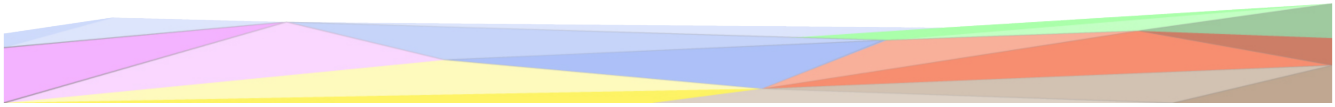
UNIVERSITY
OF TRENTO - Italy

Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

Trentino Tourist Facilities

Document Data:	Reference Persons:
December 21, 2022	Roberta Peracchio, Hamid Omidi



Index:

Revision History:

Revision	Date	Author	Description of Changes
0.1	09.11.2022	Fausto Giunchiglia	Document created

1 Introduction

This document provides a deep insight into the process we followed to create and complete this project.
The iTelos methodology was followed in all the steps, which are described section by section as follows:

- Section 2: Statement of the purpose
- Section 3: The sources used in this project
- Section 4: Description of the scenarios and the personas, including the Competency Questions
- Section 5: Inception phase
- Section 6: Informal modeling
- Section 7: Formal modeling
- Section 8: Creation of the KGC
- Section 9 and 10: Outcome exploitation and final comments

For all the material present in this report, please refer to this web page and the linked github repository inside of it: `https://rorosonoio.github.io/KGE---Trentino-tourist-facilities/`

2 Purpose and Domain of Interest (DoI)

2.1 Project Purpose

The aim of this project is to create a KG that can provide information about Trentino's tourist-related facilities. In another word, the final KG can be used to provide a general-purpose service helping tourists to find information about the various tourism-related hospitality facilities in the region of Trentino. In a nutshell, the purpose of this project is described as:

"A service which helps the users to know about different tourist facilities in Trentino."

2.2 Project Domain of Interest

The Domain of interest of this project is the *Trentino* region and the different types of *tourists* that want to visit Trentino (domestic and inbound) in *2022*. The tourists can be considered as a wide range of people from foreign tourists to people who live in different cities of this region, who want to use the tourist facilities of Trentino. From another point of view, tourists can contain different types of tourism with different goals, such as:

- Business Tourism: Business tourism is the provision of facilities and services to the millions of delegates who annually attend meetings, congresses, exhibitions, business events, incentive travel and corporate hospitality;
- Cultural Tourism: Movements of persons for essentially cultural motivations, such as study tours, performing arts, cultural tours, festivals and other cultural events, visits to sites and monuments and trips to study nature, folklore or art;
- Backpacking/Youth Tourism: The activity of traveling while carrying your clothes and other essential things in a backpack, usually not spending very much money and staying in places that are not expensive.

3 Data Sources

This section reports and describes the input resources used in this project:

3.1 Knowledge sources

To cover this project's needs in the knowledge resource part, we referred to schema.org, which already has a pre-defined wide range of resource schemas. Schema.org is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet. This is one of the most flexible which can be used in different concepts.

Therefore, we tried to map our vocabulary domain to schema.org as much as possible to keep it more flexible and more understandable for users. Moreover, this knowledge resource is one of the most flexible and can be used in different concepts. Therefore, we tried to map our vocabulary domain to schema.org as much as possible to keep it more flexible and more understandable for future users.

Furthermore, we use schema.org for developing the ontology part. In other words, we try to use the predefined and well-known structure for the ontology part to optimally map the concepts within the ontology to the real world. In this way we can improve the generalizability and usability of the knowledge graph.

3.2 Data Resources

The data resources used in this project have been extracted from different sources. The most important resource for this project is the ”Open Data Trentino” platform which contains a wide range of information and data sets related to the Trentino region services.

In the following the data sets and the resources that are used in this project have been explained:

- The data regarding accommodations is included in ”Esercizi Alberghieri.xml” (source: https://dati.trentino.it/dataset/esercizi-alberghieri) and ”Esercizi Extra Alberghieri.xml” (source: https://dati.trentino.it/dataset/esercizi-extra-alberghieri), which contain respectively the structures who give a hotel treatment and the structures who do not (that is, structures like campsites or holiday house, which do not provide assistance or services).
- The other facilities were extracted from both ”Punti di interesse del Trentino.json” (source: https://dati.trentino.it/dataset/punti-di-interesse-del-trentino), which included data about several different structures and places in the area, and ”Botteghe storiche del Trentino.csv” (source: https://dati.trentino.it/dataset/botteghe-storiche-del-trentino), which include local businesses which have been active for at least 50 years.
- For accessing the bus facilities information, we use the open data shared by Trentino Trasporti, the local public transport company, on https://www.trentinotrasporti.it to collect the dataset related to the locations of the bus stop. Therefore with this data set, we can expand the facilities to both urban and extra-urban bus stops located in the Trentino region.

- The skiing areas were gathered from the website https://www.skiresort.it/comprensori-sciistici/trentino/, from which only the ones with information on the current season were taken.
- The train stations were extracted from Open Street Maps.

The original datasets are accessible online from the below link: `https://github.com/rorosonoio/KGE---Trentino-tourist-facilities/tree/main/Datasets/Informal%20Modeling`.

In order to be able to work nicely with these resources, data manipulation and cleaning was performed in order to be useful for our purpose. We have deleted inconsistent data, columns which were not going to be used in the future and null columns.

4 Purpose Formalization

In this section The scenarios, persona related to the scenarios, Competency Questions which will be extracted from scenarios , and Entities identified have been shown.

4.1 Scenarios

1. A family decided to visit the Trentino for some days during the summer. This family contains two parents and two children. **Ludovico Trevvi** is the father of this family and wants to rent an appropriate holiday house for a week in Trento for his family. They are also a green lifestyle family and want to use green mobility as much as possible. So they have the plan to go to some natural attraction place in the Trentino region by bus. They also want to use the bicycle when they visit the city. The parents are also thinking about finding some entertainment for their children during this trip so they want to find the closest amusement park to the holiday houses they will rent. Moreover, They want to enjoy some Trentino cuisine during their time in Trentino.
2. There are some perfect skiing areas in the Trentino, that will host several people during the winter. Young people from different parts of European countries come to use this golden time of skiing in nature. **Jannis Zeus** loves skiing and decided to come to Trento to enjoy it. He also needs a hotel to stay close to the slopes, but he does not want to spend a lot of money, as he will spend most of the time outside. Other option for him might be choosing a restaurant to eat in city center and he will try to take bus to the that place too keep expenses low. In the evenings he goes to some bars to drink and enjoy their vacations in dolomite's nature.
3. Trentino has one of the main roads from Italy to other European countries. **Ben Muller** decided to make a road trip with his friends from Berlin to Rome. They will use Ben's car, but they will share the fuel price. They also plan to stay in Trentino for a maximum of 4 days. They do not want to spend a lot of money, so they prefer to stay in campsites around the cities. They Also prefer find gas stations close to the campsites which makes the trip more convenient for them. Also, they plan to buy souvenirs from each city they will stop in. During stopping days they will spend time in bars and meet new people.
4. The famous Economics Festival takes place in Trento at the beginning of Semptember. The businessman **Germano Rossi** will attend this event, so he will need a hotel to stay in Trentino. He likes the comfort of private transport, so he will use taxis to move around the city. He wants to stay one extra day, in order to visit some museums and also meet some business men for a work dinner, and at the last night he wants to meet some old friends from Trento and drink with them, but he need to choose a closest bar as he has a flight back to his city.
5. **Hamid Amiri** is an exchange student at Verona University and he would like to have an adventurous holiday in Trentino. He is planning to move around campsites and rifugios while hiking in the mountains and appreciates the natural beauties that this region has to

offer. As he usually uses the train for his trips he wants to fine some natural attractions which are close to the train stations in the Trentino region. As he has a student discount on his trip by bus, he will use the bus to check the different parts of the city.

6. A group of friends would like to visit Trentino on holiday. Since they are six people, **Gianluca Taverna** has decided to rent a holiday house. Since they do not have a car, they would like to find an accommodation close to a train station, and move with buses around, too. Gianluca and his friends want to collect souvenirs and buy local products, as well as eat in local restaurants and try nice cafes. They also want to spend some fun time in Trentino and have some drinks together.

4.2 Personas

1. **Ludovico** is 55 year old Italian man who works for a software developing company. He always use his job vacations during summer to visit a new place in Italy with his family. He believes in green lifestyle and he is into sports.
2. **Jannis** is a 22 year old Greek man who studies Data science in Germany. He loves skiing and during winter he allocates some days to go skiing in famous locations.
3. **Ben** is a 26 year old German guy who has had several experiences with road trips. He loves drinking and spending lots of time with his friends.
4. **Germano** is a 60 year old Italian businessman, who enjoys visiting museums and is very dedicated to his job. Whenever he travels he enjoys a comfortable stay.
5. **Hamid** is a 23 year old Iranian exchange student. He loves being in nature and, whenever he has free time, he enjoys some adventure.
6. **Gianluca** is a 20 year old Italian man who likes travelling with his friends and experience the local culture of the places he visits. He collects souvenirs.

4.3 Competency Questions (CQs)

Person	No.	Question
Ludovico	1.1	How can he find a holiday house in the Trentino region for his family?
Ludovico	1.2	Where is the closest bus stop to their holiday house?
Ludovico	1.3	Where are the bicycle rental stops in Trentino region?
Ludovico	1.4	Where can they find an amusement park for children in the city, which is close to their holiday house?
Ludovico	1.5	Which natural attraction they should visit?
Ludovico	1.6	Which bus stop is near the natural attraction they want to visit?
Ludovico	1.7	Which restaurant they can choose to go?
Ludovico	1.8	Which bus stop is close to the restaurant they chose to eat?
Ludovico	1.9	Which bicycle stop is near to the restaurant they chose to eat?
Jannis	2.1	Which skiing infrastructures are available in the Trentino Region?
Jannis	2.2	Which is the closest bus stop to the skiing area?
Jannis	2.3	Where can he find a hotel near to the skiing sites?
Jannis	2.4	Which bar is near the skiing sites that he chooses?
Jannis	2.5	Which ski area has the longest extension?
Jannis	2.6	Which natural attraction is the closest to the skiing site that he choose?
Jannis	2.7	Which restaurant he will go for eating which is close to the skiing sites?
Ben	3.1	Where are the gas stations during their journey in Trentino?
Ben	3.2	Where can he find a gas station close to the campsite?
Ben	3.3	Which campsite can he stay in?
Ben	3.4	Where can they buy some souvenirs in Trentino?
Ben	3.5	Where is the closest souvenir shop to the campsite they are in?

Person	No.	Question
Ben	3.6	Which bar is the closest one to the place they are stay in (campsite)?
Ben	3.7	Where can he eat during his road trip? (which restaurant?)
Germano	4.1	Which hotel should Germano book for his trip?
Germano	4.2	Which museums are present in the city?
Germano	4.3	Which taxi agency he should call for travelling in the city?
Germano	4.4	Where can he find a restaurant for his meeting?
Germano	4.5	Which bar he can choose to drink with his fiends?
Germano	4.6	Which bar is the closest one to the Germano's hotel?
Hamid	5.1	Which natural attraction is close to train station?
Hamid	5.2	Which natural attractions can he visit?
Hamid	5.3	Which campsites are close to the train station?
Hamid	5.4	Which bus stop is close to the campsite?
Hamid	5.5	Which bus stop is close to the natural attraction?
Hamid	5.6	Which bus stop close to the train station he will get out for make an trip with both bus and train?
Gianluca	6.1	Which holiday houses they can choose?
Gianluca	6.2	Which restaurants they can go?
Gianluca	6.3	Which bar they can go?
Gianluca	6.4	Which souvenir shops are in the area?
Gianluca	6.5	Which museums can he and his friends visit?
Gianluca	6.6	Where is the bus stop closer to his holiday house?
Gianluca	6.7	Which bike sharing station is closest to their holiday house?
Gianluca	6.8	Which train station is close to their holiday house?

4.4 Entities identified

- **1.1**: Core entities: Holiday house
- **1.2**: Common entities: Bus stop
- **1.3**: Common entities: Bike rental
- **1.4**: Core entities: Amusement park
- **1.5**: Contextual entities: Natural attraction
- **1.6**: Common entities: Bus stop
- **1.7**: Common entities: Restaurant
- **1.8**: Common entities: Bus stop
- **1.9**: Common entities: Bike rental
- **2.1**: Contextual entities: Skiing infrastructure
- **2.2**: Common entities: Bus stop
- **2.3**: Common entities: Hotel
- **2.4**: Common entities: Bars and pubs
- **2.5**: Contextual entities: Skiing infrastructure
- **2.6**: Contextual entities: Natural attraction
- **2.4**: Common entities: Restaurant
- **3.1**: Common entities: Gas station
- **3.2**: Common entities: Gas station
- **3.3**: Core entities: Campsite
- **3.4**: Contextual entities: Souvenir shop
- **3.5**: Contextual entities: Souvenir shop
- **3.6**: Common entities: Bars and pubs
- **3.7**: Common entities: Restaurant
- **4.1**: Common entities: Hotel
- **4.2**: Core entities: Museum
- **4.3**: Common entities: Taxi company
- **4.4**: Common entities: Restaurant
- **4.5**: Common entities: Bars and pubs

- **4.6**: Common entities: Bars and pubs
- **5.1**: Contextual entities: Natural attractions
- **5.2**: Contextual entities: Natural attractions
- **5.3**: Core entities: Campsite
- **5.4**: Common entities: Bus stop
- **5.5**: Common entities: Bus stop
- **6.1**: Core entities: Holiday houses
- **6.2**: Common entities: Restaurants
- **6.3**: Common entities: Bars and pubs
- **6.4**: Contextual entities: Souvenir shops
- **6.5**: Core entities: Museum
- **6.6**: Common entities: Bus stop
- **6.7**: Common entities: Bike sharing
- **6.7**: Common entities: Train station

5 Inception

5.1 Resource collection

The competency questions cover mainly three areas: accommodation, food and entertainment (or activities). Some of the links of the datasets (hotels, historic shops, and hospitality facilities) were provided for the project by the project owner (gathered from the link within the project description). These datasets, that can be accessed on open Data Trentino portal, will cover a small part of the project so we need more data sets to satisfy our goals in CQ parts.

For another part of the datasets, we gather the points of interest of Trentino (Punti di interesse del Trentino.json) and extracted the facilities which are matched our needs.

Also, bus stops are another part of the datasets; we downloaded the data from the Trentino Trasporti website, which offers two files. These files contain both urban and extra-urban bus stops for this project.

Name of the file	Source	Coverage
`Esercizi Alberghieri.xml`	Open Data Trentino	Accommodation
`Esercizi Extra Alberghieri.xml`	Open Data Trentino	accommodation
`Punti di interesse del Trentino.json`	Open Data Trentino	All parts (Not bus stops)
`Botteghe storiche del Trentino.csv`	Open Data Trentino	Historic shops
`google transit urbano tte.zip`	Trentino Trensporti	Bus stops(urban)
`Scraping from a website`	skiresort.it	skiing infrastructure

5.2 Resource filtering and Data cleaning

The main obstacle was to distinguish between facilities and points of interest, since they often may overlap. Moreover, all datasets were in Italian, so translation was also performed on the column names and the categories.
The main filtering was done by choosing the rows with the most consistent information: facility name, location (latitude, longitude, address and city), category and contact information (phone number and/or email).
The most inconsistent row was the contact information, as it was not made available by some infrastructures. Moreover, some entities such as natural attractions (lakes or hiking routes) do not have a owner, and as a consequence no contacts.
The ”category” column was the one that has been worked on the most. The column itself was available in every dataset, but it showed numerous inconsistencies. First of all, from all the unique values of the column, new bigger categories were created, from which the most relevant for our purpose were chosen. The following are the more sparse categories present in all datasets, to be later merged in their own datasets to follow the pre-defined schema:

- hotel, restaurant, skiing infrastructure, ice rink, amusement park, bar, healthcare services, shop, grocery shop, typical products, holiday house, internet point, souvenir shop, gas station, natural attraction, wellness center, farmhouse, ice cream and pastries, bed and breakfast, museum, pub, skiing school, bike rent, campsite, car park, residence, cable car, train station, taxi, ski rental, rifugio, archaeological site, vehicle rental, travel agency, sport school.

Moreover, the bus stop data did not have the name of the city which is considered important information for the other datasets. However, we considered that latitude and longitude were enough to perform further functions later in the project, so we did not apply any more changes.

5.3 Resource knowledge definition

As we mentioned earlier in the knowledge sources section, schema.org has been used in this project as it is enough flexible to cover the vocabulary of this project. By navigating through schema.org we can find a similar schema to the concepts that we are going to use in this project. For satisfying the Knowledge graph we tried to use the knowledge in the specific domain and schema partially.

On one hand, we can search for a property of an entity that we want to satisfy the project, and at the same time, we can be sure that using this well-known schema can help us to keep our knowledge graph more stable based on the things that were designed before.

For example, we try to use some properties and their type of them given by schema.org. It helps us to be more precise. For example, by searching for Bars and Pubs we recognized that the more flexible model would be the combination of both data sets as in the schema.org it represented like BarOrPub.

Specifically, we use some of the schemas which are the following:

For *restaurants*, we use the Restaurant schema in the mentioned resource which is accessible on Thing > Organization > LocalBusiness > FoodEstablishment > Restaurant. For *bars* and *pubs* (as it is mentioned) we use the BarOrPub on Thing > Organization > LocalBusiness > FoodEstablishment > BarOrPub. For others we tried to being inspired by:

- Thing > Organization > LocalBusiness > LodgingBusiness > Hotel
- Thing > Place > CivicStructure > Campground
- Thing > Place > LocalBusiness > LodgingBusiness > Resort > SkiResort
- Thing > Organization > LocalBusiness (Touristc shops)
- Thing > Place > CivicStructure > BusStop
- Thing > Place > LocalBusiness > AutomotiveBusiness > GasStation
- Thing > Place > CivicStructure > TrainStation
- Thing > Intangible > Service > Taxi
- Thing > Place > LocalBusiness > EntertainmentBusiness > AmusementPark

5.4 Resource formatting

All the data sets format converted to the CSV in the appropriate way of using for this project. Moreover, all the extracted categories from POI (points of interest) have been split into different data sets.

The datasets have been cleaned and optimized for the coverage of the purpose (CQs) of the KG. After collecting the data from the resources the ontology of the entity associated with the data set has been developed. In this step, we used the protege to make the schema of the datasets.

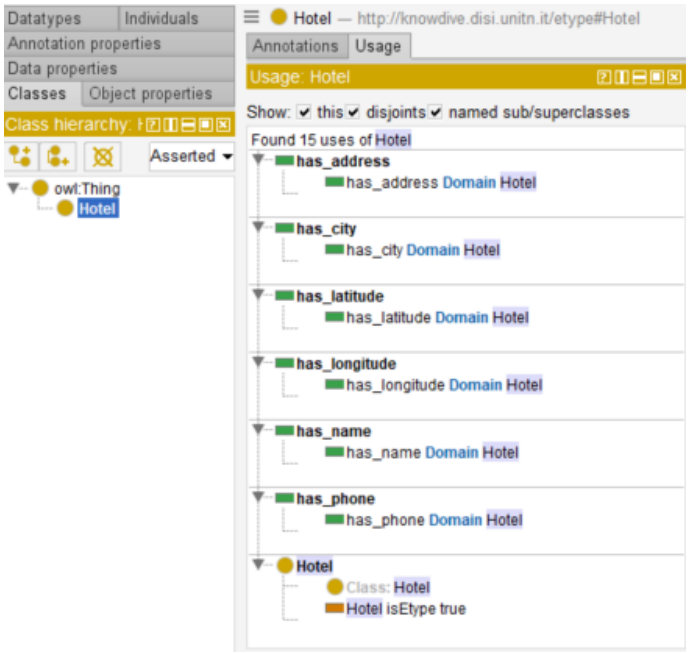
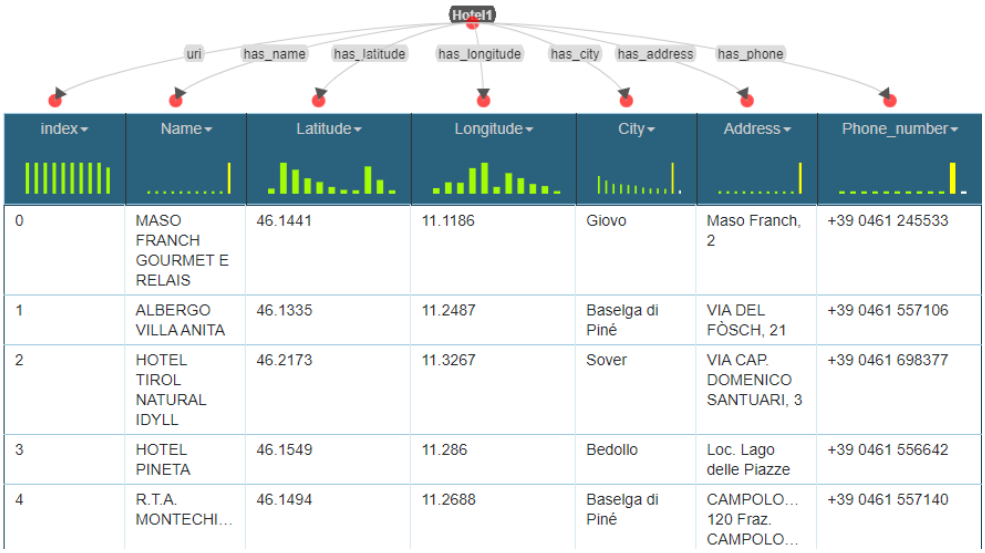


Figure 1: Protègè
An example of making an ontology for a dataset.

After that, the schema made by protègè was mapped to the dataset with karma. An example can be seen in the figure below.



index	Name	Latitude	Longitude	City	Address	Phone_number
0	MASO FRANCH GOURMET E RELAIS	46.1441	11.1186	Giovo	Maso Franch, 2	+39 0461 245533
1	ALBERGO VILLA ANITA	46.1335	11.2487	Baselga di Piné	VIA DEL FÒSCH, 21	+39 0461 557106
2	HOTEL TIROL NATURAL IDYLL	46.2173	11.3267	Sover	VIA CAP. DOMENICO SANTUARI, 3	+39 0461 698377
3	HOTEL PINETA	46.1549	11.286	Bedollo	Loc. Lago delle Piazze	+39 0461 556642
4	R.T.A. MONTECHI…	46.1494	11.2688	Baselga di Piné	CAMPOLO… 120 Fraz. CAMPOLO…	+39 0461 557140

Figure 2: Karma
Using Karma to assign the data schema to the Hotel dataset.

6 Informal Modeling

6.1 ER model description

The graph was built keeping in mind the spatial context, that is the Trentino region, while the actions are performed from the tourist. In fact, except from the link between Everything and Trentino region, all object relations depart from the perspective of the tourists.
In this graph, the following entities are present:

- Common entities: Trentino region, train station, restaurant, hotel, holiday house, bar and pub, gas station, bicycle station, bus stop.
- Core entities: museum, campsite
- Contextual entities: local products shop, souvenir shop, tourist, skiing infrastructure

6.1.1 Common

Trentino region
Describes the spatial context of the project.
Relations:

- SpatialPartOf related to Everything
- some entities are LocatedIn this region

Attributes:

- latitude (float)
- longitude (float)
- startTime (dateTime)
- endTime (dateTime)

Train station
A train station present in the region of Trentino.
Relations:

- Near an accommodation
- LocatedIn the region
- a tourist makes the action of BookTrip

Attributes:

- latitude (float)
- longitude (float)

- address (string)
- city (string)
- name (string)
- phone (string)

Taxi Company
A taxi company operating in the region of Trentino.
Relations:

- The tourist makes the action of ReserveTaxi
- LocatedIn the region

Attributes:

- latitude (float)
- longitude (float)
- address (string)
- city (string)
- name (string)
- phone (string)

Hotel
A hotel operating in the region of Trentino.
Relations:

- The tourist makes the action of BookPlace
- LocatedIn the region
- is Near another entity

Attributes:

- latitude (float)
- longitude (float)
- address (string)
- city (string)
- name (string)

Holiday house
A holiday house operating in the region of Trentino.
Relations:

- The tourist makes the action of BookPlace
- LocatedIn the region
- is Near another entity

Attributes:

- latitude (float)
- longitude (float)
- address (string)
- city (string)
- name (string)
- phone (string)

Restaurant
A restaurant operating in the region of Trentino.
Relations:

- The tourist makes the action of EatAt
- LocatedIn the region
- is Near another entity

Attributes:

- latitude (float)
- longitude (float)
- address (string)
- city (string)
- name (string)
- phone (string)

Bar or pub
A bar or a pub operating in the region of Trentino.
Relations:

- is Near another entity

Attributes:

- latitude (float)
- longitude (float)

- address (string)
- city (string)
- name (string)
- phone (string)

Bicycle station
A bike rent station present in the region of Trentino.
Relations:

- is Near another entity

Attributes:

- latitude (float)
- longitude (float)
- address (string)
- city (string)
- name (string)

Bus stop
A bus stop present in the region of Trentino.
Relations:

- is Near another entity

Attributes:

- latitude (float)
- longitude (float)
- address (string)
- city (string)
- name (string)

Gas station
A souvenir shop operating in the region of Trentino.
Relations:

- is LocatedIn the region
- is Near another entity

Attributes:

- latitude (float)

- longitude (float)
- address (string)
- city (string)
- name (string)
- phone (string)

6.1.2 Core

Amusement park
An amusement park present in the region of Trentino.
Relations:

- is Near another entity

Attributes:

- latitude (float)
- longitude (float)
- address (string)
- city (string)
- name (string)
- phone (string)

Museum
A museum operating in the region of Trentino.
Relations:

- The tourist makes the action of Visit

Attributes:

- latitude (float)
- longitude (float)
- address (string)
- city (string)
- name (string)
- phone (string)

Campsite
A campsite present in the region of Trentino.
Relations:

- The tourist makes the action of BookPlace
- is Near another entity

Attributes:

- latitude (float)
- longitude (float)
- address (string)
- city (string)
- name (string)
- phone (string)

Tourist
A person visiting the region of Trentino.
Relations:

- The tourist makes the action of EatAt
- ReserveTaxi
- BookPlace
- ShopFrom
- BookTrip
- Visit

Attributes:

- name (string)
- surname (string)
- age (int)

6.1.3 Contextual

Skiing infrastructure
Skiing infrastructures operating in the region of Trentino.
Relations:

- is Near another entity

Attributes:

- latitude (float)
- longitude (float)

- address (string)
- city (string)
- name (string)
- length (float)

Natural attraction
A natural attraction present in the region of Trentino.
Relations:

- LocatedIn the region
- is Near another entity

Attributes:

- latitude (float)
- longitude (float)
- city (string)
- name (string)

Souvenir shop
A souvenir shop operating in the region of Trentino.
Relations:

- The tourist makes the action of ShopAt
- is Near another entity

Attributes:

- latitude (float)
- longitude (float)
- address (string)
- city (string)
- name (string)
- phone (string)

Local products shop
A local products shop operating in the region of Trentino.
Relations:

- The tourist makes the action of ShopAt
- is Near another entity

Attributes:

- latitude (float)
- longitude (float)
- address (string)
- city (string)
- name (string)
- phone (string)

6.2 Teleology building

The teleology of this project consists of different classes, each one with different attributes. These classes are connected to each other based on the CQs of the project, which were extracted from the scenarios. Considering the spatial context of the KG (time and space), the Trentino region was added to the graph, which shows the space and time. Actually, time has been defined in the attributes of the Trentino region.
As can be seen, the chart's starting point is Everything, that is connected to the Trentino region. In fact, the Trentino region entity is our contextual everything. This means that in our context (tourism) we have this comprehensive entity that can cover all other entities that we are going to use. By doing that, technically we are trying to make our KG world more precise and limited. Moreover, as we can see in the previous graph, we have some connections that can fall into different categories. We have located which cover the space context of the entities.
For example, the Hotel is located in the Trentino region and it means we are considering the Hotels that are located in the Trentino. We say Hotel(s) because we also mention the cardinality of this connection.
There are also some green lines that are drawn just to make the graph more readable. And all of them are used to connect the tourist entity to another entity based on the action that a tourist can take. As with all the previous object properties that we had, these relations are extracted from CQs.

These types of connections show some actions that one class can take over another one. For instance, a Tourist can book a place or can visit a museum. The important point here is that this relation (which are the object properties) is limited to the DoI and of course the CQs of the project. Considering this statement is important, because lots of different relations can be defined between these different classes in other concepts.
The colors also play an important role here, as they will show how much an entity (class) is engaged with the purpose of the project. So again this is a good property of this methodology that provides us with another aspect or point of view of KG.
There are also some green lines that are drawn. It is worth saying, some of the relations (object properties) are common between some scenarios which seem normal. For example in two scenarios we need to find a bar near the hotel where a tourist resides, in these cases, we just show the relation one time and this connection will satisfy all other same situations. Based on

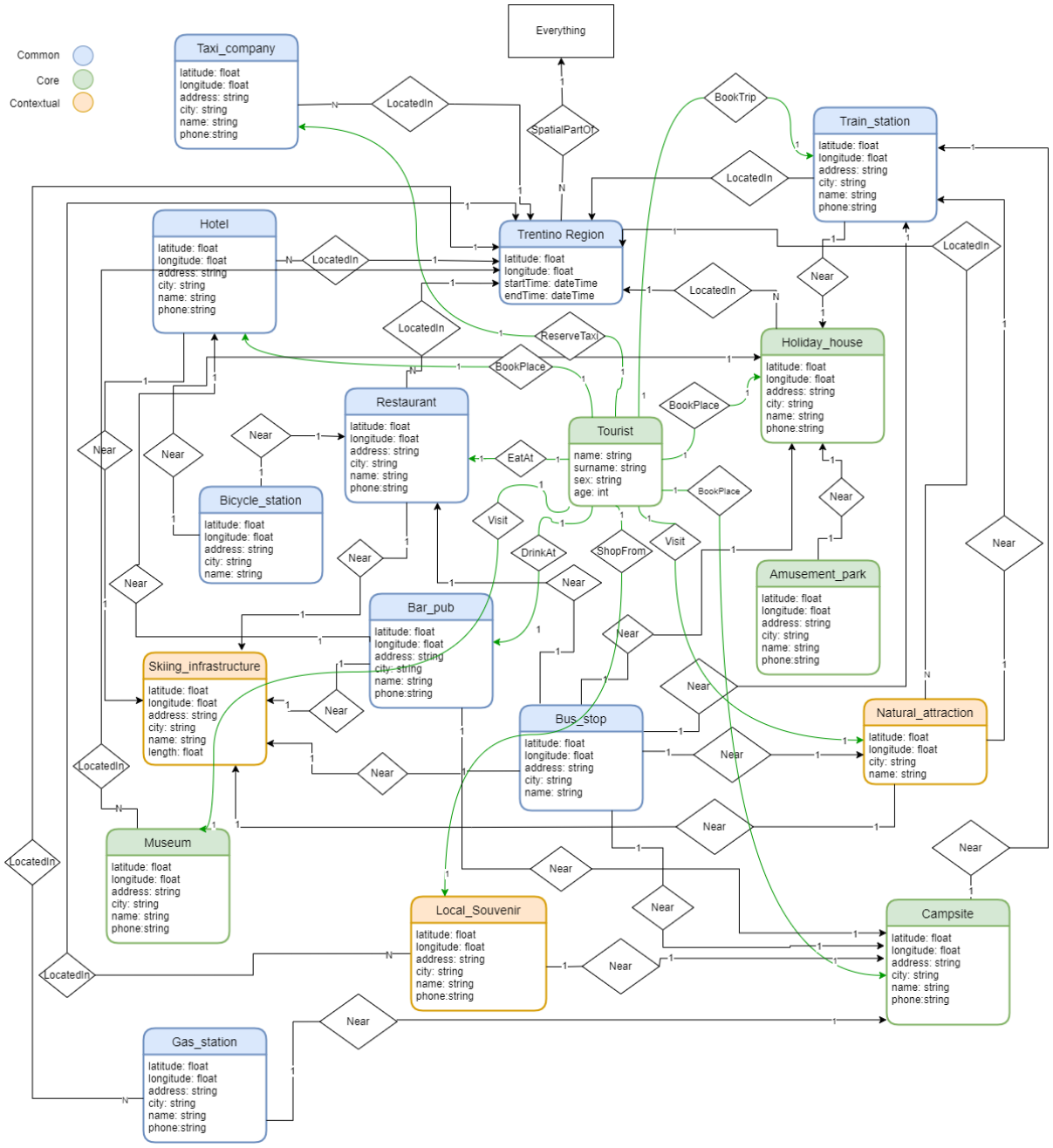
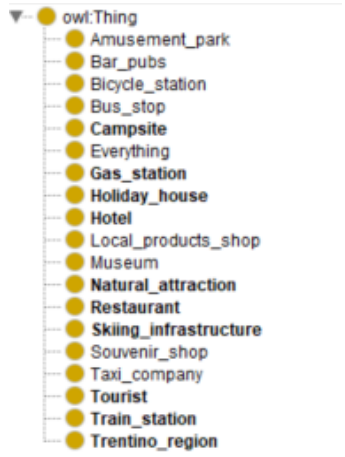


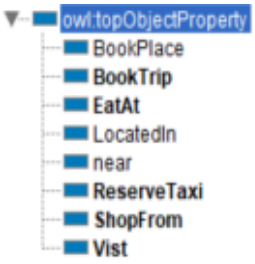
Figure 3: Teleology

The green lines were drawn to make the chart more readable and they do not have any other meaning.

the things we say so far, we can find every single answer to CQs in this chart for this context. Implementation of this phase was in the protègè. All the dataset ontologies were made before have been uploaded to the protege to make the integrated teleology.



(a) Teleology
Teleology in protègè.



(b) Teleology
Object properties.

Figure 4: Work on protègè

6.3 Data filtering and alignment

After adding the different ontologies, all the object properties between them have been defined based on the domain and the range of the connections.

In the informal modeling phase, some further choices were made regarding the division and choice of the datasets. As they were numerous, they were discarded or merged.
For example, the activity of skiing was represented by three datasets: ski rental, skiing infrastructure and skiing school. In the end, they have represented as the entity ”skiing infrastructure”, containing a list of the skiing areas present in the territory; skiing areas always have businesses such as skiing schools and ski rental shops.
As for merging, bars and pubs were united, as they were considered both as a business where a client can go for a drink during a night out.
Regarding data properties, in all phases of the data gathering, we stay focused on making datasets that have the same meaning for the same attributes. Also, we tried to use the common name for the attributes.

6.4 Issues

At the end of the inception phase, we finalized 21 different entities, based on what we found useful for each scenario; our aim was to make each one as different as possible and cover all types of tourism. However, we noticed that the entity graphs would be too crowded, making it very difficult to understand the flow of the connection.
For the sake of clarity, we decided to focus on fewer entities and improve the understanding of our thinking process.

7 Formal Modeling

7.1 ETG generation

For this phase, the following graph was produced.

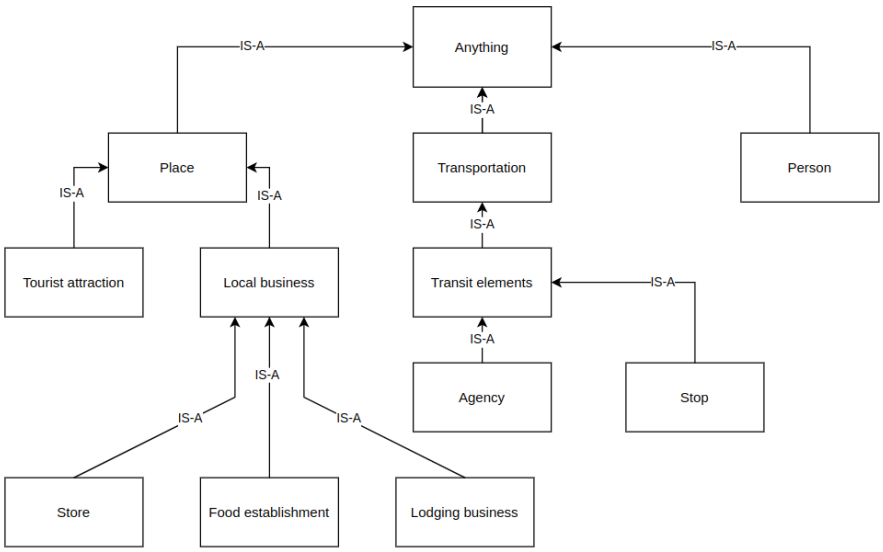


Figure 5: Ontology

Our model presents many kinds of entities, because the purpose, asking for tourist facilities, includes transportation, businesses (like shops or food facilities) and the tourist themselves. More precisely, the Place concept includes both tourist attractions and local businesses; local businesses include stores, food establishments and lodging businesses. Then, transportation includes stops for certain transportation methods and the companies dedicated to this function. However, we could not find a consistent way to represent Transport as ontology on the resource schema.org, so we based this part of the ontology from the slides offered by the professor. Lastly, the concept of tourist was not present on schema.org. However, the actions performed by a tourist are not different from the ones performed by the concept of Person, so we decided to assign this concept to our entity Tourist.

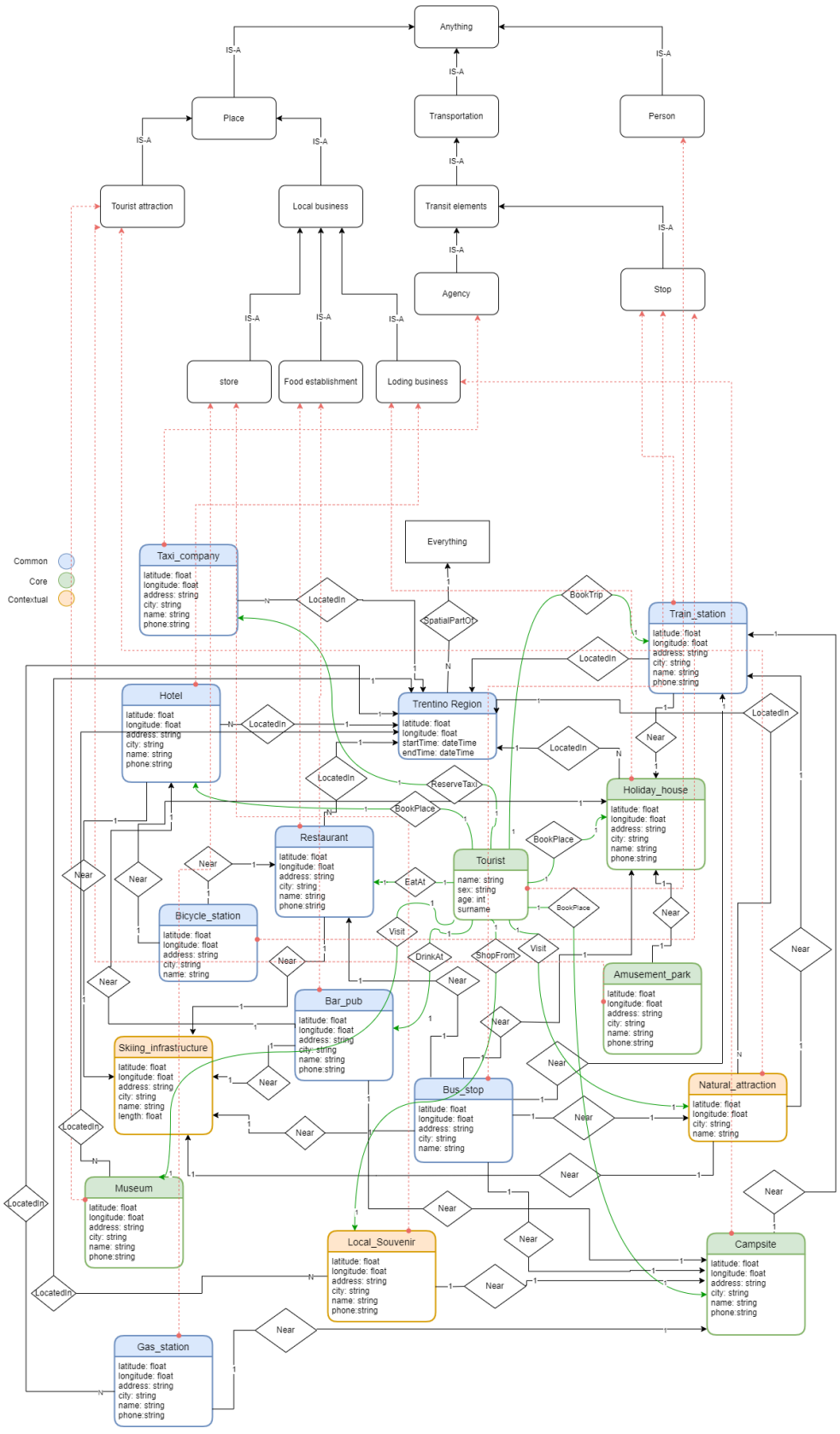


Figure 6: Complete teleontology

As for the teleontology, the entities are connected to the respective ontologies through a red line; again, the change of color and style is only to ensure better understanding. In addition, a ”simplified” graph was produced, in order to show how the single entities were connected to the ontology.

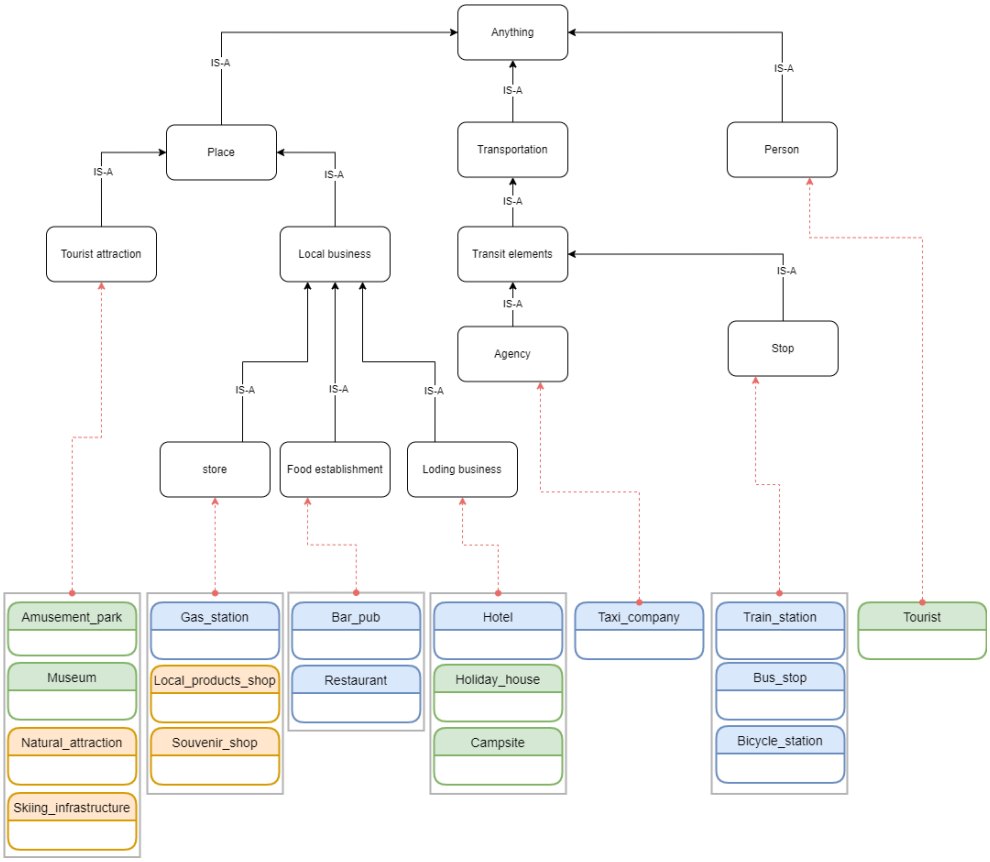


Figure 7: ”Simplified” teleontology, the only aim of this scheme is to better communicate to the reader the connections of the entities to the ontology

7.2 Language alignment

In order to formalize the terminology used in our diagrams, the teleontology file, in owl format, was uploaded on the KOS platform. Through the platform, we were able to formalize the object properties with universal definitions, which increase the re-usability of the resources. The language chosen for the definitions is English, so the language is consistent throughout the whole project.

As output, it produced the following rdf file.

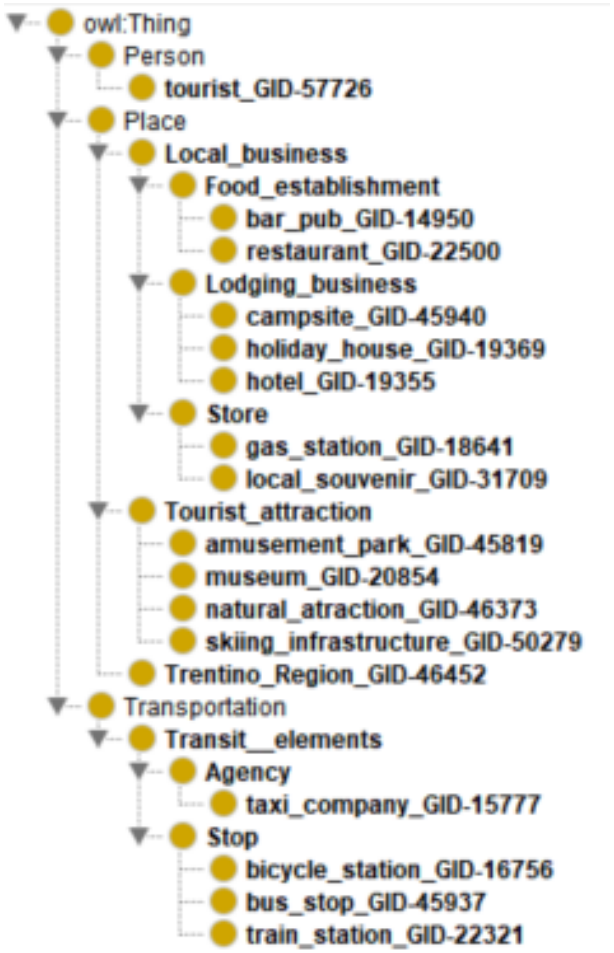


Figure 8: Language alignment scheme produced through KOS

7.3 Datasets alignment

After formalizing the actual connections between entities and their definitions, we need to create columns inside the datasets, to make concrete connections within entities.
The entity which performs the most relevant actions is the Tourist, so in the respective dataset there will be additional columns relating to the hotel they booked, the restaurant they will eat at and what they will visit. Moreover, many entities such as holiday houses and restaurants, in

the competency questions, require a bus stop or a bike rental stop in the vicinity, so the closest stops will be added as a new column.
All of these new columns in datasets will then be of use for adding connections between the datasets and build the final graph.

7.4 Issues

As mentioned earlier in the report, the amount of entities is large, which could make it harder to the reader to understand the diagram. However, we provided the diagram in Figure 7 in hope that it allows to understand better our thinking process.

8 KGC

8.1 Data integration

As a result of previous phases, each entity type is represented by a single dataset. The connections to another dataset are represented in the data by a new column (for example, in the dataset ”holiday house”, n_train represents the nearest train station to every holiday house).
As for the identifier, which connects the value of the n_... column to its dataset, we created an id for each entity, in order to use it as the uri in Karma. The id is simply the name of the entity type and the respective row number. An example below will be shown.

id	Name	Latitude	Longitude	n_bus
natural_attractions0	Itinerario Riserva Naturalistica del Laghestel	46.129722	11.243889	bus_stop177
natural_attractions1	Giro Dos Alt - Cime Alte	46.140278	11.256389	bus_stop179
natural_attractions2	Sentiero di Segonzano	46.181111	11.263333	bus_stop237
natural_attractions3	Giro dei Laghi	46.134444	11.249167	bus_stop178
natural_attractions4	Lago di Serraia	46.137857	11.256437	bus_stop179
natural_attractions5	Itinerario Lago di Lases - Lago di Santa Colomba	46.141667	11.218056	bus_stop226
natural_attractions6	Itinerario Monte Corno	46.230556	11.300833	bus_stop103
natural_attractions7	Itinerario Masi di Grumes - Cauria	46.223611	11.264444	bus_stop2335
natural_attractions8	Dosso di Costalta	46.125278	11.253611	bus_stop824
natural_attractions9	Itinerario Monte Calvo	46.125278	11.253611	bus_stop824
natural_attractions10	Dosso di Segonzano - da Bedollo	46.161111	11.301944	bus_stop185
natural_attractions11	Lago Santo di Cembra	46.196000	11.207999	bus_stop95
natural_attractions12	Dos de la Mot	46.108056	11.239444	bus_stop832
natural_attractions13	Lago di Lases	46.140599	11.221290	bus_stop226
natural_attractions14	Lago di Santa Colomba	46.124363	11.180677	bus_stop218
natural_attractions15	Trekking delle cave	46.142500	11.193056	bus_stop228

Figure 9: An example of how our id system works, using the natural attractions dataset.

8.2 Entity matching

Since our entities have always been separate to each other, we did not face an issue in semantic heterogeneity; to be precise, no singular entity was present in two different datasets. As mentioned in the previous subsection, the connections were made by us by adding columns referring to another dataset identifier.
For example, in one of our CQs it was needed to provide the closest bus stop to a holiday house, so a new column was added, providing the identifier of the closest bus stop for each of the rows.

8.3 Evaluation

Theoretically wise, we were able to answer to all the CQs presented in the informal modelling phase. Our entities and properties answer all the queries we proposed. This means that the

etype coverage for the CQs is:

$$COV_E(CQ_E) = \frac{|CQ_E \cap ETG_E|}{CQ_E} = 1$$

And the property coverage is:

$$COV_P(CQ_P) = \frac{|CQ_P \cap ETG_P|}{CQ_P} = 1$$

However, this does not imply that the project cannot be further developed, especially in the data properties. In particular, we attempted scraping of booking.com, in order to add the services that each hotel offers. Additionally, the skiing infrastructures presented more information, especially about opening times, both the time and seasonal.
Moreover, other ideas included adding the opening times of businesses and the type of food of restaurants. In the end, we decided to focus on executing correctly the phases of the project, instead of taking the time to expand it, in order to understand correctly the iTelos methodology.

Now onto the coverage on the reference ontologies (RO), we calculate again the entity and property connectivity. The etype coverage is:

$$COV_E(RO_E) = \frac{|RO_E \cap ETG_E|}{RO_E} = 0.8$$

Some etypes suggested by the ontologies were discarded, since we did not find them interesting or suitable for the competency questions.

The property coverage is:

$$COV_P(RO_P) = \frac{|RO_P \cap ETG_P|}{RO_P} = 0.45$$

The low value does confirm what mentioned before, which is that there is a need to expand the data properties. However, some properties given in the reference schema are of no use for us. For example, the tourist in ontology is represented by Person, as there is no ontology specifically for a tourist; for this reason, many data properties are very general or refer to contexts which are not useful for our purpose.

Then the final graph connectivity is evaluated. In the following formulas is calculated respectively the Entity Connectivity (E), the Object Property Connectivity (Op) and the Data Properties Connectivity (Dp) over the etypes T:

$$\sum_{n=1}^{N} E(T_k) = 5560$$
$$\sum_{n=1}^{N} Op(T_k) = 3026$$
$$\sum_{n=1}^{N} Dp(T_k) = 24000$$

9 Outcome Exploitation

All previous steps have resulted in a KG graph as we expected. For the visualization and also exploiting part, we used the graph DB, which is one of the famous tools for saving and retrieving the queries based on the RDFs as we used them in this project. In the following, you can see the related information to this project.

9.1 KG Information

In this project, we use 15 E-types classes and for each one, we have its dataset So they have their own purpose-specific properties. Moreover, 10 object properties (connection type between classes) have been used in this KG, and also we used 27 data properties. The number of entities can be seen below for each E-type class:

256 for bars and pubs, 379 for holiday houses, 13 for amusement parks, 121 stops for bicycle stations, 3942 for bus stops, 25 for campsites, 20 gas stations, 259 hotels, 61 local shops and souvenir stores, 41 museums, 46 natural attractions, 285 restaurants, 17 skiing infrastructures, 3 taxi companies, and 47 train stations.

Furthermore, we can access more information about the KG by using different features of the graph DB tools. For example we can see the class relationships from the section that has the same name on the graph DB. All the relationships are shown in the following figure.

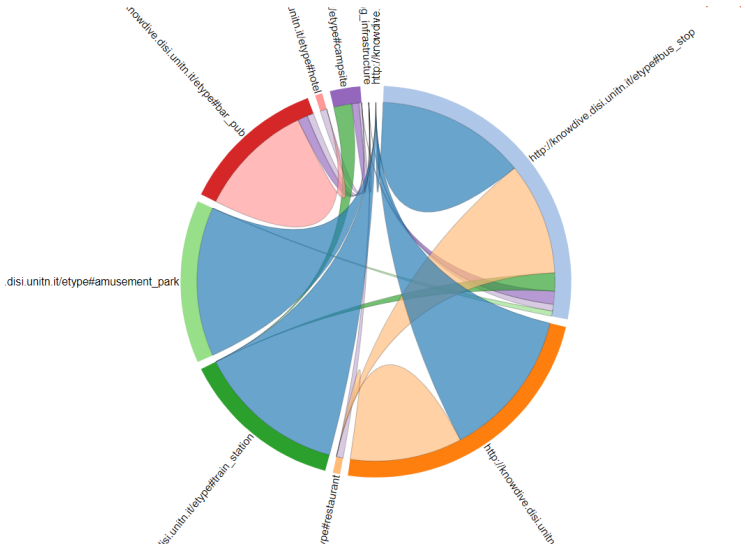


Figure 10: An example class relationship.

9.2 KGE Exploitation

This KG is exploitable based on the search on graph DB with SPARQL and also from the visualization section of the graph DB. Based on this, we can satisfy the CQs that have been made in the previous sections. This means that the graph that is representative of this KG can cover our needs for the tourist facilities as our scenarios try to simulate them.

In a nutshell this KG can be used for:
In the following figure, you can see the example of the visualization part:

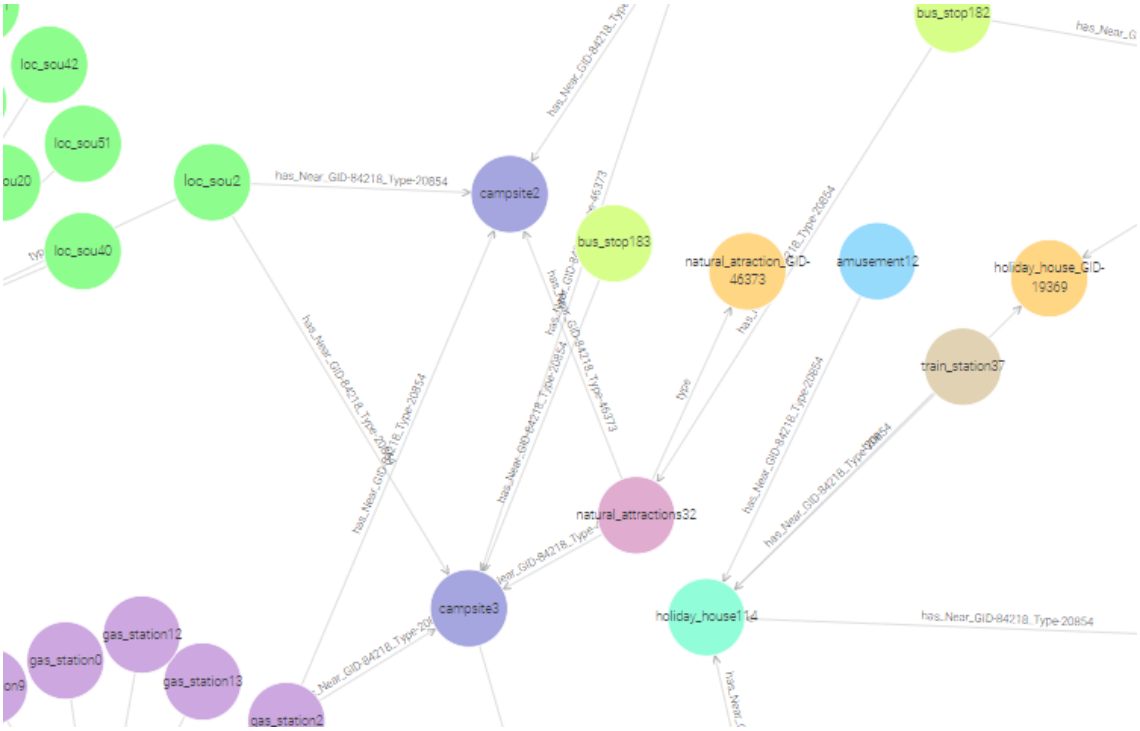


Figure 11: An example exploitation in the Visualization part.

This figure shows some connectivity between some entities. to be more precise in this specific case, we can find the near bus stop, natural attractions, restaurant, and hotel for a skiing infrastructure based on our graph which could be related to the Jannis persona in our scenarios.

As we can expect, we can expand this graph and check all the possibilities for our personas in different cases. It means that we can expand each node and see all the possibilities throughout this space which is created by the purpose of this project.

Another way for exploiting this KG is using Sparql (which is more precise and faster for finding a single CQ answer): For example the bellow figure show the query for the firat cq:

How can he find a holiday house in the Trentino region for his family?

```
PREFIX prop: <http://knowdive.disi.unitn.it/etype#>
PREFIX ds: <http://localhost:8080/source/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

select ?name ?lat ?long ?city ?address where {
    ?holiday_houses rdf:type prop:holiday_house_GID-19369;
                    prop:has_name_GID-2_Type-50279 ?name;
                    prop:has_latitude_GID-46264_Type-20854 ?lat;
                    prop:has_longitude_GID-46270_Type-18641 ?long;
                    prop:has_city_GID-45969_Type-50279 ?city;
                    prop:has_address_GID-36367_Type-15777 ?address
}
```

Figure 12: SPARQL query for exploiting the KG for the first CQ.

This query will retrieve all the holiday house name, city which they are located in, latitude and longitude, and the address of that holiday house. The bellow figure shows the outcome (a part of) of this query:

	name	lat	long	city	address
1	"CASA TITTI"	"46.0524"	"11.4578"	"Borgo Valsugana"	"VIA MAZZINI, 4"
2	"MOSER MARIA RICCARDA - CASA LIBA"	"46.00282704263 18"	"11.2464380264282"	"Calceranica Al Lago"	"Corso Alpini, 41 "
3	"MARTINELLI ANNA MARIA"	"46.0042"	"11.2458"	"Calceranica Al Lago"	"VIA AL LAGO, 11"
4	"APPARTAMENTO AL LAGO"	"46.00395235792 67"	"11.258647441864"	"Calceranica Al Lago"	"VIALE VENEZIA, 10"

Figure 13: Outcome of exploiting the KG for the first CQ.

Another example query can be seen in the following figure:

```
PREFIX prop: <http://knowdive.disi.unitn.it/etype#>
PREFIX ds: <http://localhost:8080/source/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

select ?hh_name ?hh_lat ?hh_long ?bus_stop_name ?bus_stop_lat ?bus_stop_long where {
    ?holiday_house prop:has_name_GID-2_Type-50279 "Alex";
                   prop:has_name_GID-2_Type-50279 ?hh_name;
                   prop:has_latitude_GID-46264_Type-20854 ?hh_lat;
                   prop:has_longitude_GID-46270_Type-18641 ?hh_long.
    ?near_things prop:has_Near_GID-84218_Type-20854 ?holiday_house;
                 rdf:type prop:bus_stop_GID-45937;
                 prop:has_name_GID-2_Type-50279 ?bus_stop_name;
                 prop:has_latitude_GID-46264_Type-20854 ?bus_stop_lat;
                 prop:has_longitude_GID-46270_Type-18641 ?bus_stop_long;
}
```

Figure 14: SPARQL query for exploiting the KG for the second CQ.

This query will retrieve the closest bus stop to the tourist holiday houses. and the result can be seen in the bellow figure:

	hh_name	hh_lat	hh_long	bus_stop_name	bus_stop_lat	bus_stop_long
1	"Alex"	"46.09804"	"11.112821"	"Gardolo 4 Nov. Canova"	"46.098861"	"11.112921"

Figure 15: Outcome of exploiting the KG for the second CQ.

Other query is for finding the closest bicycle station to the holiday house as can be seen in the following figure:

```
PREFIX prop: <http://knowdive.disi.unitn.it/etype#>
PREFIX ds: <http://localhost:8080/source/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

select ?hh_name ?hh_lat ?hh_long ?bike_name ?bike_lat ?bike_long where {
    ?holiday_house prop:has_name_GID-2_Type-50279 "Alex";
                   prop:has_name_GID-2_Type-50279 ?hh_name;
                   prop:has_latitude_GID-46264_Type-20854 ?hh_lat;
                   prop:has_longitude_GID-46270_Type-18641 ?hh_long.
    ?near_things prop:has_Near_GID-84218_Type-20854 ?holiday_house;
                 rdf:type prop:bicycle_station_GID-16756;
                 prop:has_name_GID-2_Type-50279 ?bike_name;
                 prop:has_latitude_GID-46264_Type-20854 ?bike_lat;
                 prop:has_longitude_GID-46270_Type-18641 ?bike_long;
}
```

Figure 16: SPARQL query for exploiting the KG for the third CQ.

The result can be seen in the bellow figure:

	hh_name	hh_lat	hh_long	bike_name	bike_lat	bike_long
1	"Alex"	"46.09804"	"11.112821"	" Canova"	"46.0992211"	"11.1092928"

Figure 17: Outcome of exploiting the KG for the third CQ.

The bellow query will retrieve all natural attractions:

```
PREFIX prop: <http://knowdive.disi.unitn.it/etype#>
PREFIX ds: <http://localhost:8080/source/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

select ?name ?lat ?long where {
    ?natural_attraction rdf:type prop:natural_atraction_GID-46373;
                        prop:has_name_GID-2_Type-50279 ?name;
                        prop:has_latitude_GID-46264_Type-20854 ?lat;
                        prop:has_longitude_GID-46270_Type-18641 ?long.
} limit 10
```

Figure 18: SPARQL query for exploiting the KG for the third CQ.

The result can be seen in the bellow figure:

	name	lat	long
1	"Ecomuseo dell'Argentario"	"46.0886"	"11.1822"
2	"Biotopo Le Grave"	"46.12397"	"11.178523"
3	"Lago delle Piazze"	"46.153711"	"11.278925"
4	"Sorgente Ferruginosa"	"46.125278"	"11.253889"
5	"CENTRO VISITATORI DEL PARCO NATURALE DI PANEVEGGIO"	"46.3074061"	"11.699201"
6	"Itinerario Malga Vernera - Rifugio Vasoni"	"46.217778"	"11.3275"
7	"Lago delle Buse"	"46.179277"	"11.32375"
8	"Dos de la Mot"	"46.108056"	"11.239444"
9	"Dosso di Segonzano - da Bedollo"	"46.161111"	"11.301944"
10	"Itinerario Riserva Naturalistica del Laghestel"	"46.129722"	"11.243889"

Figure 19: Outcome of exploiting the KG for the third CQ.

As it has been shown we can exploit the KG based on the SPARQL queries on graph DB.

10 Conclusions & Open Issues

The project deadlines set during the course were respected and everything was achieved on time, regarding the theoretical concepts, the data preparation and the report.
However, we faced some issues with the software proposed for the ontology and the knowledge graph creation:

- We created the teleology by building the single entity types singularly, and then calling the URI of each for building the final file, as we were instructed. However, this rdf file did not work on Karma on either of our computers, trying both Windows and Linux operating systems. This means that for a long time we could not progress in that area of the project, despite having tried many different solutions. In the end, we created the teleology in one single file, which we are aware is not the correct way, but it allowed us to progress further in the project.
- GraphDB was not able to upload some of our ttl file, apparently for an error created by the character ” present in some names or addresses of the entities.
- The other issues ar the data sets. Of course, the datasets in this implementation were limited to the scenarios that we made. Therefore, by expanding more scenarios we can cover more datasets and needs in the real world. Also, in some datasets, we have missing data (Like addresses and some phone numbers) that can be filled in the future steps to make this knowledge graph more sound. Of course we can expand the entities, data properties and object properties based on the more complecated scenarios.

These obstacles slowed us down in the process of building this project, but in the end we were able to overcome them.

10.1 Conclusions

The purpose was satisfied, as we are able to present information to perspective tourists willing to come to Trentino; our graph is able to present information based on people’s preferences and taste. If we did not have issues with the software, we could have expanded the project by introducing more specific data properties regarding each entity type, to enrich the graph.

However, what we considered the most important thing was to have a correct and working structure, which is what we had our focus on since the beginning.

Therefore, generally speaking we cold Use iTelos methodology step by step during this course and apply different phases for that. At the end we were able to exploit the KG that we made based on these steps and answer to all the CQs.

Therefore, we could use iTelos methodology step by step during this course and apply different phases for the data and the concepts we were engaged with. In the end, we were able

to exploit the KG that we made based on these steps and answer all the CQs. In a nutshell, we could use iTelos methodology for designing and applying the Knowledge graph for tourist facilities in Trentino region.