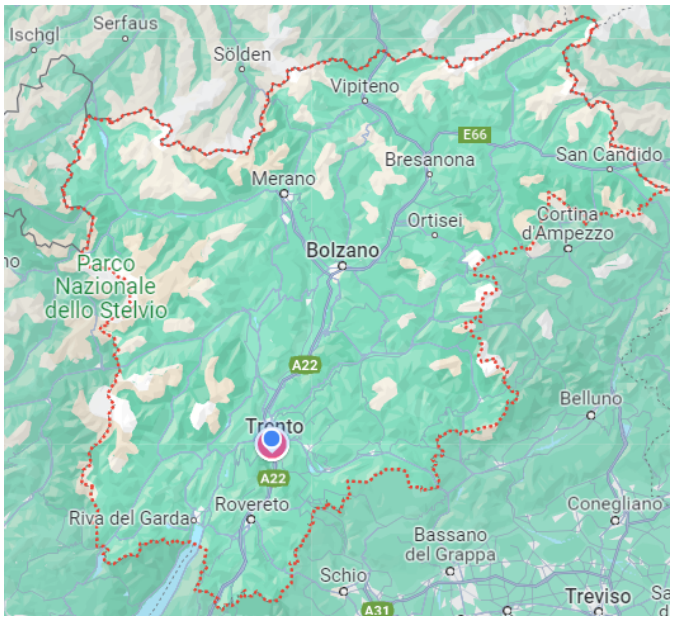



Figure 1: Trentino Autonomous Province

# 3 Project Development

To achieve the project's purpose described before, we divide the project development into two main subsections:

### 3.1 Data Production

In the data production phase, we focus on the role of the data producer and how to acquire the necessary data to later build the KG given the purpose.

As data producers, our mission is to make this data as available and reusable as possible, (compliant with the quality and reusability guidelines defined by iTelos. 6*, or at least 5*). We will document and publish the project result on Github and on the KnowDive group so that future researchers or tourists can easily access it.

### 3.2 Data Composition

As data consumers, our mission is to compose the high quality formal resources produced into the final Knowledge Graph. We will perform this task using the iTelos methodology, after thoroughly defining our Language, Knowledge and Data. The quality of this Knowledge graph will be evaluated based on the Competency questions that will be developed in the next section.

## 4 Purpose Formalization

In this section, with the goal of identifying the possible use cases of our resources, we first divide the purpose formalization into 3 different subsections: scenarios definition, personas and competency questions (CQs). Once we have listed down the competency questions, we fill a Purpose Formalization sheet (PFsheet) to extract the concepts identifying the information entities and their properties. Finally, the last part of the purpose formalization and the first phase of the iTelos is to shape a ER model to representate formally the initial purpose.

### 4.1 Scenarios definition

In this section, we outline three distinct scenarios designed to cater to diverse preferences and interests in exploring the natural wonders of Trentino. Each scenario offers a unique perspective, guiding individuals through different levels of adventure and nature immersion:

1. Challenging Adventure: a plan to enjoy adventurous hike in Trentino's mountains, looking for challenging trails and stunning views.
2. Moderate Nature Experience: an exploration of Trentino's hills, combining nature with cultural experiences, seeking easy to moderate trails.
3. Relaxing Scenic Strolls: gentle walks around picturesque locations in Trentino, focusing on easy trails and accessible spots.

### 4.2 Personas

In this section, we introduce six diverse personas, each bringing a unique perspective and set of preferences to the exploration of Trentino's natural treasures.These personas provide a diverse lens through which we can tailor scenarios and recommendations, ensuring a personalized and enriching experience for each individual exploring Trentino.

| Id | Origin | Nature expertise | Age | Description |
|---|---|---|---|---|
| 1 | Italy,Lombardy | High | 55 | **Antonio** is a local outdoor enthusiast, familiar with Trentino's trails, seeks challenging hikes and hidden gems. He enjoys skiing after picking it up from his business trips to the Alps. He has a car. |
| 2 | Italy,Toscana | Medium | 28 | **Giulia** is a young professional traveling with her boyfriend, Matio. Giulia enjoys art, history and good coffee .She is visiting Trentino for the first time, seeks diverse trails and wants detailed information. Matio loves biking |
| 3 | Italy,Toscana | Low | 27 | **Mario** is a young professional traveling with her wife, Julia, and children by car. A tourist with minimal nature experience, wants picturesque spots, child-friendly and needs guidance on transportation options. |
| 4 | Czech Republic, Brno | High | 23 | **Veronica** is doing her Erasmus for 6 months in Trento and wants to thoroughly explore everything the region has to offer. Time is not a limitation on her travels, but money is. She's into climbing and doing challeging hikes. |
| 5 | Germany, Munich | Low | 32 | **Dalim** is a novice hiker on a weekend vacation, interested in scenic spots and gentle walks, prefers easily accessible locations. |
| 6 | Poland, Krakow | Medium | 26 | **Tymoteusz** is a nature lover, keen on exploring Trentino's beauty at a relaxed pace, enjoys both hills and mountains. He loves swimming and beautiful rocks |

### 4.3 Competency questions

In this section, we articulate a set of key competency questions that serve as the foundation for unlocking the information within the Trentino Knowledge Graph. These questions are strategically designed to cater to the diverse needs and interests of individuals exploring Trentino's natural wonders.

- What are the top three challenging trails in Trentino?
- Can you suggest a moderate hill trail with altitude lower than 2000 m?
- Are there any beginner-friendly walks near Trento?
- How can I reach the Dolomites?
- Can you provide a list of must-visit lakes in the Trentino region?
- Who can I ask for general touristic info in each comune?
- Who can I ask for information when I go to Lago di Santa Giustina?

- What public transportation options are available from Trento to the starting point of a beginner-friendly trail?
- Is there any easy hike close to Pozza di Fassa?
- Where can I buy sleeping bags and a basic tent close to my accommodation?
- What caves can be visited in Rovereto in a day trip?
- What public transportation is available from Trento to popular hiking destinations?
- How can I get to the longest ski resort?
- How can I access the different ski stations in Trento?
- Is there a bus I can take to reach the Dolomites?
- What are the top three mountains accessible by car?
- Which are the mountains that are close to the lakes?
- Can you provide a list of mountain shelter or camping sites along their chosen mountain trails?
- What is the closest gas station to every ski station?
- Which natural attraction is close to this train station (e.g., Levico Terme)?

#### 4.3.1 PF Sheet

In this table, we present a comprehensive overview of the relationships between scenarios, personas, competency questions (CQs), entities, properties, focus, and popularity within the context of our Trentino Knowledge Graph project. The table encapsulates the intricate web of connections, guiding the design and implementation of our knowledge graph.

| Scenarios | Personas | CQs | Entities | Properties | Focus | Popularity |
|---|---|---|---|---|---|---|
| 1 | 1,2,4,6 | 18 | accommodation | id, accommodationCategory, latitude, longitude, city, address | Core | Contextual |
| 2 | 1,3,5 | 3,16 | tourist_trip | id, name, touristtype, trip_origin, arrival_time, height | Core | Core |
| 1,2,3 | 2,4,5,6 | 8,12,15 | station | id, name, latitude, longitude, city, address, station_type | Common | Contextual |
| 1,2,3 | 2,4,5 | 10,13 | Store | id, name, latitude, longitude, city, address | Contextual | Common |
| 1,2,3 | 2,3,4,5,6 | 6,7 | Person_guide | id, name, email, affiliation (city) | Core | Common |
| 1,2,3 | 2,3,4,5,6 | 6,7 | Person_tourist | name, birthdate, gender, nationality, description (nature_expertise), accommodation | Contextual | Common |
| 1,2,3 | 1,2,3,4,5,6 | 1,2,4,5,9 | Landform | id, address, city, name, latitude, longitude, landform_type | Common | Contextual |
| 1,2,3 | 1,3 | 19 | GasStation | id, name, latitude, longitude, city, address | Common | Contextual |
| 1,2,3 | 1 | 14 | SkiResort | id, name, latitude, longitude, city, address, length | Core | Contextual |
| 1,2,3 | 1,2,3,4,5,6 | 5,6,9,11 | City(Region) | id, name | Common | Core |

#### 4.3.2 ER model

In the landscape of Trentino's natural tourism, we have designed a robust Entity-Relationship (ER) model to encapsulate the information tpresented . The ER model is the backbone of our knowledge graph, crafted to establish meaningful connections between entities, personas, scenarios, and competency questions. It serves as the architectural blueprint that not only structures our data but also provides a comprehensive framework for users to navigate and explore the diverse facets of Trentino's natural tourism.

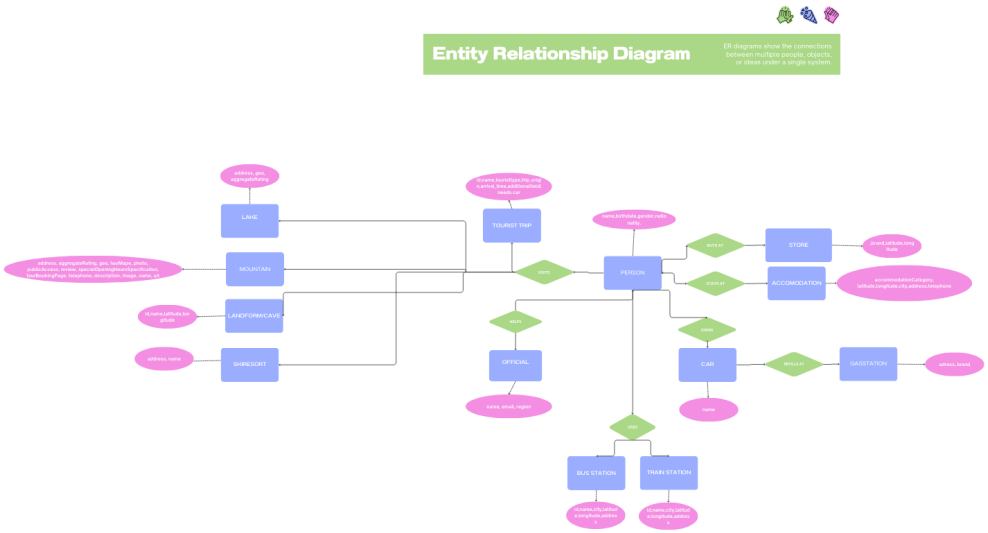



Figure 2: ER model

# 5 Information Gathering

In this phase of the project both, the Data Producer and the Data Consumer, play an important role. From the producer side, the aim is to collect informal resources from sources with an higher level of heterogeneity and from the consumer side, the aims is to collect formal resources to be composed with the objective of building the final KG.

## 5.1 Source identification

The original sources of information for the knowledge and the data layer are the following:

- Knowledge layer
  - Formal: Open Street Map - Trentino Territory Lightweight Ontology
  - Standard references:
    - SCHEMA.ORG describes the hierarchy of different Entity types and data types.
    - SCHEMA.ORG LOV schema of the schema.org vocabulary
    - GTFS STATIC is the General Transit Feed Specification, that allows to get data in the correct format. This link describes the format and structure of the files.
    - GTFS LOV describes the GTFS vocabulary.
    - GTFS UPGRADED (Subashish) describes the GTFS ontology
    - GEOSPATIAL ONTOLOGY (Subashish) is a data model for the geospatial domain.

    * TIME ONTOLOGY is a vocabulary for temporal entities such as time intervals, their properties and relationship.

- Data layer
  - Formal: Trentino OSM places
    This dataset contains information scraped from Open Street Maps about the Trentino region: springs, peaks, cave entrances and other natural wonders are listed here.
  - Semi-Formal:KGE22 - Trentino Tourist Facilities This source contains individual datasets of bus,train and bike stops; taxi companies; bars and restaurants; campsites, hotels and vacation houses;gas stations and souvenir shops; and museums, ski stations and natural attractions.
  - Informal:
    * ISTAT Turismo This source contains contact emails of experts in the ISTAT declaration (economic activity classification) by municipality.
    * OPEN DATA TRENTINO: This source contains 6.556 different datasets of the Trentino province of different data formats and topics.

### 5.2 Resources collection and scraping

However, as stated in our purpose, we mainly aim to cover everything related to nature tourism (lakes, mountains, hikes, ski and the transportation to these places), so we extracted the data that fits our needs the best in the following files:

| File name | Source | Topic | Classification |
|---|---|---|---|
| camping_site.csv | Trentino Tourism KG 2022 | Accomodation | Common |
| ski_acommodation.csv | Trentino Tourism KG 2022 | Accomodation | Common |
| holiday_house.csv | Trentino Tourism KG 2022 | Accomodation | Common |
| bus_station.csv | Trentino Tourism KG 2022 | Station | Core |
| train_station.csv | Trentino Tourism KG 2022 | Station | Core |
| gas_station.csv | Trentino Tourism KG 2022 | Gas Station | Core |
| skiresort.csv | Trentino Tourism KG 2022 | Ski resort | Contextual |
| person_guide.csv | ISTAT Turismo | Local information reference people | Common |
| region.csv | ISTAT Turismo | Region | Common |
| mountain.csv | OSM Places | Landform | Contextual |
| lake.csv | Custom | Landform | Contextual |
| caves.csv | OSM Places | Landform | Contextual |
| store.csv | OSM Places | Mountain equipment stores | Core |
| tourist_trip.csv | Komoot | Trails and hikes | Contextual |
| person_tourist.csv | Custom | Our personas | Contextual |

The datasets and cars were created by us, the latter was eliminated because it no longer helped fulfill our purpose (it made more sense as a data attribute).All resources were distributed except for those published online on the Komoot website. We tried several approaches, we asked the owners for the dataset, but they only provide it to companies. Next we tried to scrap the data from the source with BeautifulSoup4, but it was not possible. So we moved on to the next approach, creating our own data using the information provided on the Komoot website.

The progress of the scraping attempts can be seen in the following notebook.

### 5.3 Data cleaning and formatting

The data cleaning process involved the following tasks:

1. Deleting columns of entities that are not relevant to our purpose

   This step also included deleting entries of dataset containing information not related to Trentino territory such as data regarding Bolzano.

2. Modifying word accents

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Numero | Codice comune ISTAT | Comune | Azienda per il turismo | Referente sta e-mail | | | | | |
| 2 | 3 | 22003 | Aldeno | Trento, Monte Bondone e Altopiano | Pierangela Pe pierangela.pedron@trento.info | | | | | |
| 3 | 2 | 22002 | Albiano | Trento, Monte Bondone e Altopiano | Pierangela Pe pierangela.pedron@trento.info | | | | | |
| 4 | 5 | 22236 | Altopiano della Vigolana | Altipiani cimbri e Vigolana | Verena Pasca office@altpecimbra.it | | | | | |
| 5 | 8 | 22006 | Arco | Garda Trentino, Valle di Ledro, Terr | Marica Cattoi marica.cattoi@gardatrentino.it claudia.ventura@gardatrentino.it | | | | | |
| 6 | 33 | 22042 | Carisolo | Madonna di Campiglio, Pinzolo, Val | Claudia Trogg sales@campiglidolomiti.it | | | | | |
| 7 | 30 | 22038 | Canal San Bovo | San Martino di Castrozza, Primiero e | Wilma Loss infovanoi@sanmartino.com | | | | | |
| 8 | 26 | 22034 | Caldonazzo | Valsugana, Tesino e Valle dei Moche | Anna Dorigon anna.dorigoni@visitvalsugana.it | | | | | |
| 9 | 28 | 22036 | Campitello di Fassa-Ciampedel | Val di Fassa | Nicoletta Artc infocanazei@fassa.com | | | | | |
| 10 | 27 | 22035 | Calliano | Rovereto, Vallagarina e Monte Bald | Matteo Zocc zocca@visitrovereto.it | | | | | |
| 11 | 32 | 22040 | Capriana | Val di Fiemme e Val di Cembra | Susanna Bone susanna.bonelli@visitfiemme.it | | | | | |
| 12 | 31 | 22039 | Canazei-CianacÃ±i | Val di Fassa | Nicoletta Artc infocanazei@fassa.com | | | | | |
| 13 | 29 | 22037 | Campodenno | Val di Non | Elisa Widman info.coredo@visitvaldinon.it | | | | | |
| 14 | 25 | 22033 | Caldes | Val di Sole | Mavie Cattoi mavie.cattoi@visitvaldisole.it | | | | | |
| 15 | 61 | 22083 | FiavÃ | Garda Trentino, Valle di Ledro, Terr | Marica Cattoi marica.cattoi@gardatrentino.it | | | | | |
| 16 | 68 | 22093 | Giustino | Madonna di Campiglio, Pinzolo, Val | Claudia Trogg sales@campiglidolomiti.it | | | | | |
| 17 | 65 | 22090 | Frassilongo-Garait | Valsugana, Tesino e Valle dei Moche | Anna Dorigon anna.dorigoni@visitvalsugana.it | | | | | |

Figure 3: Example of input dataset

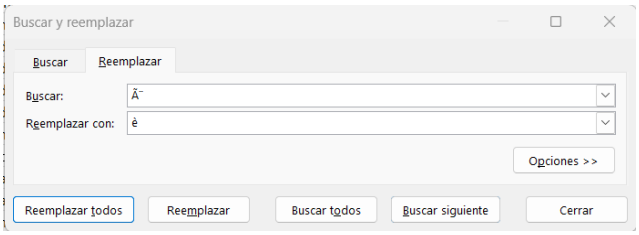



Figure 4: Substitution by ASCII characters

3. Removing repeated entries and null values
4. Transforming data from the form POINT(longitude, latitude) into two different columns and using this information to create 2 new columns: city and address

   A vast majority of our data is location-based, and the formating and information available is very varied amongst entities. In an effort to normalise, we are using the geopy python library to fill in the blanks and enrich current information, adding properties such as city and address.

   In order to use it, we have modified the coordinates format (in different ways for the OSM and KGE2022 sources) to provide 2 floating points with the correct decimal numbers.

   Once this transformation was completed, we were able to get the city and address of some of our entities (bus stops, mountains, lakes...) using the longitude and latitude. The complete and annotated description of the geographical transformations can be found on this notebook.

5. Removing data that is not related with Trentino area.

Regarding formatting, for reusability and iTelos compliance reasons, all the resources have been aligned following the CSV format standard, instead of txt or xlsx formats.

### 5.4 Knowledge modeling and formal resource generation

Once we have all the datasets cleaned and formatted into csv files and in order to generate formal resources first we need to associate our datasets to a schema.

While cleaning the datasets and naming the entities, we had in mind the schemas provided in schema.org, which made the process of modeling quite easy. We have limited the properties shown as information that is both present in the datasets and relatable to an existing schema. We used the following schemas:

| Concept | Schema | Route | Properties |
|---|---|---|---|
| Campsites,shelters, holiday houses | Accomodation | Thing > Place > Accommodation | id, accommodationCategory, latitude, longitude, city, address, telephone |
| Panoramic viewpoints | TouristTrip | Thing > Intangible > Trip > TouristTrip | id, name, touristType, trip_origin, arrival_time, height |
| Our personas, guides and information contacts | Person | Thing > Person | name, birthdate, gender, nationality, description, affiliation, email |
| Mountain equipment stores | Store | Thing > Organization > LocalBusiness > Store / Thing > Place > LocalBusiness > Store | id, name, longitude, latitude |
| Mountains, lakes and Caves | Landform | Thing > Place > Landform | id, name, longitude, latitude, city, address |
| Public transportation: trains and buses | Civic Structure | Thing > Place > CivicStructure | id, name, longitude, latitude, city, address |
| Gas station | GasStation | Thing > Organization > LocalBusiness > AutomotiveBusiness > GasStation | id, name, longitude, latitude, city, address |
| Ski resort | SkiResort | Thing > Place > LocalBusiness > SportsActivityLocation > SkiResort | id, name, longitude, latitude, city, address, length |
| Region | AdministrativeArea | Thing > Place > AdministrativeArea | id, name |

The tool used for modeling the schemas is Protégé and the outcome is a RDF-OWL schema file (see Fig. 5). Once we have the schema, we map the data layer and the knowledge layer using Karma (see Fig. 6), producing RDF files whose content is copied into a ttl file.

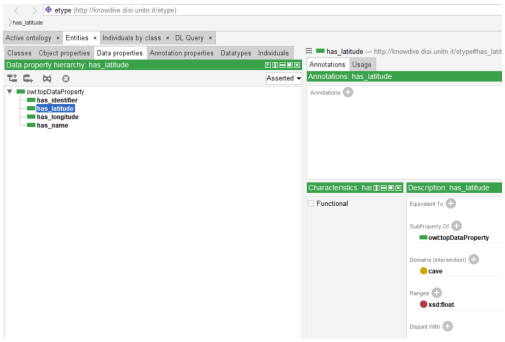


Figure 5: Protegé

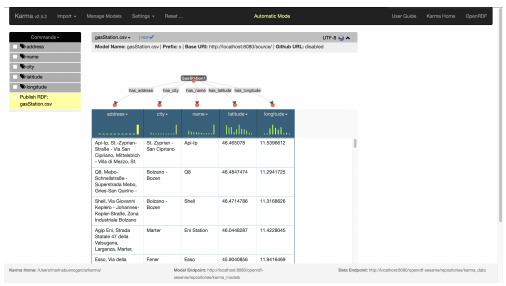


Figure 6: Karma

Throughout this phase, our PFsheet and ER mapping have suffered modifications as we have worked with the data in detail. The hikes are not comprised of natural resources (lakes,mountains,caves) and instead are an alternative element that tourists can visit. We also realized having the opening hours of the stations was nonsensical in some cases (like bus stops in the street). We are aware that iTelos is an itherative metodology and as such, we are constantly revising and makeing sure every new step is coherent with the previous ones.

# 6 Language Definition

The aim of this section is to fix the language (concepts and words) used to represent the information required to satisfy the project purpose. In order to do so, first we need to identify all the concepts for our specific purpose (entity types, data and object properties), found in our PFsheet. Afterwards, we performed the UKC alignment of these concepts, and the Language resource building. The result of this process, the formal concepts definitions, have been collected and presented in the tables 1 and 2.

Throughout all our decisions, we had in mind the balance between creating new definitions (that could serve better our individual context) or using existing ones (less catered to our needs, but more reusable). As a consequence of this, and of basing the knowledge layer on existing schema.org types, most of the concepts were already formalized in the UKC, except for the accommodation, transportation and formation type and some object properties such as acts_in and part_of (highlighted in blue). We performed dataset filtering, removing all the elements (entities, attribute, ETypes and properties) which haven't been defined before by any of the concepts formalized.

When defining the concepts, we had the opportunity of clearing up definitions that were uncertain from schema.org for our specific purpose. For instance, additionalfield was the only relevant parameter in schema.org to determine if a route was accessible by car. The word lemma "accessible" conveys the information more successfully. Other example of this is the word lemma "jurisdiction" referring to the provinces a tourism official is responsible for, defined in the knowledge layer as "affiliation". More examples can be found highlighted in purple in both tables.

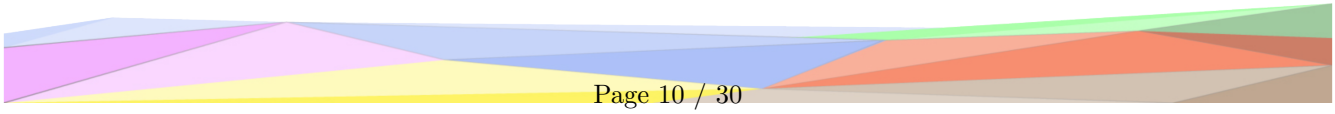
Our role as data consumers is more crucial in this step. As data producers, we have made sure to correctly define all the concepts within the entity type. As data consumers, we have analyzed in parallel concepts of all entity types to ensure coherency, reusability and clarity. Finally, prioritising reusability.

We were able to observe how the same entity type in the Knowledge layer, person, needed two different word lemmas, "tourist" and "guide".

On the other hand, two different entity types accommodated the same lemma and definition. This is the case of "hike" and "person", that both reflect the level of expertise in hiking required (or possessed) as a data property.

Above all, some definitions are more straightforward than others, but all convey the meaning of the word lemma it is referred to, and all provide an essential frame to understand and explain the Knowledge Graph that will be created.

Thus, the heterogeneity at language level of natural tourism in trentino territory has been handled. This has been done by defining a purpose-specific domain language (based on natural language), composed by concepts formally defined and uniquely identified (associated to a purpose-specific namespace). Additionally, the purpose-specific language resource for the final KG has been created. Finally, the data resources have been filtered and aligned with the language's concepts defined for the final KG.



| Word lemma | Schema.org | Concept label | Description |
|---|---|---|---|
| accommodation | accommodation | accommodation_GID-14254 | living quarters provided for public convenience |
| accommodation type | accommodationCategory | has_accommodation_type_GID-14001 | specific category of housing that individuals can choose from when seeking a place to stay |
| latitude | latitude | has_latitude_GID-46263 | the angular distance between an imaginary line around a heavenly body parallel to its equator and the equator itself |
| longitude | longitude | has_longitude_GID-46270 | the angular distance between a point on any meridian and the prime meridian at Greenwich |
| city | city | has_city_GID-45988 | an incorporated aministrative district established by state charter |
| address | address | has_address_GID-45803 | the place where a person or a place can be found or communicated with |
| hike | tourist_trip | hike_GID-1458 | a long walk usually for exercise or pleasure |
| identifier | id | has_identifier_GID-39085 | a symbol that establishes the identity of the one bearing it |
| name | name | has_name_GID-2 | a language unit by which a person or a thing is known |
| expertise | touristtype, description | has_expertise_GID-30843 | skillfulness by virtue of possessing special knowledge |
| origin | trip_origin | has_origin_GID-45883 | the place where something begins, where it springs into being |
| duration | arrival_time | has_duration_GID-80582 | the period in time during which something continues |
| guide | person_tourismofficial | guide_GID-34478 | someone that offers basic information or instruction |
| email | email | has_email_GID-33745 | (computer science) a system of worldwide electronic communication in which a computer user can compose a message at one terminal that is generated at the recipient's terminal when he logs in |
| jurisdiction | affiliation | has_jurisdiction_GID-46238 | in law; the territory within which power can be exercised |
| tourist | Person_tourist | tourist_GID-57726 | someone that travels for pleasure |
| birthday | birthdate | has_birthday_GID-81257 | the date on which a person was born |
| nationality | nationality | has_nationality_GID-74177 | the status of belonging to a particular nation by birth or naturalization |

Table 1: Concept identification I

| Word lemma | Schema.org | Concept label | Description |
|---|---|---|---|
| gender | gender | has_gender_GID-33966 | a grammatical category in inflected languages governing the agreement between nouns and pronouns and adjectives; in some languages it is quite arbitrary but in Indo-European languages, it is usually based on sex or animatedness |
| store | Store | store_GID-23209 | a mercantile establishment for the retail sale of goods or services |
| gas station | GasStation | gas station_GID-18641 | a service station that sells gasoline |
| ski resort | SkiResort | ski resort_GID-46553 | a resort with lodging or facilities for skiing |
| length | description | has_length_GID-28259 | the linear extent in space from one end to the other; the longest horizontal dimension of something that is fixed in place |
| formation | Landform | formation_GID-49530 | (geology) the geological features of the earth |
| transportation | CivicStructure | transportation_GID-24838 | a facility consisting of the means and the equipment for the movement of passengers or goods |
| close | is_closest_to | close_GID-84218 | not far distance in time or space or degree or circumstances |
| ~~visit~~ | ~~visits~~ | ~~visit_GID-6352~~ | ~~the act of going to see someone or place or thing for a short time~~ |
| ~~buy at~~ | ~~buys_at~~ | ~~buy at_GID-112546~~ | ~~do one's shopping at;do business with; be a customer or client of~~ |
| stay at | stays_at | stay at_GID-113409 | reside temporarily |
| ~~refill at~~ | ~~refils_at~~ | ~~refill at_GID-14002~~ | ~~action or opportunity to replenish or top up something~~ |
| act in | acts_in | act in_GID-14003 | action or process of participating, engaging, or behaving within a particular manner or context |
| part of | part_of | part of_GID-14002 | inclusion within a whole or a segment of something larger or more comprehensive |
| height | height | has_height_GID-28272 | elevation especially above sea level or above the earth's surface |
| region | region | region_GID-46452 | the extended spatial location of something |
| formation_type | additionalType | has_formation_type_GID-14004 | (geology) the different types of geological features of the earth |
| transportation_type | additionalProperty | has_transportation_type_GID-14005 | types of facility consisting of the means and the equipment for the movement of passengers or goods |

Table 2: Concept identification II

In the table 2, we can see that some rows are strikethrough. Originally, we meant them to be object properties until we realized that that kind of data was later going to be calculated when doing the queries to the final knowledge graph. This changes can be appreciated more in the following section.

# 7 Knowledge Definition

## 7.1 EER Model

The correct development of this phase is essential to achieve a seamless integration of the data in the next one. Therefore, we have very closely examined our ER model and made some necessary changes. For the EER, we have also included the cardinality relationships(1 to 1, N to 1, 1 to N and M to N).

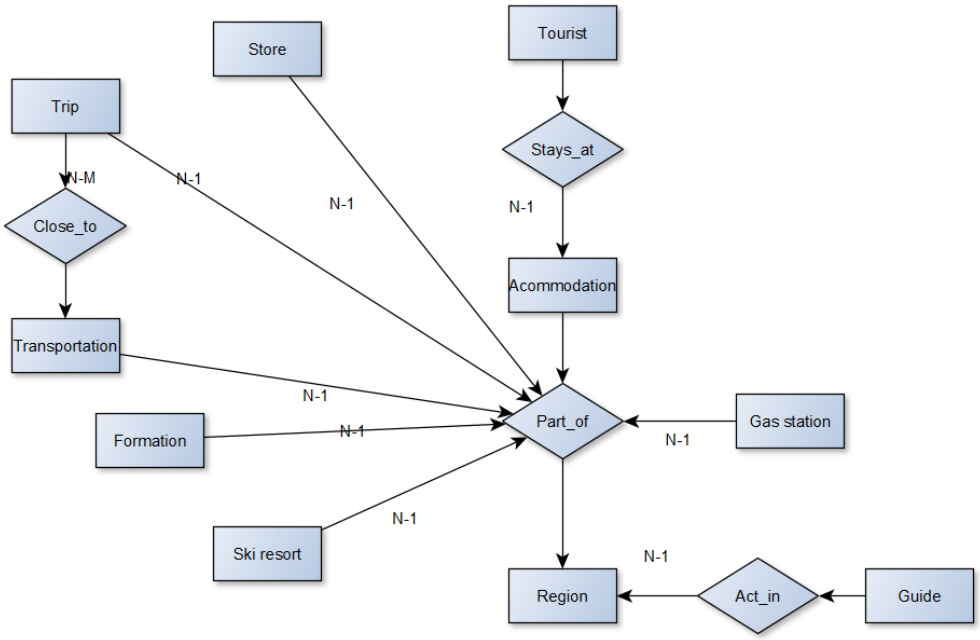



Figure 7: EER model (phase 1)

- Creating Entity types: We created father entity types to organize our information and collect common properties: transportation (father of bus and train station), and formation (father of cave, lake and mountain).Also, we were missing a region entity type to connect to every landform,hike,acommodation,shop and station within Trentino.
- Eliminating Entity types: The Entity type car does not advance our purpose, and so we have decided to eliminate it and substitute it for a data property in persona defining if the person possesses a vehicle.
- Modifying relationships: Now, the tourist guide is not connected to the tourist, but rather to the physical landmark they have information about.The tourist, instead of the car, refills at a gas station.

We are aware of the three main limitations of ontologies: unknown origin of data and object properties, unknown step between ER and EER, and missing situational context. We will try to

reduce them as much as possible by accompanying our ER and KG with documentation, like this report, and by providing the appropiate metadata.

### 7.2 Phases of the EER Model

As mentioned before, we originally defined object properties such as close_to, buy_at, visits and refills_at. However, we later realized that those are functionalities that we can calculate querying our KG but we didn't need and we shouldn't, for reusability purposes, calculate in advance.

Figure 14 shows the first phase of our final knowledge graph with all the entity types and correct object properties. And figure 15 shows the phase 2 of our KG with all the parameters (in pink) that can be calculated thanks to out KG.

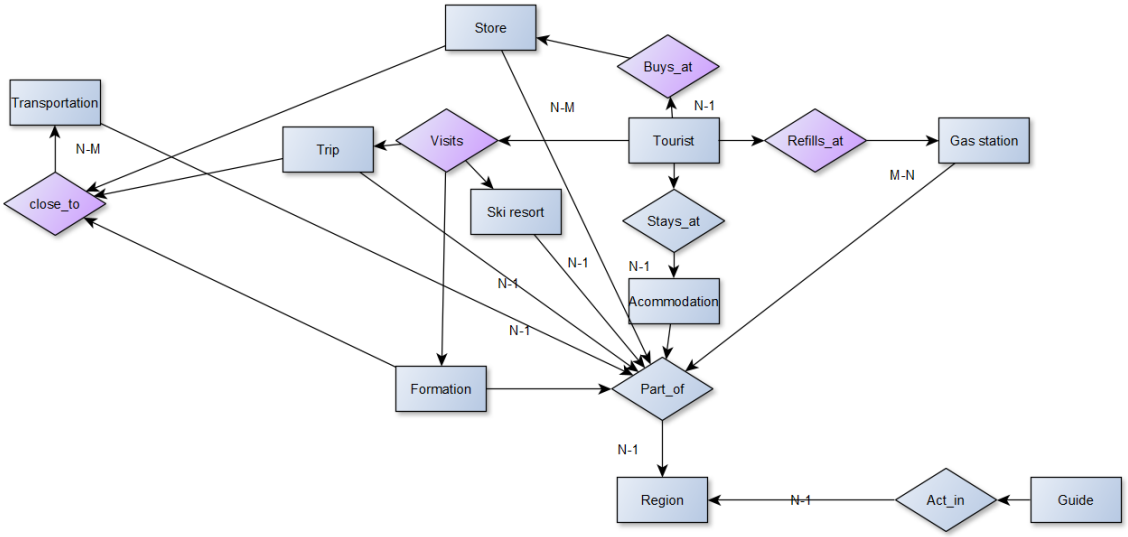



Figure 8: Updated EER model (phase 2)

### 7.3 Top-Down: Ontology

In the producer side, the aim is to model an interoperable ontology for each dataset using Protegé. . At the consumer side, the objective is to model a single unique interoperable (reusable) ontology, for the single composed final KG. Both tasks are based on the reuse of already existing formal (standard) ontologies.We performed both tasks, making sure our Language was alligned, an each entity type had its Language definition. We also showcased the two entity hierarchies present in our dataset: transportation and formation.

The ontologies for guide, hike and transportation were created by us, the other 12 individual ontologies were based in schema.org and renamed according to the word lemmas defined in Section 6. The ontologies of formation and cave are based on the same schema.org ontology:landform. All of these ontologies together allow the consumer side to model the reality of tourism and territory in Trentino: who the tourist is, what they can visit and how, where they can stay, shop and who can they ask for help.

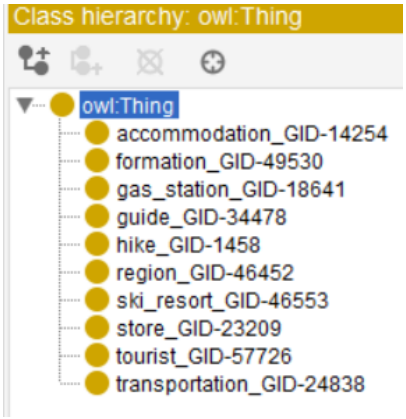



Figure 9: Ontology

### 7.4 Bottom-up : Teleology

For the teleology, we revised our 20 competency questions and from there revised the etypes and properties relevant to be modelled. This step was done using Protegé.

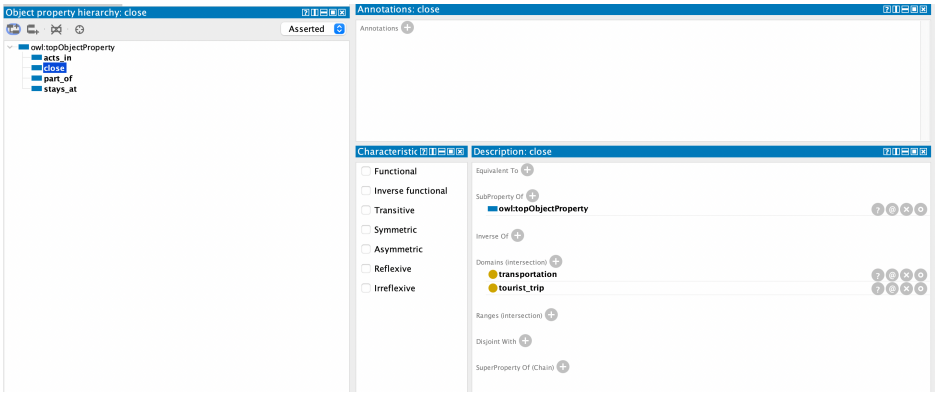
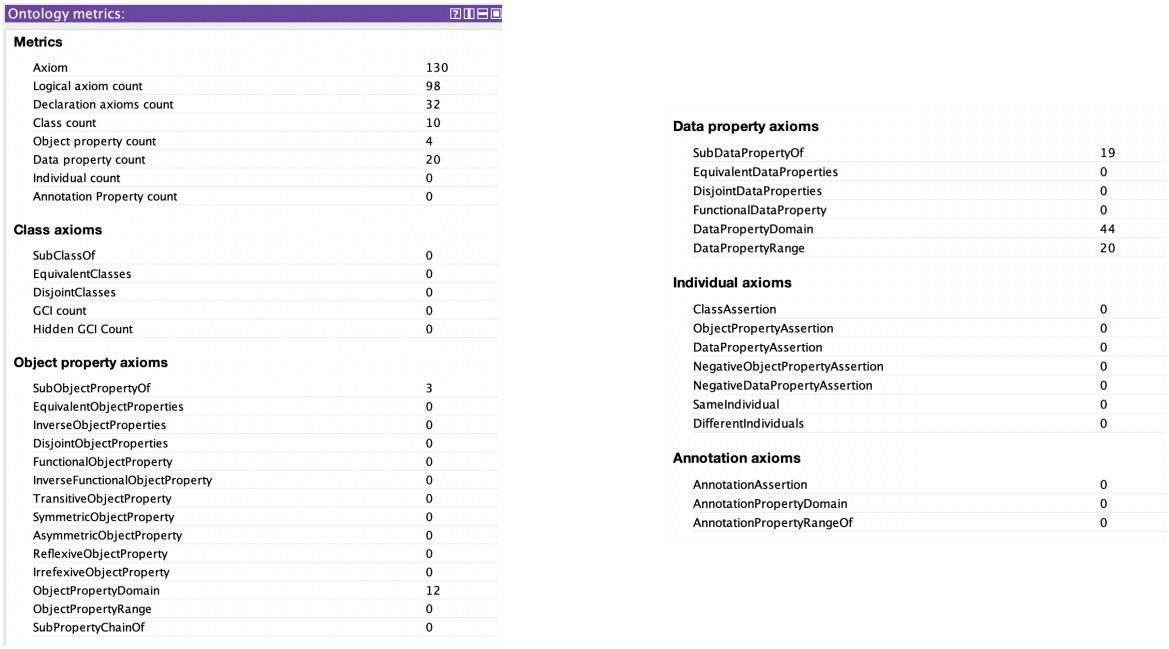



Figure 10: Teleology

The updated list of 20 competency questions is:

- What are the top three challenging trails in Trentino?
- Can you suggest a moderate hill trail with altitude lower than 2000 m?
- Are there any beginner-friendly walks near Trento?
- How can I reach the Dolomites?
- Can you provide a list of must-visit lakes in the Trentino region?
- Who can I ask for general touristic info in each comune?
- Who can I ask for information when I go to Lago di Santa Giustina?
- What public transportation options are available from Trento to the starting point of a beginner-friendly trail?

- Is there any easy hike close to Pozza di Fassa?
- Where can I buy sleeping bags and a basic tent close to my accommodation? (I'm Antonio)
- How many caves, mountains, and lakes can be visited in Baselga di Pine in a day trip?
- What public transportation is available from Trento to popular hiking destinations?
- How can I get to the longest ski resort?
- How can I access the different ski stations in Trento?
- Is there a bus I can take to reach the Dolomites?
- What are three mountains ONLY accessible by car?
- Which mountains are close to the lakes?
- Can you provide a list of mountain shelters or camping sites along popular mountain trails?
- What is the closest gas station to every ski station?
- Which natural attraction is closest to the train station Levico Terme?



Ontology metrics:

| Metrics | |
|---|---|
| Axiom | 130 |
| Logical axiom count | 98 |
| Declaration axioms count | 32 |
| Class count | 10 |
| Object property count | 4 |
| Data property count | 20 |
| Individual count | 0 |
| Annotation Property count | 0 |

| Class axioms | |
|---|---|
| SubClassOf | 0 |
| EquivalentClasses | 0 |
| DisjointClasses | 0 |
| GCI count | 0 |
| Hidden GCI Count | 0 |

| Object property axioms | |
|---|---|
| SubObjectPropertyOf | 3 |
| EquivalentObjectProperties | 0 |
| InverseObjectProperties | 0 |
| DisjointObjectProperties | 0 |
| FunctionalObjectProperty | 0 |
| InverseFunctionalObjectProperty | 0 |
| TransitiveObjectProperty | 0 |
| SymmetricObjectProperty | 0 |
| AsymmetricObjectProperty | 0 |
| ReflexiveObjectProperty | 0 |
| IrrefexiveObjectProperty | 0 |
| ObjectPropertyDomain | 12 |
| ObjectPropertyRange | 0 |
| SubPropertyChainOf | 0 |

| Data property axioms | |
|---|---|
| SubDataPropertyOf | 19 |
| EquivalentDataProperties | 0 |
| DisjointDataProperties | 0 |
| FunctionalDataProperty | 0 |
| DataPropertyDomain | 44 |
| DataPropertyRange | 20 |

| Individual axioms | |
|---|---|
| ClassAssertion | 0 |
| ObjectPropertyAssertion | 0 |
| DataPropertyAssertion | 0 |
| NegativeObjectPropertyAssertion | 0 |
| NegativeDataPropertyAssertion | 0 |
| SameIndividual | 0 |
| DifferentIndividuals | 0 |

| Annotation axioms | |
|---|---|
| AnnotationAssertion | 0 |
| AnnotationPropertyDomain | 0 |
| AnnotationPropertyRangeOf | 0 |

Figure 11: Characteristics of the teleology

The teleology created is able to accommodate all these questions. Between the main metrics of our teleology (Fig 11) we can count 8 entity types, 19 different data properties and 8 object properties. All the data properties have a specified range, as seen in data property range. The full array of metrics can be seen in the attached image.

We also ensured to perform part of the dataset alignment for the next phase, aligning the dataset previously collected, with the modelling choices operated in the knowledge Layer activity.

## 7.5 Middle-Out : Teleontology

In the teleontology phase, we directly import and join both the ontology and teleology to produce a single file, as a consumer. This is done as well using Protegé. We perform Knowledge annotation checking every concept with the Language Annotation spreadsheet.

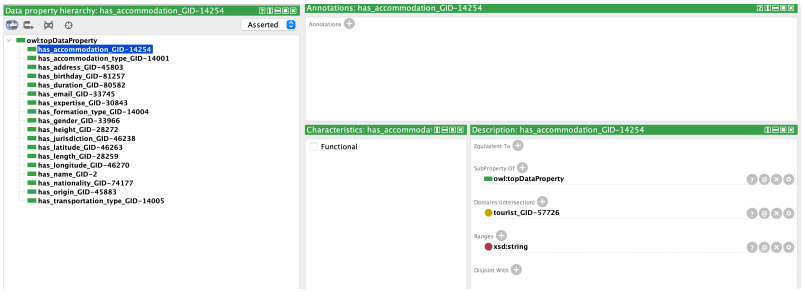



Figure 12: Teleontology

The metrics for the teleontology (Fig 13) are very similar to those of the teleology, with some exceptions. As expected, two new entity types have been added (transportation and formation, both at the top of the hierarchy) and the logical axiom count has increased. The count of data and object properties remains the same.

Ontology metrics:

| Metrics | |
|---|---|
| Axiom | 131 |
| Logical axiom count | 99 |
| Declaration axioms count | 32 |
| Class count | 10 |
| Object property count | 5 |
| Data property count | 19 |
| Individual count | 0 |
| Annotation Property count | 0 |

| Class axioms | |
|---|---|
| SubClassOf | 0 |
| EquivalentClasses | 0 |
| DisjointClasses | 0 |
| GCI count | 0 |
| Hidden GCI Count | 0 |

| Object property axioms | |
|---|---|
| SubObjectPropertyOf | 4 |
| EquivalentObjectProperties | 0 |
| InverseObjectProperties | 0 |
| DisjointObjectProperties | 0 |
| FunctionalObjectProperty | 0 |
| InverseFunctionalObjectProperty | 0 |
| TransitiveObjectProperty | 0 |
| SymmetricObjectProperty | 0 |
| AsymmetricObjectProperty | 0 |
| ReflexiveObjectProperty | 0 |
| IrrefexiveObjectProperty | 0 |
| ObjectPropertyDomain | 14 |
| ObjectPropertyRange | 0 |
| SubPropertyChainOf | 0 |

| Data property axioms | |
|---|---|
| SubDataPropertyOf | 18 |
| EquivalentDataProperties | 0 |
| DisjointDataProperties | 0 |
| FunctionalDataProperty | 0 |
| DataPropertyDomain | 44 |
| DataPropertyRange | 19 |

| Individual axioms | |
|---|---|
| ClassAssertion | 0 |
| ObjectPropertyAssertion | 0 |
| DataPropertyAssertion | 0 |
| NegativeObjectPropertyAssertion | 0 |
| NegativeDataPropertyAssertion | 0 |
| SameIndividual | 0 |
| DifferentIndividuals | 0 |

| Annotation axioms | |
|---|---|
| AnnotationAssertion | 0 |
| AnnotationPropertyDomain | 0 |
| AnnotationPropertyRangeOf | 0 |

Figure 13: Characteristics of the teleontology

Comparing the final product of the phase with the reference ontologies, we can see a more rich description of each entity type with its data and object properties, and its word lemma description.

With this, the heterogeneity at knowledge level has been handled. This effort will facilitate the mapping between data and knowledge layer in the iTelos Data Definition Phase.

# 8 Data Definition

This section is dedicated to the description of the Data Definition phase. This phase begins with the data resources cleaned and aligned and the teleontologies. The objective from the producer side is to obtain one Knowledge Graph for each dataset, and at the consumer side the objective is to obtain one full Knowledge Graph. In this phase, we are merging the knowledge and the data layers into a single structure.This means dealing with meaning heterogeneity present in the data values, to be associated to the teleontology.

Since the last phase, we came to a realization about the nature of object properties that has lead to significant changes in the EER and teleontology of the KG. We had mapped several object properties, expecting to obtain the data values that connect them while making queries within the Knowledge Graph. However, this is a mistake, as it is impossible to map properties before having the data first. Therefore, we have created two phases within our Knowledge Graph: the first one shows the current object properties. The second one illustrates in purple the future object properties that will be obtained in a later phase. Before moving on to Data Definition tasks, we make sure to update the different datasets, language sheet, ontologies, teleology and teleontology according to the first phase.

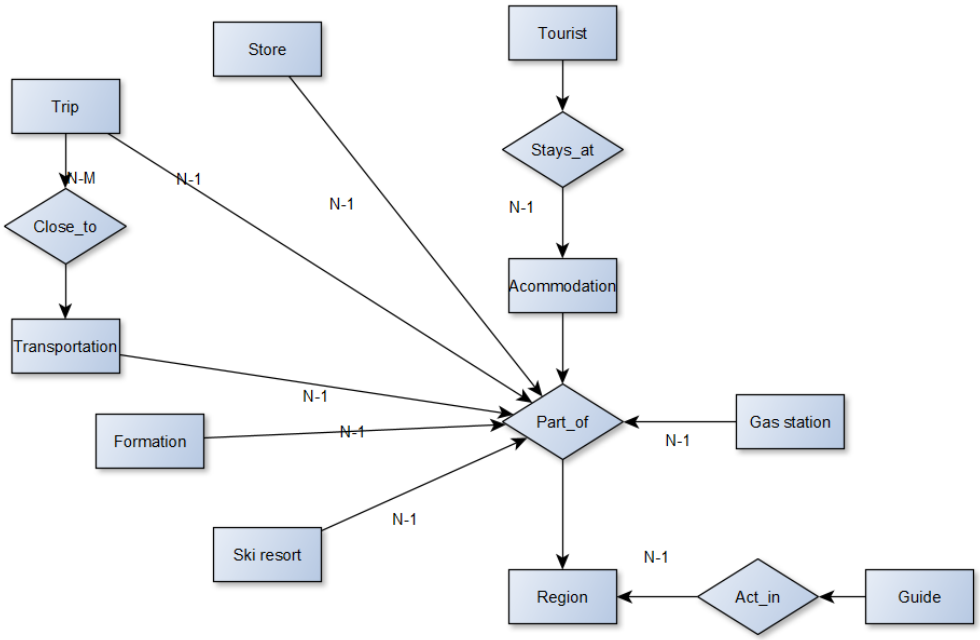



Figure 14: Phase 1 EER model

Data Definition sub activities:

- **Producer activities**: the producer activities aim at merging the knowledge layer of a single

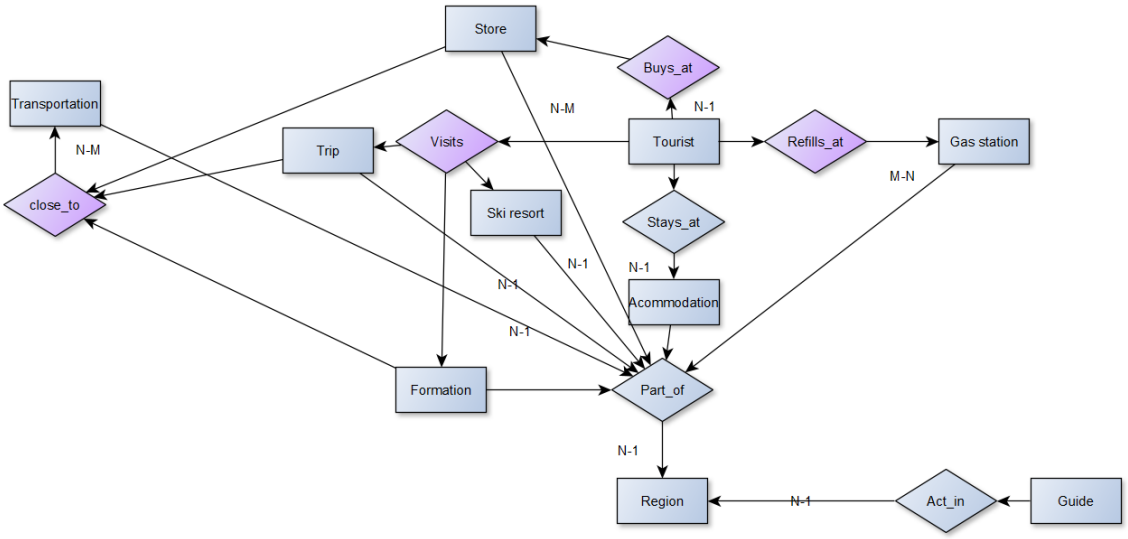



Figure 15: Phase 2 EER model

dataset with the data values present within such a dataset.

– Entity identification: For every entity type, we have defined a URI (Unique Respurce Identifier), of the type URN. The only exception to this is the entity type tourist, whose identifier is the name of the tourist. The naming has followed the structure name_number, starting at number one. Entity identification has allowed us to formally identify the entities in the datasets.

| **Concept** | **Identifier** |
|---|---|
| Acommodation | acom_n |
| Gas station | gas_n |
| Landform | landf_n |
| Tourist | Name |
| Guide | guide_n |
| Tourist trip | tourist_trip_n |
| Region | region_n |
| Ski resort | ski_n |
| Transport station | station_n |
| Store | store_n |

– Data mapping : In this section, we are concretely merging the information representation defined in the teleontology, with the relative information values in the datasets. The activity is composed by many mapping operations that concretely implement the solution to the entity matching problem. The output is 10 distinct Knowledge Graphs, one for each dataset, in RDF-Turtle (ttl) files, which are the main output of the last iTelos phase. We have performed this operation using Karma, and made sure to include the URIs of each with the format mentioned on the previous section. As it is the Producer side, the KGs files will remain separate, in order to be exploited for other purposes.

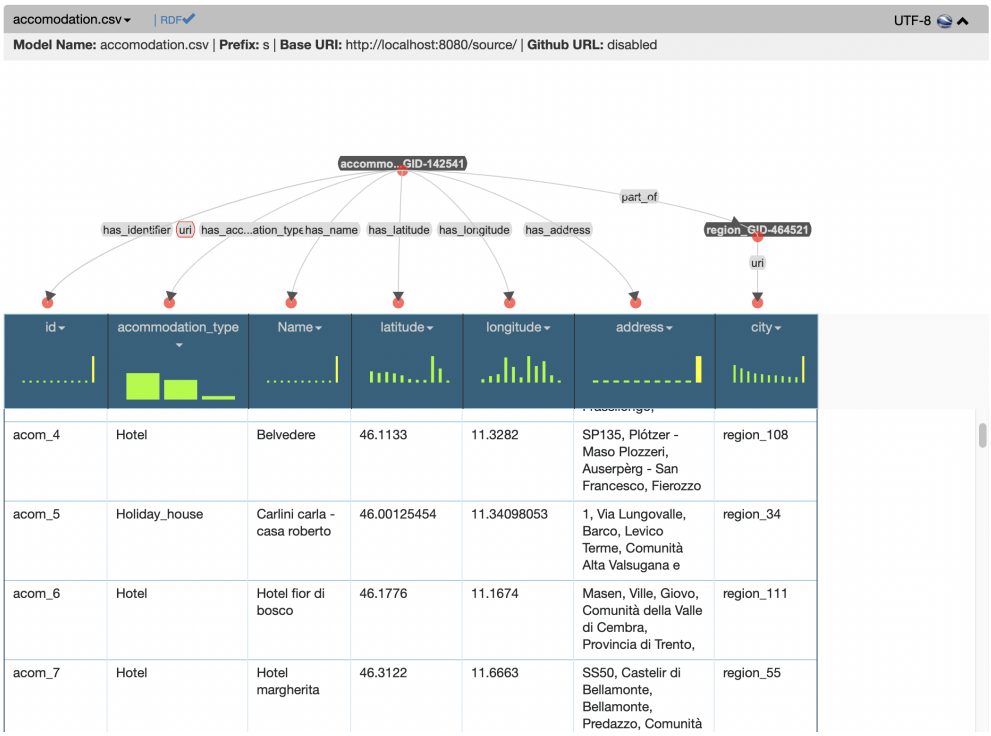



Figure 16: Example of data mapping

At consumer side, the Data Definition phase aims at producing the final KG, suitable to satisfy the requirements extracted from the user purpose (Competency Questions). Our final KG, will be both highly reusable and purpose-specific, due to the language alignment with the UKC and the generation of the teleontology adopted to structure its information, respectively.

- Entity matching: This step corrects the incoherences of real world entities represented through different properties, and properties values, within different datasets.
  - Schema layer: The information of what region was each place located in was stored in the object property part_of, and the data property has_city. Therefore, we eliminated the latter.
  - Data layer: We performed entity matching to modify the values of the object property part_of, that related 9 entity types with the entity_type region. This was necessary because the list of Regions obtained from ISTAT Turismo, did not coincide completely with those obtained by the geopy python library when including the coordinates of a place. Examples of corrections we have done include eliminating all kinds of commas, double spaces and accents. More complex corrections included finding the region of a village or city, as seen on the illustration (The region was originally mapped as Maiano, and had to be corrected to Cles).
- Entity identification: The entity types maintain their identifiers, as shown in the producer side. We make sure the identifiers are unique within the set of entity types.
- Entity mapping: We unite the different Knowledge Graphs created at the producer dimension, by copying the individual Knowledge Graphs into a single text file with extension

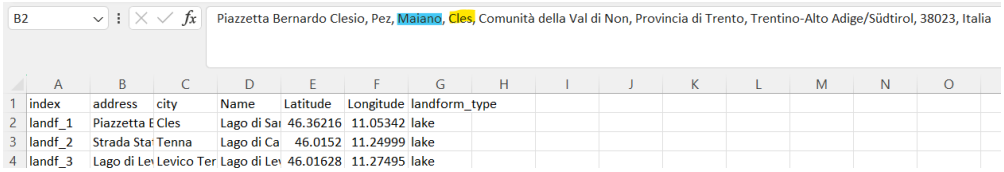



Figure 17: Example of data matching

RDF. This file is specific to our purpose.

In the last iTelos phase we have done the handling of the meaning heterogeneity, by entity matching and identification, The merging of the knowledge and data layer, and the generation of the final process output.

# 9 Evaluation

The last phase of the iTelos methodology consists on the evaluation of the quality of the final KG. The different evaluation activities includes the assessment of the KG's information statistics (number of etypes and properties, number of entities for each etype...), the calculation of different evaluation metrics for both the knowledge and data layer and finally the execution of different queries to test the suitability of the KG to answer the competency questions and therefore satisfy the project purpose.

## 9.1 Statistics final Knowledge Graph

| Variable | Concept | Value | Reference |
|---|---|---|---|
| $Cov_E$ | Number of etypes extracted from the CQs | 10 | Fig. 14 |
| $T_E$ | Number of etypes of the Teleontology | 10 | Number of class in Fig.11 |
| $Cov_P$ | Number of properties extracted from the CQs | 22 | |
| $T_P$ | Number of properties of the Teleontology | 25 | Number of object (4) and data properties (21) in Fig.11 |
| $RO_E$ | Number of etypes extracted from the ROs | 791 | |
| $RO_P$ | Number of properties extracted from the ROs | 0 | |

- We have the same number of etypes extracted from the CQs and in the Teleontology for two reasons. First, because we have come up with a lot of competency questions (20) that covered almost every topic related to our aim, tourism in trentino area, and therefore we have decided not to add more redundant entity types. And second, we haven't created hierarchies so there's no extra etypes.
- The number of properties extracted from the competency questions is slightly lower that the number of properties of the Teleology because we have extra information such as the gender, nationality or birthdate of the tourists.

- The Reference Ontology is the Trentino OSM LWontology.

## 9.2 Knowledge layer evaluation

The aim of evaluating the knowledge layer is to understand how much portion of knowledge (shaped as etypes and properties) is covered by the KG.

### 9.2.1 Primary goal - Purpose-based evaluation

To evaluate how much the final KG is able to satify the competency queries we calculate:

- the etype coverage ($Cov_E$) of the Teleontology $T$, given a set of CQs:

$$Cov_E(CQ_E) = \frac{|CQ_E \cap T_E|}{CQ_E} = \frac{10}{10} = 1$$

  So the teleontology covers 100% the entities extracted from the CQs.

- the property coverage ($Cov_P$) of the Teleontology $T$, given a set of CQs:

$$Cov_P(CQ_P) = \frac{|CQ_P \cap T_P|}{CQ_P} = \frac{25}{25} = 1$$

  So the teleontology covers 100% the properties (object and data) extracted from the CQs.

### 9.2.2 Secondary goal - Reusability evaluation

To evaluate how much reusable is the final KG we calculate:

- the etype cogerage ($Cov_E$) of the Teleontology $T$, given a set of reference ontologies (RO):

$$Cov_E(RO_E) = \frac{|RO_E \cap T_E|}{RO_E} = \frac{10}{791} = 0,013$$

- the property cogerage ($Cov_P$) of the Teleontology $T$, given a set of reference ontologies (RO):

$$Cov_P(RO_P) = \frac{|RO_P \cap T_P|}{RO_P} = \frac{25}{0} = undefined$$

## 9.3 Data Layer evaluation

The aim of evaluating the KG's data layer is to understand how connected the KG is. As the coverage calculated before, the connectivity of a KG can be evaluated over 2 dimensions:

- Entity connectivity: it evaluates the grade of connection between the different entities in the KG.
- Property connectivity: It evaluates the grades of connection between each single KG's entity and its properties values.

## 9.4 Query execution

In this sections we show the effectiveness and results of exploiting the KG by querying 5 out of the 20 CQs. To carry out this evaluation we have written queries in SparQL that have been executed in GraphDB.

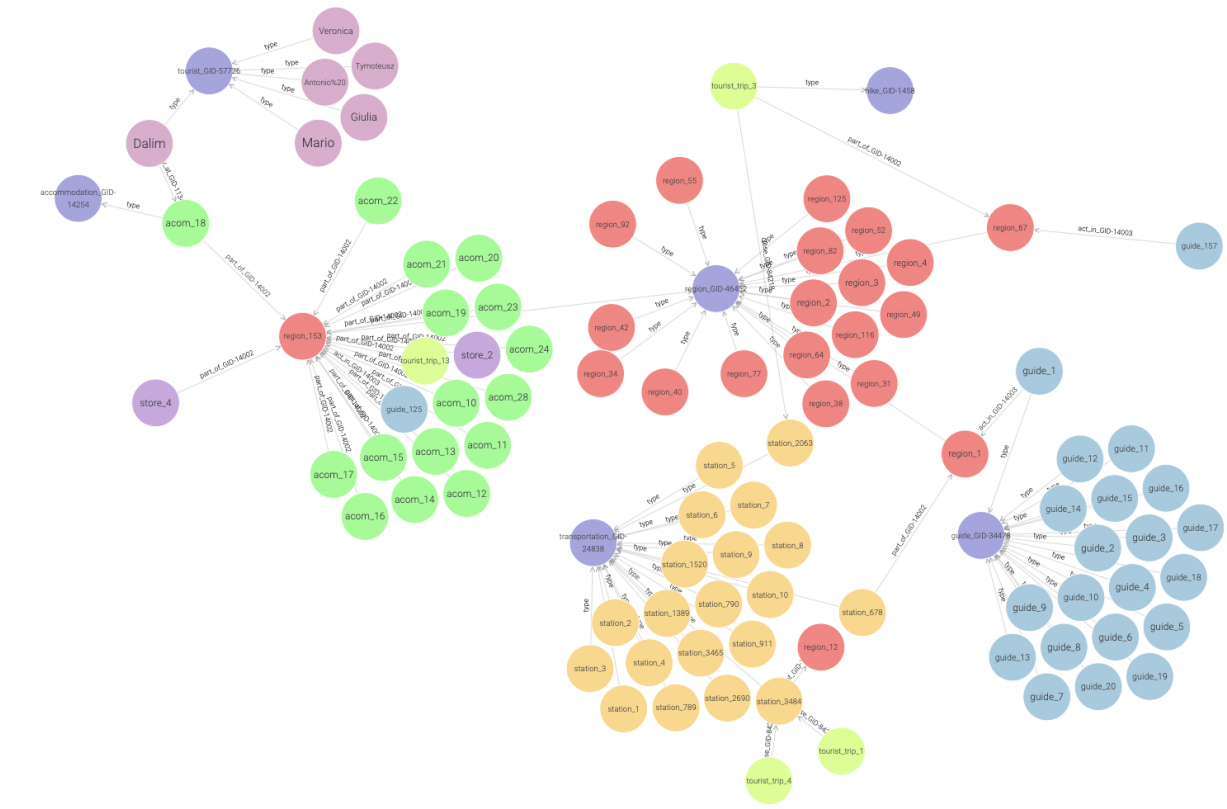



Figure 18: Final KG from GraphDB

- Query 1 (CQ 1): What are the top three challenging trails in Trentino?

  By most challenging trails we mean the trails that need a high level of expertise.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>

SELECT ?name ?origin ?expertise ?time ?height ?region
WHERE {
  ?hike rdf:type etype:hike_GID-1458 ;
        etype:has_name_GID-2 ?name ;
        etype:has_duration_GID-80582 ?time ;
        etype:has_height_GID-28272 ?height ;
        etype:has_expertise_GID-30843 ?expertise ;
        etype:close_GID-84218 ?stationid ;
        etype:part_of_GID-14002 ?regionid .
  FILTER(?expertise = "high") .
  ?regionid etype:has_name_GID-2 ?region .
```

```
  ?stationid etype:has_name_GID-2 ?origin .
}
LIMIT 3
```

Listing 1: Your SPARQL Query

| | name | origin | expertise | time | height | region |
|---|---|---|---|---|---|---|
| 1 | "Val di Ledro, Wasserfall Via del ponale loop from Riva del Garda" | "Bastione" | "high" | "03:25" | "470" | "Riva del Garda" |
| 2 | "Bastione Santa Barbara loop from Sant'Alessandro" | "Dimaro Ferrovia" | "high" | "04:50" | "780" | "Riva del Garda" |
| 3 | "Lago di Tenno â€“ TÃ¼rkisfarbener Tenno See loop from Gavazzo" | "Arco Piazza Italia" | "high" | "05:03" | "840" | "Tenno" |

Figure 19: Query 1

- Query 2 (CQ 10): Where can Antonio buy sleeping bags and a basic tent close to my accommodation?

  The idea is to get the longitude and latitude of the accommodation where Antonio is staying and calculating the distance between that geographical point and all the stores. Finally, select the one with the lowest distance.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX omgeo: <http://www.ontotext.com/owlim/geo#>

SELECT ?name ?address ?distance
WHERE
{
  ?tourist rdf:type etype:tourist_GID-57726 ;
           etype:has_name_GID-2 ?touristname ;
           etype:stay_at_GID-113409 ?acommodation .
  FILTER(?touristname = "Antonio ") .

  ?acommodation etype:has_latitude_GID-46263 ?acomm_latitude ;
                etype:has_longitude_GID-46270 ?acomm_longitude .

  ?store rdf:type etype:store_GID-23209 ;
         etype:has_name_GID-2 ?name ;
         etype:has_latitude_GID-46263 ?store_latitude ;
         etype:has_longitude_GID-46270 ?store_longitude ;
         etype:has_address_GID-45803 ?address .
  BIND(omgeo:distance(?acomm_latitude, ?acomm_longitude, ?store_latitude, ?
      store_longitude) as ?distance) .
}
ORDER BY ASC(?distance)
LIMIT 1
```

Listing 2: SPARQL Query for Tourist Antonio

| | name | address | distance |
|---|---|---|---|
| 1 | "Snow House" | "Snow House, 13, Via Crivelli, Fontanabotte, Zivignago, Pergine Valsugana, ComunitÃ  Alta Valsugana e Bersntol, Provincia di Trento, Trentino-Alto Adige/SÃ¼dtirol, 38057, Italia" | "3.938738901433638"^^xsd:float |

Figure 20: Query 2

- Query 3 (CQ 11): How many caves, mountains and lake can be visited in Baselga di Pine in a day trip?

  The idea is to group by the types of formation and count them for a specific region, in this case Baselga di Pine.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX omgeo: <http://www.ontotext.com/owlim/geo#>

SELECT ?formation_type (COUNT(?name) as ?count)
{
  ?formation rdf:type etype:formation_GID-49530 ;
             etype:has_formation_type_GID-14004 ?formation_type ;
             etype:has_name_GID-2 ?name ;
             etype:part_of_GID-14002 ?region .

  ?region etype:has_name_GID-2 ?name_region ;
          FILTER(?name_region = "Baselga di Pine").
}
GROUP BY (?formation_type)
```

Listing 3: SPARQL Query for Formation Types in Baselga di Pine

| | name | address | type |
|---|---|---|---|
| 1 | "Lago di Levico" | "Lago di Levico, LocalitÃ  al Lago, Levico Terme, ComunitÃ  Alta Valsugana e Bersntol, Provincia di Trento, Trentino-Alto Adige/SÃ¼dtirol, 38056, Italia" | "lake" |

Figure 21: Query 3

- Query 4 (CQ 14): How can I access the different skis stations in Trento?

  The idea is to calculate the distance of all the ski resorts and all the transportation stations and for each ski resort take the one that is the closest.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX omgeo: <http://www.ontotext.com/owlim/geo#>

SELECT ?ski_name ?distance ?station_name ?station_type
WHERE
{
    ?ski_resort rdf:type etype:ski_resort_GID-46553 ;
                etype:has_name_GID-2 ?ski_name ;
                etype:has_latitude_GID-46263 ?ski_latitude ;
                etype:has_longitude_GID-46270 ?ski_longitude .

    ?station rdf:type etype:transportation_GID-24838 ;
             etype:has_name_GID-2 ?station_name ;
             etype:has_latitude_GID-46263 ?station_latitude ;
             etype:has_longitude_GID-46270 ?station_longitude ;
             etype:has_transportation_type_GID-14005 ?station_type  .

```

```
    BIND(omgeo:distance(?ski_latitude, ?ski_longitude, ?station_latitude, ?
        station_longitude) as ?distance) .

    {
        SELECT ?ski_name (MIN(?distance) AS ?mindistance)
        WHERE
        {
            ?ski_resort rdf:type etype:ski_resort_GID-46553 ;
                        etype:has_name_GID-2 ?ski_name ;
                        etype:has_latitude_GID-46263 ?ski_latitude ;
                        etype:has_longitude_GID-46270 ?ski_longitude .

            ?station rdf:type etype:transportation_GID-24838 ;
                     etype:has_name_GID-2 ?station_name ;
                     etype:has_latitude_GID-46263 ?station_latitude ;
                     etype:has_longitude_GID-46270 ?station_longitude ;
                     etype:has_transportation_type_GID-14005 ?station_type  .

            BIND(omgeo:distance(?ski_latitude, ?ski_longitude, ?station_latitude, ?
                station_longitude) as ?distance) .
        }
        GROUP BY (?ski_name)
    }

    FILTER(?mindistance = ?distance) .
}
LIMIT 5
```

Listing 4: SPARQL Query for Ski Resorts and Closest Transportation Stations

| | ski_name | distance | station_name | station_type |
|---|---|---|---|---|
| 1 | "San Martino di Castrozza" | "0.08514792715984919"^^xsd:float | "S.Martino di C.-Im.Col Verde" | "bus" |
| 2 | "Madonna di Campiglio" | "0.2864411797943001"^^xsd:float | "M.Campiglio Des Alpes" | "bus" |
| 3 | "Passo San Pellegrino" | "0.5985344616477905"^^xsd:float | "P.So S.Pellegrino-Segg.Cost." | "bus" |
| 4 | "Belvedere/Col Rodella" | "1.8102107285667173"^^xsd:float | "Penia Alta" | "bus" |
| 5 | "Paganella/Andalo" | "3.2507951498166046"^^xsd:float | "Andalo-Maso Toscana" | "bus" |

Figure 22: Query 4

- Query 5 (CQ 20): Which natural attraction is closest to this train station Levico Terme?

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX omgeo: <http://www.ontotext.com/owlim/geo#>

SELECT ?name ?address ?type
WHERE
{
  ?landform rdf:type etype:formation_GID-49530 ;
            etype:has_name_GID-2 ?name ;
            etype:has_address_GID-45803 ?address ;
            etype:has_latitude_GID-46263 ?landform_latitude ;
```

```
            etype:has_longitude_GID-46270 ?landform_longitude ;
            etype:has_formation_type_GID-14004 ?type .

  ?station rdf:type etype:transportation_GID-24838 ;
           etype:has_latitude_GID-46263 ?station_latitude ;
           etype:has_longitude_GID-46270 ?station_longitude ;
           etype:has_name_GID-2 ?station_name.
  FILTER(?station_name = "Levico Terme") .

  BIND(omgeo:distance(?landform_latitude, ?landform_longitude, ?station_latitude, ?
      station_longitude) as ?distance).
}
ORDER BY ASC(?distance)
LIMIT 1
```

Listing 5: SPARQL Query for Nearest Landform to Levico Terme

| | formation_type | count |
|---|---|---|
| 1 | "cave" | "2"^^xsd:integer |
| 2 | "mountain" | "45"^^xsd:integer |

Figure 23: Query 5

# 10 Metadata Definition

- Language resources metadata description

| | |
|---|---|
| **DatLicense** | Apache-2.0 license |
| **DatURL** | `https://github.com/mbueno-g/KGE---Tourism-in-Trento/blob/main/Phase%203%20-%20Language%20Definition/Language%20spreadsheet.csv` |
| **DatKeyword** | tourism, trentino |
| **DatPublisher** | Lucia Trillo Carreras, Marina Bueno Gracia |
| **DatCreator** | Lucia Trillo Carreras, Marina Bueno Garcia |
| **DatOwner** | Lucia Trillo Carreras, Marina Bueno Garcia |
| **DatLanguage** | English |
| **DatSize** | 32 concepts |
| **DatName** | Language spreadsheet.csv |
| **DatPublication** | 21/01/2024 |
| **DatDescription** | Description of all the language used in the Natural Trentino Knowledge graph |
| **DatVersion** | 2 |
| **DatDomain** | tourism |
| **DatFileFormat** | csv |

- Knowledge resources metadata description

| DatLicense | DatCreator | DatOwner | DatLanguage | DatName | DatPublicationTimestamp | 0 | DatDomain |
|---|---|---|---|---|---|---|---|
| CC Attribution-ShareAlike License | Innsbruck university | schema.org | english | accomodation | 01/09/2024 | An accommodation is a place that can accommodate human beings | tourism |
| CC Attribution-ShareAlike License | Tourism Structured Web Data Community Group | schema.org | english | tourist_trip | 01/09/2024 | A created itinerary of visits to one or more places of interest | tourism |
| CC Attribution-ShareAlike License | schema.org | schema.org | english | station | 01/09/2024 | A public structure, such as a town hall or concert hall. | transport |
| CC Attribution-ShareAlike License | schema.org | schema.org | english | Store | 01/09/2024 | A retail good store. | tourism |
| CC Attribution-ShareAlike License | IPTC rNews properties | schema.org | english | Person_tourismofficial | 01/09/2024 | A person (alive, dead, undead, or fictional). | tourism |
| CC Attribution-ShareAlike License | schema.org | schema.org | english | Landform general | 01/09/2024 | A landform or physical feature | tourism |
| CC Attribution-ShareAlike License | schema.org | schema.org | english | GasStation | 01/09/2024 | A gas station. | tourism |
| CC Attribution-ShareAlike License | schema.org | schema.org | english | SkiResort | 01/09/2024 | A ski resort. | tourism |
| CC Attribution-ShareAlike License | schema.org | schema.org | english | City(Region) | 01/09/2024 | The region in which the locality is, and which is in the country. | territory |

- Data resources metadata description

| DatLicense | DatKeyword | DatOwner | DatLanguage | DatSize | DatName | DatPublication | DatVersion | DatDomain | DatFileFormat |
|---|---|---|---|---|---|---|---|---|---|
| Apache-2.0 license | bus,transport,trentino, trento | Roberta Peracchio,Hamid Omidi | english | 231 KB | bus.csv | 21/12/2022 | 1 | transport | csv |
| Apache-2.0 license | train,transport,trentinto, trento | Roberta Peracchio,Hamid Omidi | english | 11.8 KB | train_station_2.csv | 21/12/2022 | 2 | transport | csv |
| Apache-2.0 license | gas station, cars, transport, trentino, trento | Roberta Peracchio,Hamid Omidi | english | 2.22 KB | gas_station.csv | 21/12/2022 | 1 | transport | csv |
| Apache-2.0 license | acommodation, holiday, hotel,trentino, trento | Roberta Peracchio,Hamid Omidi | english | 29.5 KB | hotel_2.csv | 21/12/2022 | 2 | tourism | csv |
| Apache-2.0 license | acommodation, holiday, campsite,trentino, trento | Roberta Peracchio,Hamid Omidi | english | 4.22 KB | campsite_2.csv | 21/12/2022 | 2 | tourism | csv |
| Apache-2.0 license | acommodation, holiday, holiday house, trentino, trento | Roberta Peracchio,Hamid Omidi | english | 57.9 KB | holiday_house2.csv | 21/12/2022 | 2 | tourism | csv |
| Apache-2.0 license | skiing, sports, holiday, trentino, trento | Roberta Peracchio,Hamid Omidi | english | 2.53 KB | skiing_infrastructure2.csv | 21/12/2022 | 2 | tourism | csv |
| Open Data Commons Attribution License | trento, trentino, tourism, shops, outdoors, nature | Knowdive Research group | english | 2 KB | pint_outdoor_shop.txt | 06/09/2023 | 1 | tourism | txt |
| Open Data Commons Attribution License | trento, trentino, natural, mountain | Knowdive Research group | english | 240 KB | point_peak.txt | 06/09/2023 | 1 | tourism | txt |
| Open Data Commons Attribution License | trento, trentino, natural, cave | Knowdive Research group | english | 32 KB | point_cave_entrance.txt | 06/09/2023 | 1 | tourism | txt |
| Open Data Commons Attribution License | trento, trentino, natural, lake | Trentino.com | english | 3.55 KB | lakes.csv | 15/12/2023 | 1 | tourism | csv |
| CC Attribution 4.0 International (CC BY 4.0) | tourism,trento, trentino, regions, territory | Provincia autonoma di Trento | italian | 19 KB | dataset-data-61135.csv | 12/07/2022 | NA | tourism | csv |
| Apache-2.0 license | tourism,trento,hikes, nature, trentino | Komoot | english | 2.02 KB | tourist_trip.csv | 15/12/2023 | 5 | tourism | csv |
| Apache-2.0 license | tourism,trento, trentino, regions, territory, custom, personas | Marina Bueno Garcia | english | 0,35 KB | person_tourist.csv | 07/12/2023 | 2 | tourism | csv |

# 11 Open Issues

In summary, this research has successfully achieved its objective of constructing a knowledge graph that encapsulates the diverse information related to natural tourism in Trentino. The knowledge graph serves as a valuable tool for gaining insights into various aspects of visitors' options. The practical outcomes of this project signify a promising starting point, with ample opportunities for further enhancement and refinement within the context of Trentino's natural attractions. Looking forward, numerous potential directions for future work have been identified. These include the integration of additional hikes, modeling points of interest within the hikes, enriching the ontology with more references, and the incorporation of more time-specific data into the knowledge graph, possibly from diverse sources (specifying the best and worst times a year to make a certain visit). beyond mere points of interest, could provide deeper insights into their preferences and behaviors.Although the project has yielded practical outcomes, it represents just the initial phase of a broader exploration into understanding and enhancing the natural tourism experience using knowledge graphs. The extensive potential for future work promises exciting developments in advancing research within the realm of natural tourism in Trentino.