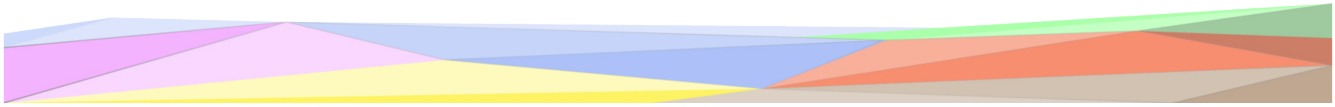
UNIVERSITY
OF TRENTO - Italy

Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

KGE 2023 - Composing Trentino Natural Places, Facilities and Transportation Ways by iTelos

Document Data:	Reference Persons:
October 9, 2023	Xiaoyue Li



Index:

1 Introduction **1**

2 Project Description **1**

2.1 Objective . . . 1

2.2 Project development . . . 2

2.2.1 Data Production . . . 2

2.2.2 Data Composition . . . 2

3 Purpose Formalization **3**

4 Information Gathering **5**

4.1 Data and Knowledge Source . . . 5

4.1.1 Informal Data and Knowledge Source from Producer . . . 5

4.1.2 Formal Data and Knowledge Source from Consumer . . . 7

4.2 Resource Collection, Processing and Scraping . . . 7

4.2.1 Informal Resource Collection, Processing and Scraping from Producer . . . 7

4.2.2 Formal Resource Collection and Scraping from Consumer . . . 8

4.3 Integrate Scraped Data with Scraped Schemas Using Karma . . . 9

5 Language Definition **11**

5.1 Formalize Etypes with Properties from resources by Producer . . . 12

5.1.1 Formalize Etypes with Properties from Punti di interesse del Trentino Resource . . . 12

5.1.2 Formalize Etypes with Properties from Trasporti pubblici del Trentino Resource . . . 13

5.1.3 Formalize Etypes with Properties from Trentino OSM Resource . . . 14

5.2 Formalize Object Properties for Composing Resources from Consumer . . . 15

6 Knowledge Definition **16**

6.1 Teleologies from Producer . . . 16

6.2 Composed Trentino Spatial Teleology and Teleontology from Consumer . . . 19

6.2.1 Trentino Spatial Teleology . . . 19

6.2.2 Trentino Spatial Teleontology . . . 20

7 Data Definition **22**

7.1 Dataset formatting by Consumer . . . 22

7.2 Entity identification by Producer . . . 23

7.3 Object property ‘On’ identification by Consumer . . . 24

7.4 Data mapping . . . 24

8 Outcome Exploitation **25**

8.1 Evaluation . . . 25

8.1.1 Coverage metric . . . 25

8.1.2 Connectivity metric . . . 27

8.2 KG exploitation 27

9 Conclusion and Open Issues **28**

Revision History:

Revision	Date	Author	Description of Changes
0.1	October 9, 2023	Author1	Document created

1 Introduction

The integration of spatial datasets, such as natural places, facilities and transportation ways, in Trentino can provide a comprehensive and accurate representation of the region's infrastructure. This information can be used for urban planning and emergency management. Having precise and up-to-date data about the facilities, transportation ways and characteristics of natural places in Trentino offers convenience for the region's residents.

The iTelos methodology, provided by Knowledge Graph Engineering (KGE), develops a process to enhance resource reusability within a specific domain. Following the iTelos methodology, we conduct a project to manage and compose knowledge and data resources about the places in Trentino.

The report is organized as follows. Section 2 defines the purpose and specific domain of interest of the project. It provides a high-level overview of low quality resources that will be created by producer, and high quality resources that will be identified and composed by consumer. Section 3 formalize our research purpose by defining a set of scenarios, personas, competency questions (CQs) to identify a set of entities, which are modeled in an Enity-Relationship (ER) model. Section 4 delineates the utilization of three distinct data and knowledge sources, namely Trentino OSM, Punti di interesse del Trentino, and Trasporti pubblici del Trentino. This section is dedicated to select, separate and scrape data from the three sources. In Section 5, we extract entity types, data properties, and object properties from the aforementioned data, which are subsequently formalized through linkage with Global Identifiers (GIDs) in the Universal Knowledge Core (UKC). Section 6 is devoted to the formalization of teleologies for the three sources, which are then integrated into a comprehensive Trentino Spatial Teleology. Concurrently, we construct the Trentino Ontology and integrate it with the Trentino Spatial Teleology to form the Trentino Spatial Teleontology. Section 7 outlines the formalization of data based on the Trentino Spatial Teleontology. Section 8 evaluate the quality of our data and knowledge results and show the Knowledge Graph (KG) exploitation. Finally, we draw the conclusion and give open issues in 9.

2 Project Description

2.1 Objective

Project purpose: The goal of this project is to build comprehensive Trentino spatial reference resources, encompassing massive *natural places*, *facilities* and *transportation ways*, to support services in Trentino territory. This information will enable individuals to locate places and gain a holistic understanding of the locations within Trentino.

Project domain of interest (DoI): Our study focuses on the annotated places, including natural places, facilities and transportation ways from OpenStreetMap in the June of 2023, points of interest in the April of 2014 and public transportation stops from the June to Septemeber of 2023 in Trentino.

2.2 Project development

2.2.1 Data Production

In this phase, producer aims to produce datasets that meet our project purpose, these resources need to be created if they that don't exist yet or exist in a bad quality. In our project, we need to create and formalize resources about natural places, facilities, and transportation ways in Trentino. The data about natural places and facilities should have point geometry features, represented by a pair of longitude and latitude coordinates, such as POINT(11.6008628 46.3124594). Natural places typically represent peaks, trees, etc., while facilities typically represent restaurants, bus stops, train stations, etc. The data about transportation ways should have line geometry features, represented by a set of pairs of longitude and latitude coordinates, such as MULTILINESTRING((11.4088526 46.4936212,11.4089414 46.4935992,11.4090736 46.493582)). Transportation ways are typically used to represent roads and railways.

2.2.2 Data Composition

Consumer identifies the existing high quality resources to satisfy our project purpose, i.e., Trentino OSM dataset (data resource) and Trentino OSM lightweight Ontology (LWOntology) (knowledge resource), which are provided by Knowledge Graph Engineering (KGE) course. They contains a large of various place resources, we use the data and knowledge about natural for integrating natural place resources purpose, point of interest and traffic for integrating facility resources purpose, railway and transport resources for integrating transportation ways resources purpose. All the selected resources from consumer will be composed with the formalized resources from producer. Here we propose data composition should be organized in two dimensions, which we refer to as *vertical composition* and *horizontal composition*. *Vertical composition* involves identifying the same natural places, or facilities, or transportation ways and composing them into one place. For example, if the 'Sun Supermarket' facility is annotated in both Dataset 1 and Dataset 2, we need to identify these instances and retain only one 'Sun Supermarket' in our composed dataset. *Vertical composition* can effectively reduce the scale of the data by eliminating redundant instances. *Horizontal composition* refers to identifying which transportation ways that natural places or facilities are located on. For example, the 'Sun Supermarket' facility is located on the 'Fifth Street', which is a transportation way. *Horizontal composition* expands the data by adding relations between natural places and facilities with transportation ways. Haversine distance, a calculation method that determines the distance between two points on a great circle, e.g., earth surface, based on their longitude and latitude, which can be utilized to index transportation ways for natural places and facilities based on their geographical coordinates, e.g., we found the transportation ways for a facility if their minimal distance is less than 0.05 km.

3 Purpose Formalization

Our project purpose is to integrate Trentino *natural places*, *facilities* and *transportation ways* resources to furnish a comprehensive Trentino spatial reference information.

To describe multiple aspects considered by the project purpose, we list a set of usage scenarios as follows:

- **Scenario 1.** Trentino has stunning natural environment, which includes a lot of peaks, the abundance of springs and trees. It is a popular destination for outdoor enthusiasts and nature lovers.
- **Scenario 2.** Trentino has efficient public transportation system, which includes an extensive network of bus stops, train stations, etc. This makes it easy for residents and visitors to travel around the region and explore all that it has to offer.
- **Scenario 3.** Trento is the capital of the autonomous province of Trentino. The city center is bustling with a lot of restaurants, bars, cafes, supermarkets and other facilities, providing people with a wide range of options for dining and shopping.
- **Scenario 4.** The Department of Information Engineering and Computer Science (DISI) is located in Povo of Trento, which has fewer facilities but provides necessary bus transportation options for students, making it easy for them to commute to and from the school.

In the scenarios defined above, we represents a set of real users with specific features included in the project purpose, whhich are listed as follows:

- **Personas 1.** Lily, a 60-year-old woman living in Obergummer, Rovereto, is an outdoor enthusiast. She loves to explore the natural attractions of Trentino.
- **Personas 2.** Anna, a 28-year-old doctoral student, plays a key role in her research group. She is responsible for organizing various activities for the group.
- **Personas 3.** Luca, a 25-year-old master student lives around the Trento train station of the center of Trento, he has a passion for cooking and tasting food.

Taking into account the personas in the scenarios defined, we create Competency Questions (CQs):

- **CQ 1.** Please recommend Lily the nearest 3 peaks from her home.
- **CQ 2.** Can you provide the roads information that Lily can drive to the 3 peak destinations?
- **CQ 3.** On Monday, Anna wants to find an opening restaurant near DISI for the party of her research group. For the transportation convenience of students, this restaurant should near at least one bus stop.
- **CQ 4.** Luca is looking for a library near his home that is easily accessible by bus stops and surrounded by restaurants.
- **CQ 5.** Luca wants to shop in the supermarket that is nearest from his home.

From the *CQs*, referring to *Personas* and *Scenarios*, we extract *Entities* with *properties*. These entities are categorized as either Common, Core, or Contextual entities by considering *Focus classification* and *Popularity classification*. The details of this work are outlined in Table 1.

Table 1: Entities extraction and classification

Scenarios	Personas	CQs	Entities	Properties	Focus classification	Popularity classification
1	1	1	Lily home	id: int, name: string, coordinates: string	Contextual	Contextual
1	1	1	peak	id: int, name: string, coordinates: string	Core	Common
1	1	2	road	id: int, name: string, coordinates: string, oneway: boolean, maxspeed: int	Common	Common
2, 3, 4	2	3	DISI school	id: int, name: string, coordinates: string, timetable: string	Contextual	Contextual
2, 3, 4	2	3	restaurant	id: int, name: string, coordinates: string, timetable: string	Core	Common
2, 3, 4	2	3	bus stop	id: int, name: string, coordinates: string, timetable: string	Core	Common
2,3	3	4	Luca home	id: int, name: string, coordinates: string	Contextual	Contextual
2,3	3	4	library	id: int, name: string, coordinates: string, timetable: string	Core	Common
2,3	3	4	bus stop	id: int, name: string, coordinates: string, timetable: string	Core	Common
2,3	3	4	restaurant	id: int, name: string, coordinates: string , timetable: string	Core	Common
3	3	5	Luca home	id: int, name: string, coordinates: string, timetable: string	Contextual	Contextual
3	3	5	supermarket	id: int, name: string, coordinates: string, timetable: string	Core	Common

Based on the Common and Core entities in Table 1, we design an Entity–relationship (ER) model as Figure 1.

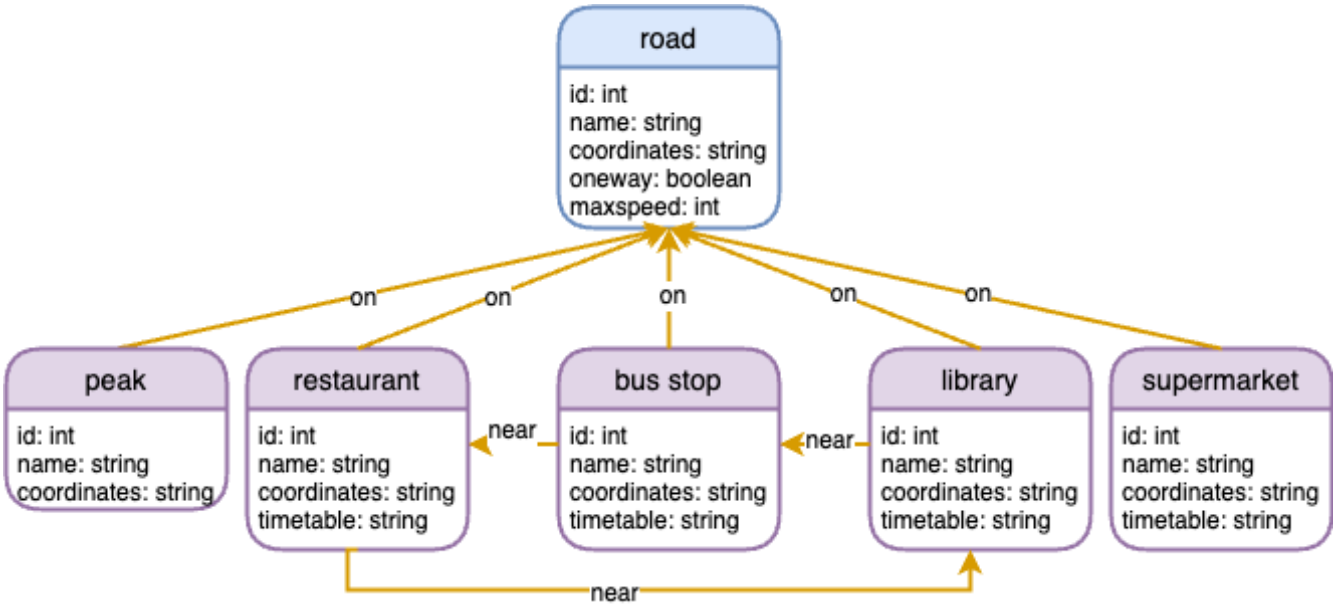


Figure 1: ER model.

4 Information Gathering

In this section, the producer aims to identify informal data and knowledge sources, which will be integrated to achieve the project's objectives. Then the producer collects informal data and knowledge from the informal sources, and processes them to extract the resources that will be used. While, the consumer aims to identify formal data and knowledge sources, then scrapes formal data and knowledge sources to integrate all resources.

4.1 Data and Knowledge Source

4.1.1 Informal Data and Knowledge Source from Producer

We identified two informal resources: Punti di interesse del Trentino and Trasporti pubblici del Trentino (formato GTFS). Their catalogs can be found in Table 2 and Table 3.

[1] https://dati.trentino.it/dataset/punti-di-interesse-del-trentino
[2] https://dati.trentino.it/dataset/trasporti-pubblici-del-trentino-formato-gtfs/resource/3394a64e-b7a6-4b6f-87ec-03f06ff7f2b7

Table 2: Punti di interesse del Trentino catalog.

Resource name	**Punti di interesse del Trentino**
Domain	Trentino (Italy)
Keywords	Points of interest
Language	Italian
Provider	Tourism and Sports Service
Data URL	POI - Trentino [1]
Data format	.json file
Data description	This dataset contains data about point of interests in Trentino, such as bars, hotels, restaurants, etc., totally 112 types of points of interest.
Knowledge URL	N/A
Knowledge description	N/A

Table 3: Trasporti pubblici del Trentino catalog.

Resource name	**Trasporti pubblici del Trentino (formato GTFS)**
Domain	Trentino (Italy)
Keywords	Public transportation
Language	Italian
Provider	Public Transport Service
Data URL	https://dati.trentino.it/dataset/trasporti-pubblici-del-trentino-formato-gtfs
Data format	.txt files
Data description	This dataset contains urban and suburban public transportation data. It describes public transportation (train, bus and cable car) static and scheduled information e.g., bus stops, routes, transfers, fares, calendar.
Knowledge URL	The MITT - OpenData Manual - v.7PDF Popular [2], we call it as Trentino GTFS Knowledge
Knowledge description	The MITT - OpenData Manual - v.7PDF Popular offers a comprehensive overview and detailed explains of the data, which is structured according to the General Transit Feed Specification (GTFS) format.

4.1.2 Formal Data and Knowledge Source from Consumer

OSM is a freely available and open geographic database that is maintained through public collaboration by volunteers. The catalog of Trentino OSM data and knowledge resources can be found in the Table 4.

Table 4: Trentino OSM data and knowledge catalog.

Resource name	Trentino OSM
Domain	Trentino (Italy)
Keywords	Space, Geography, Trentino
Language	English, Italian
Provider	Knowledge Graph Engineering (KGE) course
Data URL	OSM Places dataset [3]
Data format	.txt files
Data description	KGE course provides a cleaned and classified OSM dataset with boundary of Trentino. The Trentino OSM dataset is organized into 11 folders and a building.txt file. Each folder corresponds to a specific type of place. For instance, the 'road' folder contains records of all places categorized as roads. Furthermore, the places within each folder are further classified according to their subcategories. As an example, within the 'road' folder, there is a 'major_road' subfolder that contains records of places classified as major roads.
Knowledge URL	OSM-LO.UAN.owl [4], OpenStreetMap Data in Layered GIS Format [5]
Knowledge description	OSM-LO.UAN.owl signifies the Trentino OSM Lightweight Ontology. The Knowledge Graph Engineering (KGE) course introduces the Trentino OSM LWOntology to encapsulate the class representations and hierarchical structures of Trentino OSM locations. This ontology serves as a foundational framework for categorizing Trentino OSM locations based on their class values. The Trentino OSM LWOntology encompasses a total of 791 classes, each elucidated by a comment. These classes are arranged in a tree structure with a maximum depth of seven. The OSM Place dataset, an organized version of the OSM dataset in Trentino, aligns with the Trentino OSM LWOntology, with each discrete data file corresponding to a leaf node class within the Trentino OSM LWOntology.

4.2 Resource Collection, Processing and Scraping

4.2.1 Informal Resource Collection, Processing and Scraping from Producer

We collected the data and knowledge resources from Punti di interesse del Trentino and Trasporti pubblici del Trentino sources. **Scraped Punti di interesse del Trentino Data.** We collected the data from Punti di interesse del Trentino dataset, i.e., POI - Trentino.json file. This file contains information on points of interest, which are related to a total of 112 categories. We then processed the POI - Trentino.json file by separating it based on the values of the 'category' property, with Italian values translated into English prior to separating. The separated data were unified with .txt files. Finally, from the separated datasets, we scraped the data related

[3] https://datascientiafoundation.github.io/LiveDataTrentino/datasets/OSM%20Places/
[4] https://datascientiafoundation.github.io/LiveKnowledge/datasets/osm-lightweight-ontology/
[5] https://download.geofabrik.de/osm-data-in-gis-formats-free.pdf

with facilities as shown in Table 5. These files are related to facilities, not natural places, and transportation ways.

Table 5: Scraped Punti di interesse del Trentino data.

Files	Domain	Description	Source
bar.txt, bus station.txt, cafe.txt, library.txt, restaurant.txt, supermarketconvenience store.txt, train station.txt.	facility	Each file contains information on points of interest with one category.	Punti di interesse del Trentino dataset

Scraped Trasporti pubblici del Trentino Data. We collected the data from Trasporti pubblici del Trentino dataset are the stops.txt, routes.txt, trips.txt and stop_times.txt files. These files contain information about stops, and detailing the planned routes and trips that stop at these stops. They also provide the schedules for buses, trains, or cable cars, indicating the times they are expected to arrive at these stops. Due to we only consider the spatial data, i.e., the stops. We based on the collected files to separate the stops.txt data according to whether the stopped vehicles are buses or trains. Finally, we scraped the data related with facilities as shown in Table 6. These files are related with facilities, not natural places and transportation ways.

Table 6: Scraped Trasporti pubblici del Trentino data.

Files	Domain	Description	Source
bus stop.txt, train stop.txt	facility	The two files reprsent the information about bus stops and train stops. Deatils of properties can be found in see the MITT - OpenData Manual - v.7PDF Popular	Trasporti pubblici del Trentino dataset

Scraped Trasporti pubblici del Trentino Teleology. from the The MITT - OpenData Manual - Manual - v.7PDF Popular file, we identitied an etype ‘stop’ with a set of data properties, which are built as a Scraped Trasporti pubblici del Trentino Teleology as shown in Figure 2.

4.2.2 Formal Resource Collection and Scraping from Consumer

Consumer collected and scraped data and knowledge schema resources from the whole Trentino OSM dataset and Trentino OSM LWOntology, respectively.

Scraped Trentino OSM Data. We scraped the files with rich data related to natural places, facilities and transportation ways from the Trentino OSM data are as shown in Table 7.

Scraped Trentino OSM Ontology. Each file in the scraped Trentino OSM data abstract an etype. To organize these etypes, we scraped a Scraped Trentino OSM Ontology from the Trentino OSM LWOntology as shown in Figure 3. It is part of the Trentino OSM LWOntology and designed by Protege. We renamed the osm_place in Trentino OSM LWOntology to Trentino_place in the Scraped Trentino OSM Ontology.

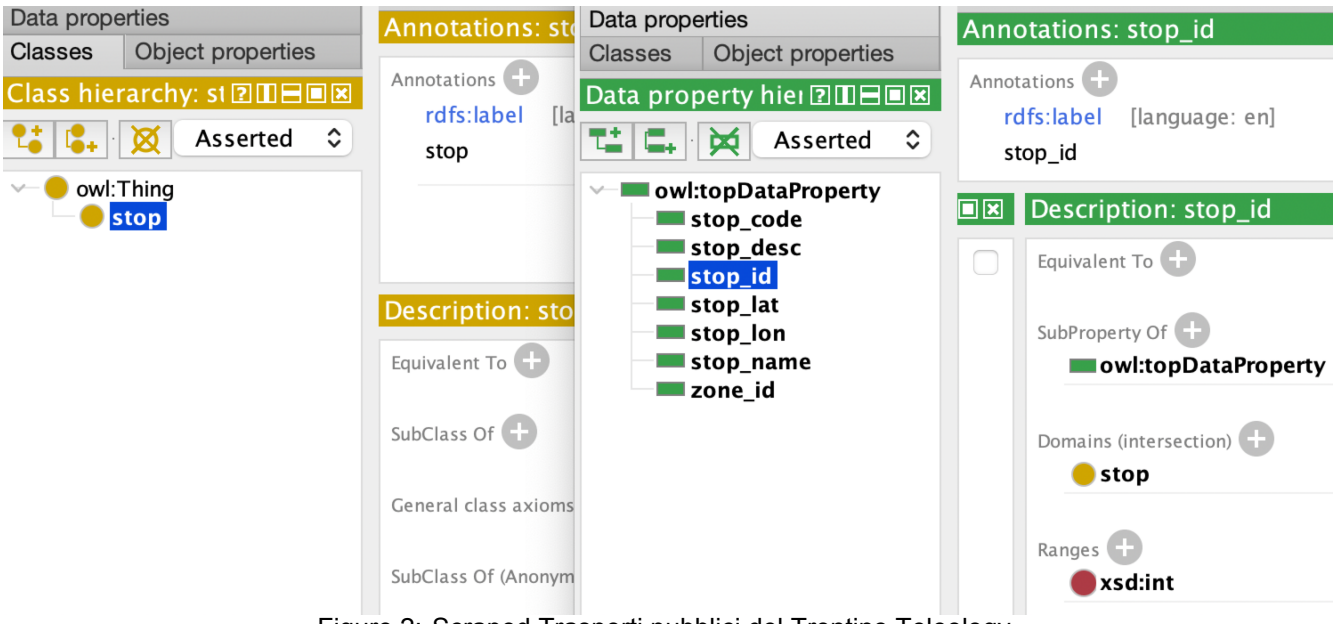


Figure 2: Scraped Trasporti pubblici del Trentino Teleology.

Table 7: Scraped Trentino OSM data.

Files	Domain	Description
point_peak.txt, point_spring.txt, point_tree.txt	point_natural place	Each file records information about places that fall under the category indicated by the file name.
point_bar.txt, point_bus_station.txt, point_bus_stop.txt, point_cafe.txt, point_library.txt, point_railway_halt.txt, point_railway_station.txt, point_restaurant.txt, point_supermarket.txt	facility	
primary.txt, secondary.txt, step.txt, tertiary.txt, very_small_road_service.txt, pedestrian.txt, bridleway.txt, cycleway.txt, footway.txt, living_street.txt, minor_road_residential.txt, path.txt.	transportation way	
funicular.txt, narrow_gauge.txt, rail.txt	transportation way	

4.3 Integrate Scraped Data with Scraped Schemas Using Karma

Karma [6] is an information integration tool that can integrate data with knowledge schema. We use Karma to map the selected data with the built ontologies or teleontologies. For example, each file in the scraped Trentino OSM data maps an etype in the Scraped Trentino OSM Ontology. As shown in Figure 4, which is the 'point_bus_stop.txt' opened by Karma, which maps the etype 'point_bus_stop' etype in the Scraped Trentino OSM Ontology.

For the Scraped Trasporti pubblici del Trentino resource, we can map 'bus stop.txt' and 'train stop.txt' files with the stop etype in the Scraped Trasporti pubblici del Trentino Teleology by Karma. For example, Figure 5 shows the 'bus stop.txt' file maps the 'stop' etype with data properties from the Scraped Trasporti pubblici del Trentino Teleology by Karma.

[6] https://usc-isi-i2.github.io/karma/

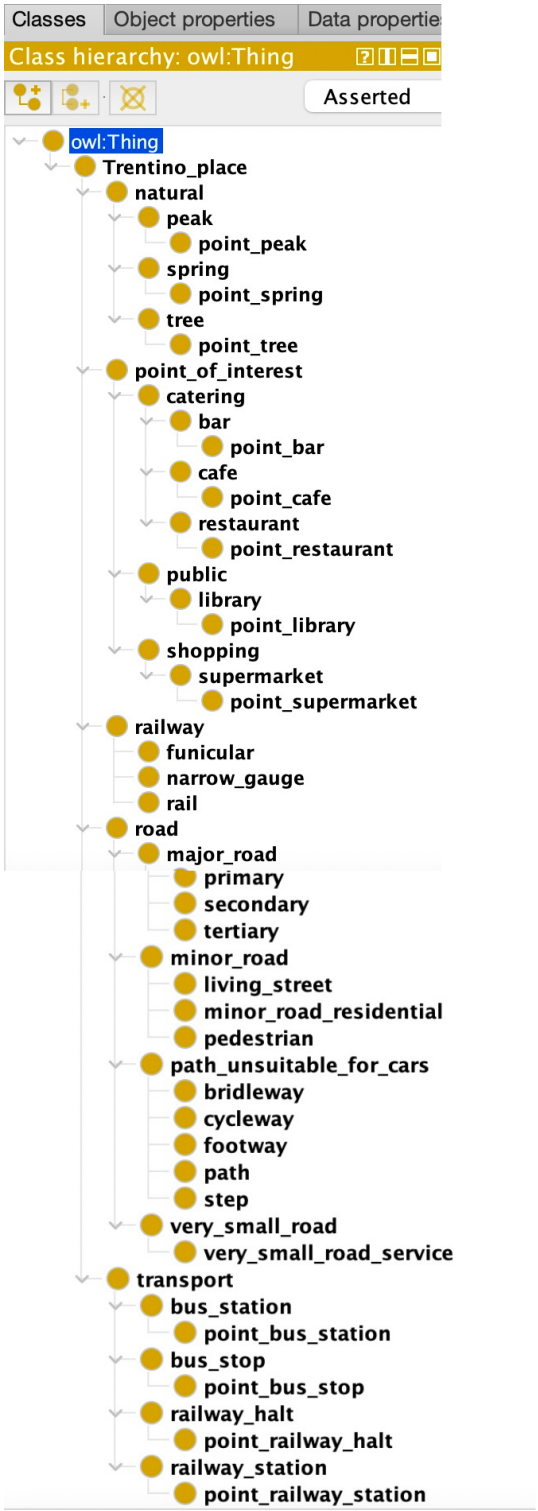
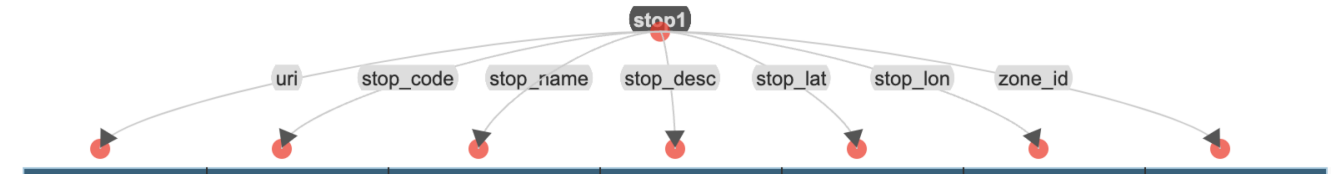


Figure 3: Scraped Trentino OSM Ontology.

gid	osm_id	code	fclass	name	coordinates
17	60959048	5621	bus_stop	Tiers, Brunnenplatz - Tires, Piazza Fontana	POINT(11.5271085 46.4679644)
18	60959057	5621	bus_stop	Tiers, Laurin - Tires, Laurin	POINT(11.5311175 46.4660347)
19	60959116	5621	bus_stop	Plafötsch	POINT(11.5779008 46.4685472)

Figure 4: The 'point_bus_stop.txt' maps the 'point_bus_stop' etype in the Scraped Trentino OSM Ontology.



stop_id	stop_code	stop_name	stop_desc	stop_lat	stop_lon	zone_id
110	1082	Grauno		46.229416	11.29825	1082
124	1081RR	Grumes		46.221406	11.294187	1081
2813	1081FO	Grumes Via Fontanelle		46.219673	11.293091	1081

Figure 5: The 'bus stop.txt' file maps the 'stop' etype with data properties from the Scraped Trasporti pubblici del Trentino Teleology by Karma.

5 Language Definition

In this section, we extracted entity types (etypes), data properties, and object properties from the three sources. We then formalized the language of these concepts for each source by mapping them to the Global Identifiers (GIDs) in the Universal Knowledge Core (UKC). In the UKC, each GID corresponds to a unique definition of a concept. If the UKC returns the GID of a concept, we link the concept with the GID, e.g., bar_GID-14950; otherwise, we assign a new GID to the

concept, ranging from 1000 to 11000. The producer aims to work with the concepts from the three resources that we have already identified in the above section. Meanwhile, the consumer aims to identify new concepts that will be used to integrate these resources.

5.1 Formalize Etypes with Properties from resources by Producer

5.1.1 Formalize Etypes with Properties from Punti di interesse del Trentino Resource

According to the scraped Punti di interesse del Trentino data resource, we extracted entity types with data properties as shown in Table 8. The supermarket etype is created because of the supermarketconvenience store.txt.

Table 8: Etypes with data properties from scraped Punti di interesse del Trentino data.

Properties	Property type	Etypes
id, latitude, longitude, city, street, timetable, Italian-Name, category, topics.	Data property	bar, bus station, cafe, library, restaurant, supermarket, train station.

The outcomes of this language formalization for entity types and data properties are saved in an excel file as shown in Figure 6.

concept labels	description
bar_GID-14950	a room or establishment where alcoholic drinks are served over a counter
bus station_GID-15745	a terminal that serves bus passengers
cafe_GID-15804	a small restaurant where drinks and snacks are sold
library_GID-20054	a building that houses a collection of books and other materials
restaurant_GID-22500	a building where people go to eat
supermarket_GID-24168	a large self-service grocery store selling groceries and dairy products and household goods
train station_GID-22321	terminal where trains load or unload passengers or goods
id_GID-10032	short for identifier
latitude_GID-46263	the angular distance between an imaginary line around a heavenly body parallel to its equator and the equator itself
longitude_GID-46270	the angular distance between a point on any meridian and the prime meridian at Greenwich
city_GID-45969	a large and densely populated urban area; may include several independent administrative districts
street_GID-24034	the part of a thoroughfare between the sidewalks; the part of the thoroughfare on which vehicles travel
timetable_GID-34825	a schedule of times of arrivals and departures
ItalianName_GID-10033	name spelled in italian
category_GID-43482	a collection of things sharing a common attribute
topics_GID-10034	primary purpose or function of a facility

Figure 6: Formalize Concepts from Punti di interesse del Trentino resource by UKC.

5.1.2 Formalize Etypes with Properties from Trasporti pubblici del Trentino Resource

According to the scraped Trasporti pubblici del Trentino data and knowledge resource, we extracted entity types with data properties as shown in Table 9. These data properties are derived from the column names in bus_stops.txt and train_stops.txt, which have been modified for clarity. For instance, 'stop_desc' column name has been changed to 'stop description' data property. Explains of 'stop id', 'stop code', 'stop name', 'stop description', 'stop latitude', 'stop longitude', 'zone id' data properties referring the explains of 'stop_id', 'stop_code', 'stop_name', 'stop_desc', 'stop_lat', 'stop_lon', 'zone_id' in the MITT - OpenData Manual - v.7PDF Popular file.

Table 9: Etypes with data properties from scraped Trasporti pubblici del Trentino data.

Properties	Property type	Etypes
stop id, stop code, stop name, stop description, stop latitude, stop longitude, zone id	Data property	bus stop, train stop

The outcomes of this language formalization for entity types and data properties are saved in an excel file as shown in Figure 7. The definitions of the concepts with new GIDs (10035 - 10042) refer their descriptions in the in MITT - OpenData Manual - v.7PDF file.

concept labels	description
bus stop_GID-45937	a place on a bus route where buses stop to discharge and take on passengers
train stop _GID-10035	a place on a train route where trains stop to discharge and take on passengers
stop id_GID-10036	contains a numerical value that uniquely identifies the stop/station. Several trips can use the same stops
stop code_GID-10037	contains a numerical value that uniquely identifies the stop/station
stop name_GID-10038	contains the name of the stop/station
stop description_GID-10039	contains the description of the stop/station (always empty)
stop latitude_GID-10040	contains the latitude of the stop/station
stop longitude_GID-10041	contains the longitude of the stop/station
zone id_GID-10042	contains information on the zone to which the stop belongs

Figure 7: Formalize Concepts from Trasporti pubblici del Trentino resource by UKC.

5.1.3 Formalize Etypes with Properties from Trentino OSM Resource

According to the Scraped Trentino OSM Ontology, we extracted the etypes: Trentino_place, natural, peak, spring, tree, point_of_interest, catering, bar, cafe, restaurant, public, library, shopping, supermarket, railway, road, major_road, minor_road, path_unsuitable_for_cars, very_small_road, transport, bus_station, bus_stop, railway_halt, railway_station, point_bar, point_bus_station, point_bus_stop, point_cafe, point_library, point_peak, point_railway_halt, point_railway_station, point_restaurant, point_spring, point_supermarket, point_tree, primary, secondary, step, tertiary, very_small_road_service, pedestrian, bridleway, cycleway, footway, living_street, minor_road_residential, path, funicular, narrow_gauge and rail. Based on the scraped Trentino OSM data and our project objective, we extracted the data properties as shown in Table 10. Explains of etypes and data properties can be found in OpenStreetMap Data in Layered GIS Format.

Table 10: Etypes with data properties from scraped Trentino OSM dataset.

Properties	Property type	Etypes
gid, osm_id, code, fclass, name, coordinates	Data property	point_bar, point_bus_station, point_bus_stop, point_cafe, point_library, point_peak, point_railway_halt, point_railway_station, point_restaurant, point_spring, point_supermarket, point_tree, primary, secondary, step, tertiary, very_small_road_service, pedestrian, bridleway, cycleway, footway, living_street, minor_road_residential, path, funicular, narrow_gauge, rail.
layer, bridge, tunnel	Data property	funicular, narrow_gauge, rail, primary, secondary, step, tertiary, very_small_road_service, pedestrian, bridleway, cycleway, footway, living_street, minor_road_residential, path
ref, oneway, maxspeed	Data property	primary, secondary, step, tertiary, very_small_road_service, pedestrian, bridleway, cycleway, footway, living_street, minor_road_residential, path

The outcomes of this language formalization for entity types and data properties are saved in an excel file osm concepts_ukc.xlsx [7], part of the file is as shown in Figure 8. The descriptions of the concepts with new GIDs (10000 - 10030, 10043 - 10057) refer their descriptions in the Scraped Trentino OSM Ontology.

[7] https://docs.google.com/spreadsheets/d/1_kPTaej_BIHm5UFxThtLM5k9pfWldL7R/edit?usp=drive_linkouid=118083013774113966850rtpof=truesd=true

concept labels	description
point_bar_GID-10000	a bar as a point place that offer drinks.
point_bus_station_GID-10001	a large bus station as a point place with multiple platforms.
point_bus_stop_GID-10002	a bus stop as a point place.
point_cafe_GID-10003	a cafe as a point place.
point_library_GID-10004	a library as a point place.
point_peak_GID-10005	a mountain peak as a point place.
point_railway_halt_GID-10006	a smaller, local railway station, or subway station as a point place.
point_railway_station_GID-10007	a larger railway station as a point place of mainline rail services.
point_restaurant_GID-10008	a normal restaurant as a point place.
point_spring_GID-10009	a spring as a point place, possibly source of a stream.
point_supermarket_GID-10010	a supermarket as a point place.
point_tree_GID-10011	a tree as a point place.
primary_GID-10012	a primary road, typically national.
secondary_GID-10013	a secondary road, typically regional.
step_GID-23917	support consisting of a place to rest the foot while ascending or descending a stairway
tertiary_GID-10014	a tertiary road, typically local.
very_small_road_service_GID-10015	a service road for access to buildings, parking lots, etc.
pedestrian_GID-10016	a pedestrian only street.
bridleway_GID-10017	a path for horse riding.
cycleway_GID-10018	a path for cycling.
footway_GID-10019	a footpath.
living_street_GID-10020	a street where pedestrians have priority.
minor_road_residential_GID-10021	a road in residential areas.
path_GID-10022	an unspecified path.
funicular_GID-15798	a railway up the side of a mountain pulled by a moving cable and having counterbalancing ascending and descending cars
narrow_gauge_GID-10023	a narrow gauge railway track.
rail_GID-22309	short for railway
gid_GID-10024	unique identifier of places with a feature in OSM dataset.
osm_id_GID-10025	the OSM Id taken from the Id of this feature (node_id, way_id, or relation_id) in the OSM database.
code_GID-35741	a set of rules or principles or laws (especially written ones)
fclass_GID-10026	class name of places in OSM dataset.
name_GID-2	a language unit by which a person or thing is known
coordinates_GID-10027	readable coordinates.
layer_GID-33577	an abstract place usually conceived as having depth
bridge_GID-15577	a structure that allows people or vehicles to cross an obstacle such as a river or canal or railway etc.
tunnel_GID-24989	a passageway through or under something, usually underground (especially one for trains or cars)
ref_GID-10028	reference number of this road ('A 5', 'L 605', ...)
oneway_GID-10029	Is this a oneway road? "F" means that only driving in direction of the linestring is allowed. "T" means that only the opposite direction is allowed. "B" (c
maxspeed_GID-10030	max allowed speed in km/h
Trentino_place_GID-10043	A place from the Trentino OpenStreetMap, Punti di interesse del Trentino and Trasporti pubblici del Trentino (formato GTFS).
natural_GID-10044	A place with natural features.
peak_GID-46388	the top point of a mountain or hill
spring_GID-50318	a natural flow of ground water
tree_GID-69557	a tall perennial woody plant having a main trunk and branches forming a distinct elevated crown; includes both gymnosperms and angiosperms
point_of_interest_GID-10045	A place as point of Interest.
catering_GID-10046	A point of interest about catering services.
bar_GID-14950	a room or establishment where alcoholic drinks are served over a counter
cafe_GID-15804	a small restaurant where drinks and snacks are sold
restaurant_GID-22500	a building where people go to eat
public_GID-10047	A public point of interest.
library_GID-20054	a building that houses a collection of books and other materials
shopping_GID-387	searching for or buying goods or services
supermarket_GID-24168	a large self-service grocery store selling groceries and dairy products and household goods
railway_GID-10048	A line place such as railway, subway, light rail, tram, ...
road_GID-22592	an open way (generally public) for travel or transportation
major_road_GID-10049	A major road.
minor_road_GID-10050	A minor road.
path_unsuitable_for_cars_GID-10051	A path is unsuitable for cars.
very_small_road_GID-10052	A very small road.
transport_GID-10053	A transport related place, such as parking lot, petrol (gas) station, ...
bus_station_GID-10054	A large bus station with multiple platforms.
bus_stop_GID-10055	A bus stop
railway_halt_GID-10056	A smaller, local railway station, or subway station.
railway_station_GID-10057	A larger railway station of mainline rail services.

Figure 8: Formalize Concepts from Trentino OSM resource by UKC.

5.2 Formalize Object Properties for Composing Resources from Consumer

Consumer generates the two object properties, i.e., 'near' and 'on', as shown in Table 11, which can compose the etypes from the three resources. An object property relates from a source etype to a target etype, such as the 'point_bar' near the 'point_peak', the 'bar' on the 'pedestrian'.

The outcomes of object property language formalization are saved in an excel file, which is as shown in Figure 9.

Table 11: Formalize Object Properties for Composing Resources.

Properties	Property type	Etypes
near	Object property	Source/Traget etypes: bar, bus station, cafe, library, restaurant, supermarket, train station, bus stop, train stop, point_bar, point_bus_station, point_bus_stop, point_cafe, point_library, point_peak, point_railway_halt, point_railway_station, point_restaurant, point_spring, point_supermarket, point_tree.
on	Object property	*Source etypes*: bar, bus station, cafe, library, restaurant, supermarket, train station, bus stop, train stop, point_bar, point_bus_station, point_bus_stop, point_cafe, point_library, point_peak, point_railway_halt, point_railway_station, point_restaurant, point_spring, point_supermarket, point_tree. *Target etypes*: primary, secondary, step, tertiary, very_small_road_service, pedestrian, bridleway, cycleway, footway, living_street, minor_road_residential, path, funicular, narrow_gauge, rail

concept labels	description
near_GID-84218	not far distant in time or space or degree or circumstances
on_GID-10031	physically in contact with and supported by (a surface).

Figure 9: Formalize Object Properties with UKC for Composing Resources.

6 Knowledge Definition

In this section, the producer aims to generate teleologies for each dataset, always considering the reusability of knowledge enhancement. Meanwhile, the consumer aims to integrate the teleologies from the producer, specifically to generate the Trentino Spatial Teleology and Trentino Spatial Teleontology.

6.1 Teleologies from Producer

The producer has developed three distinct teleologies: Trentino OSM Teleology, Trentino POI Teleology, and Trentino GTFS Teleology, representing the formalized etypes, data properties and object properties as outlined in Table 8, Table 9 and Table 10 respectively. These three teleologies were designed using Protege and are stored as .owl files. The Trentino OSM Teleology [8], Trentino POI Teleology [9] and Trentino GTFS Teleology [10] have been uploaded online.

Figure 10 depicts the Trentino GTFS Teleology in Protege, which includes two etypes: ‘bus stop_GID-45937’ and ‘train stop_GID-10035’. Each of them has seven data properties: ‘stop

[8] https://drive.google.com/file/d/1_L9WzplbkjhJWvBJ_q3ZFYtYbma4ou-5/view?usp=drive_link
[9] https://drive.google.com/file/d/1e3f5k5xCUyZljZ5YfeMzdkkAAA6lv1Aj/view?usp=drive_link
[10] https://drive.google.com/file/d/1MkyI6lelaQJSVOQViZULbi5Qg7WIHxJI/view?usp=drive_link

id_GID-10036', 'stop code_GID-10037', 'stop name_GID-10038', 'stop description_GID-10039', 'stop latitude_GID-10040', 'stop longitude_GID-10041' and 'zone id_GID-10042'. An object property 'near_GID-84218' establishes a relationship between the two etypes. The data properties in Trentino GTFS Teleology reused the GTFS schema from the MITT - OpenData Manual - v.7PDF PDF Popular file. And the etypes and object property in Trentino GTFS Teleology are developed by considering our project objective.

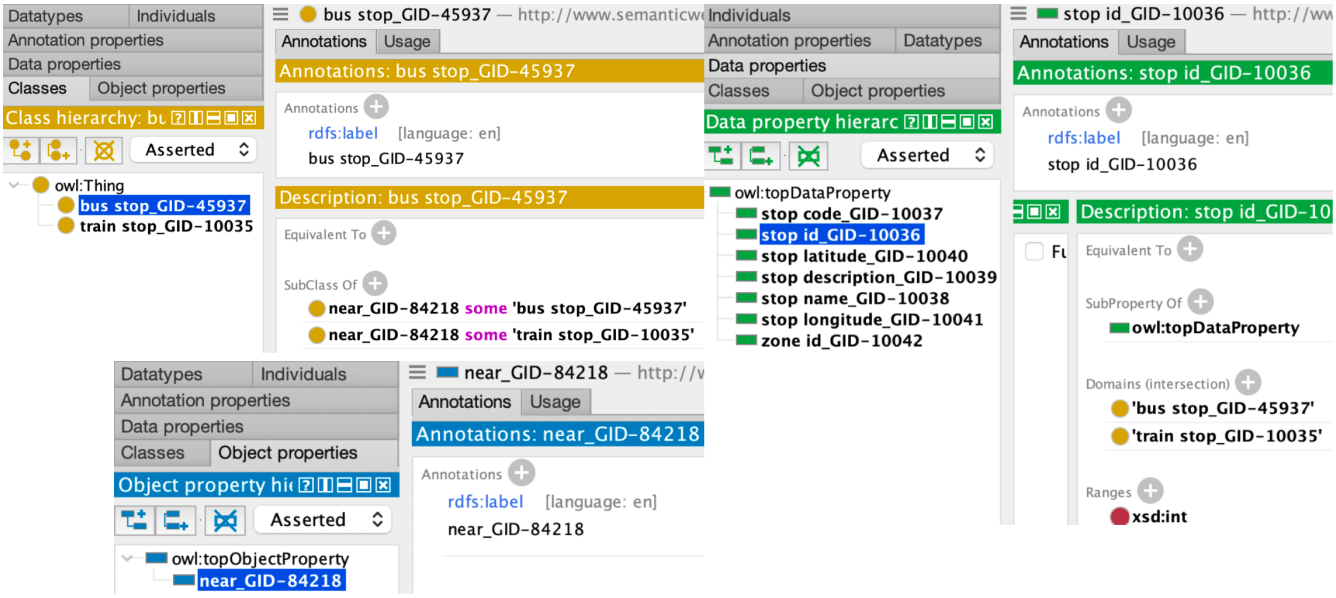


Figure 10: Trentino GTFS Teleology in Protege.

Figure 11 visualizes the Trentino OSM Teleologyin Protege. In the Trentino OSM Teleology, etypes are definied by reusing the Trentino OSM LWontology, the data properties are defined by reusing the information from the OpenStreetMap Data in Layered GIS Format file. The object properties are developed by considering our project objective.

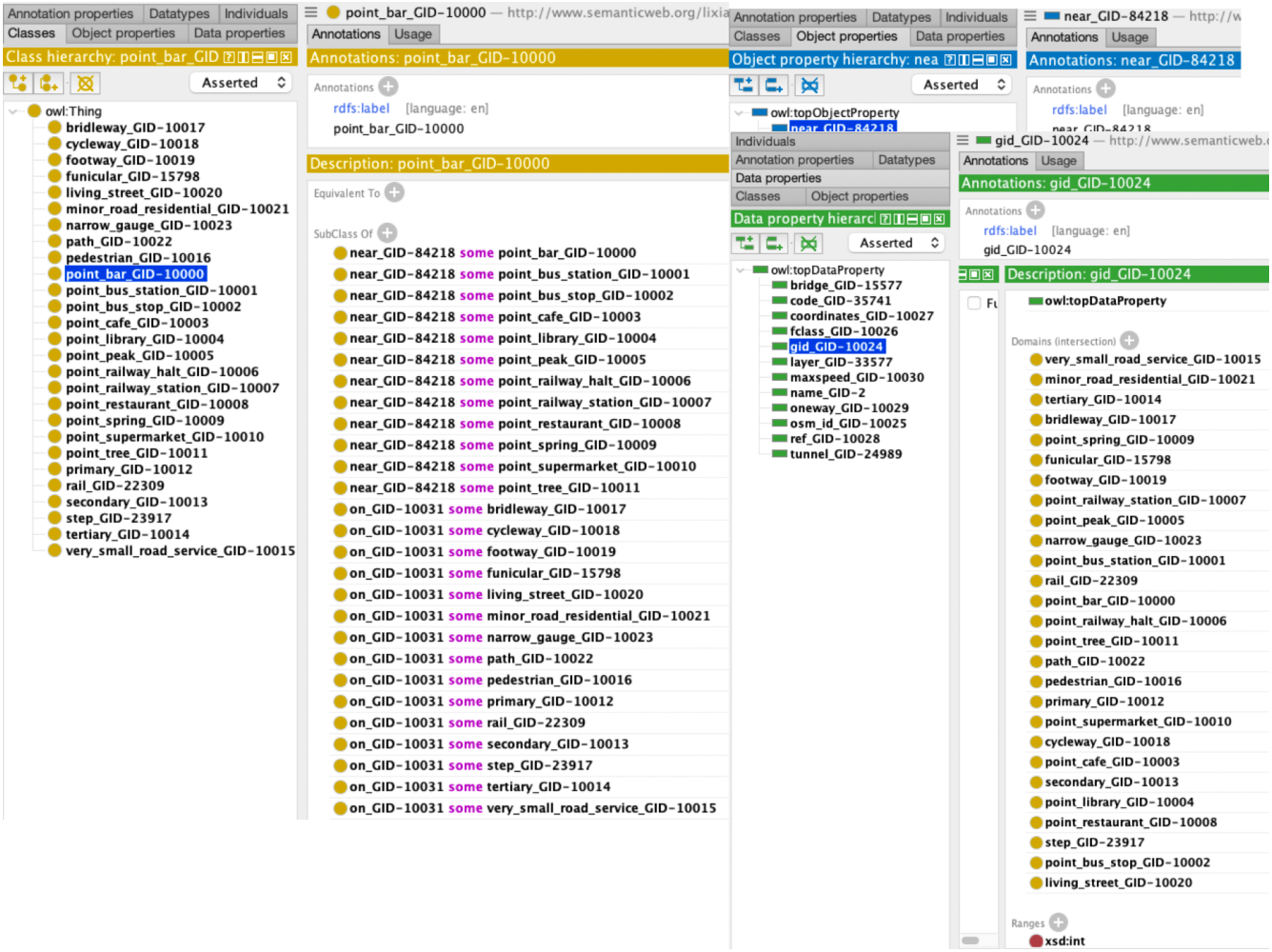


Figure 11: Trentino OSM Teleology in Protege.

Figure 12 visualizes the Trentino POI Teleology in Protege. The etypes are and data properties are defined by considering the data structure of the scraped data from fromPunti di interesse del Trentino resource. The object properties are developed by considering our project objective.

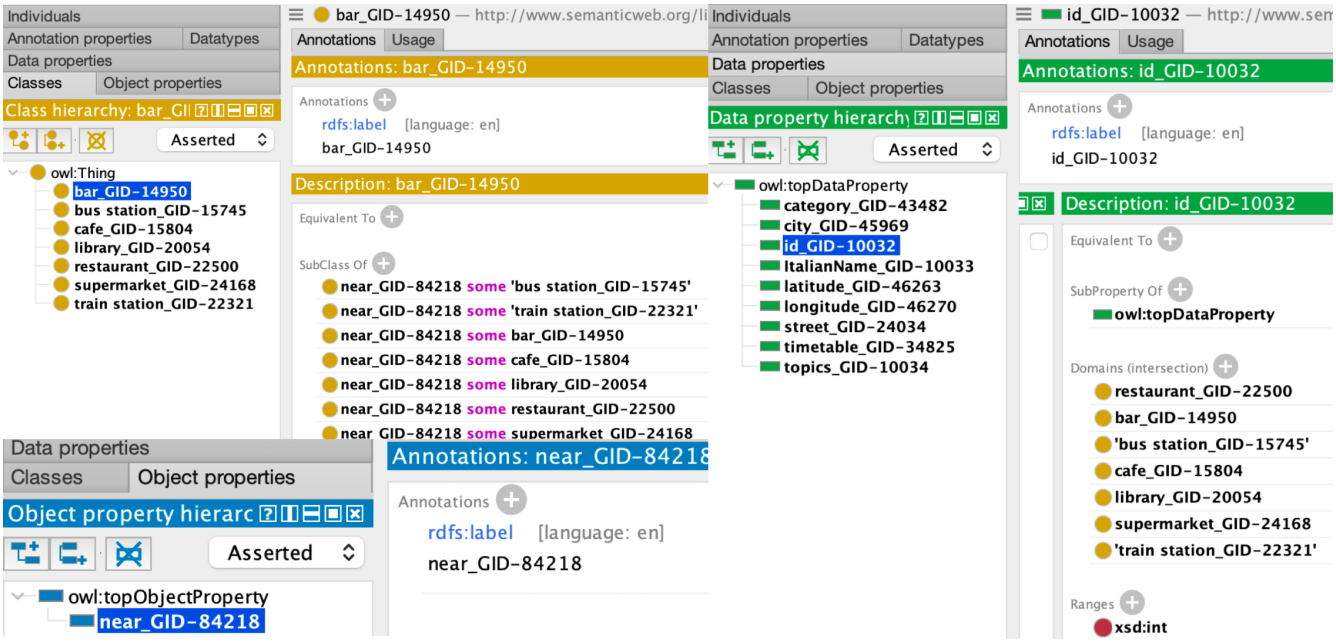


Figure 12: Trentino POI Teleology in Protege.

6.2 Composed Trentino Spatial Teleology and Teleontology from Consumer

6.2.1 Trentino Spatial Teleology

The consumer generated the Trentino Spatial Teleology by composing the Trentino OSM Teleology, Trentino POI Teleology, Trentino GTFS Teleology from producer. This composition was performed with the objective of maximizing reuse and aligning with the project goals. The etypes of the Trentino Spatial Teleology are a union of the etypes present in the three individual teleologies. During the composition process, redundant etypes and data properties were eliminated, the object properties are generating by assuming all the natural places and facilities are located on transportation ways. For instance, the 'point_bar_GID-10000' Trentino OSM Teleology and 'bar_GID-14950' in the Trentino POI Teleology represent the same etype for bar entities. In the Trentino OSM Teleology, 'point_bar_GID-10000' has data properties: 'gid_GID-10024', 'osm_id_GID-10025', 'code_GID-35741', 'fclass_GID-10026', 'name_GID-2' and 'coordinates_GID-10027'. The 'bar_GID-14950' in the Trentino POI Teleology has different data properties: 'id_GID-10032', 'latitude_GID-46263', 'longitude_GID-46270', 'city_GID-45969', 'street_GID-24034', 'timetable_GID-34825', 'ItalianName_GID-10033', 'category_GID-43482' and 'topics_GID-10034'. We reuse the schema of Trentino OSM LWOntology as much as possible. Hence, in the final Trentino Spatial Teleology, we have a single etype named 'point_bar_GID-10000', which includes data properties from all data properties of the two etype. Due to 'latitude_GID-46263' and 'longitude_GID-46270' were excluded as they duplicate information provided by 'coordinates_GID-10027'; 'ItalianName_GID-10033' was excluded as it duplicates information provided by 'name_GID-2'; 'category_GID-43482' was excluded as it duplicates information provided by 'fclass_GID-10026'; 'id_GID-10032' was excluded as it duplicates information provided by gid_GID-10024'. Hence in the final Trentino

Spatial Teleology, we have a single etype named ‘point_bar_GID-10000’, its data properties are ‘gid_GID-10024’, ‘osm_id_GID-10025’, ‘code_GID-35741’, ‘fclass_GID-10026’, ‘name_GID-2’, ‘coordinates_GID-10027’, ‘city_GID-45969’, ‘street_GID-24034’, ‘timetable_GID-34825’, and ‘topics_GID-10034’. The Trentino Spatial Teleology was designed by protege, which can be downloaded online [11]. we illustrate it in this report as shown in Figure 13.

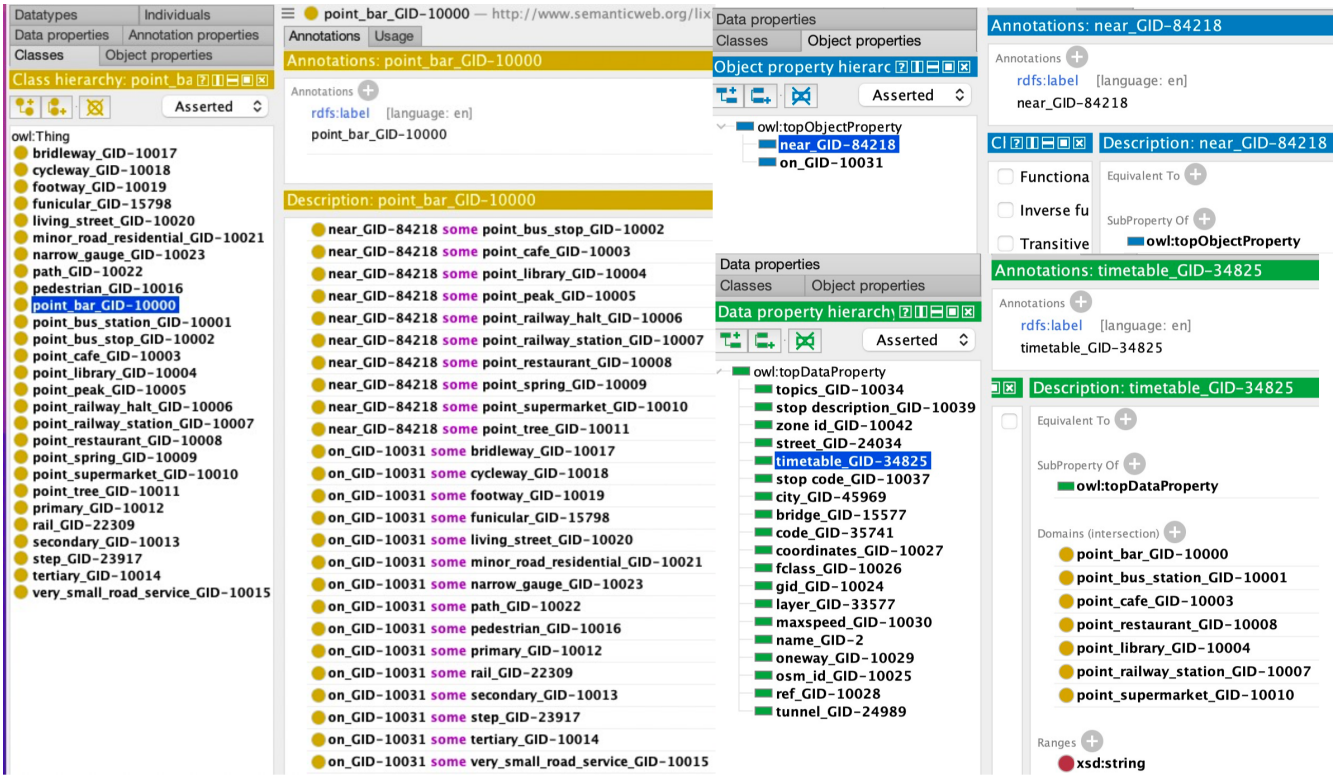


Figure 13: Trentino Spatial Teleology in Protege.

6.2.2 Trentino Spatial Teleontology

To build Trentino Spatial Ontology, consumer needs to compose Trentino OSM Ontology as shown in the Figure 3 in Section 4.2.2, Trentino POI Ontology and Trentino GTFS Ontology. We noticed that the classes of Trentino POI Ontology and Trentino GTFS Ontology have been recorded in the Trentino OSM Ontology. For example, Trentino GTFS Ontology contains bus ‘stop_GID-45937’ and ‘train stop_GID-10035’ classes, we view them are same as ‘point_bus_stop_GID-10002’ and ‘point_railway_halt_GID-10006’ in Trentino OSM Ontology. Hence, we view Trentino Spatial Ontology is the Trentino OSM Ontology where the etypes are linking with GIDs.

The Trentino Spatial Teleontology was constructed based on the Trentino Spatial Teleology and the Trentino Spatial Ontology, as illustrated in Figure 14. In the Trentino Spatial Teleontology, black arrows represent the ISA relations. The source etype of the arrow is a subclass, while the target etype of the arrow is a superclass. For instance, ‘railway’ is a

[11] https://drive.google.com/file/d/1PsU-kNDTvpq2FiWfAtlHvYns03S89z-O/view?usp=drive_link

'Trentino_place', or in other words, 'railway' is a subclass of 'Trentino_place'. Subclasses inherit the data properties and object properties of their superclass. Therefore, since 'railway' is a subclass of 'Trentino_place', 'railway' possesses the data properties of 'Trentino_place', namely, gid, osm_id, name, fclass, code, and coordinates. In the Trentino Spatial Teleontology, the etypes represented by purple nodes are used to denote natural places and facilities. On the other hand, the etypes represented by blue nodes are used to denote transportation ways. Object properties were generated based on the assumption that natural places and facilities are near each other and they are located on transportation ways. The .owl file for the Trentino

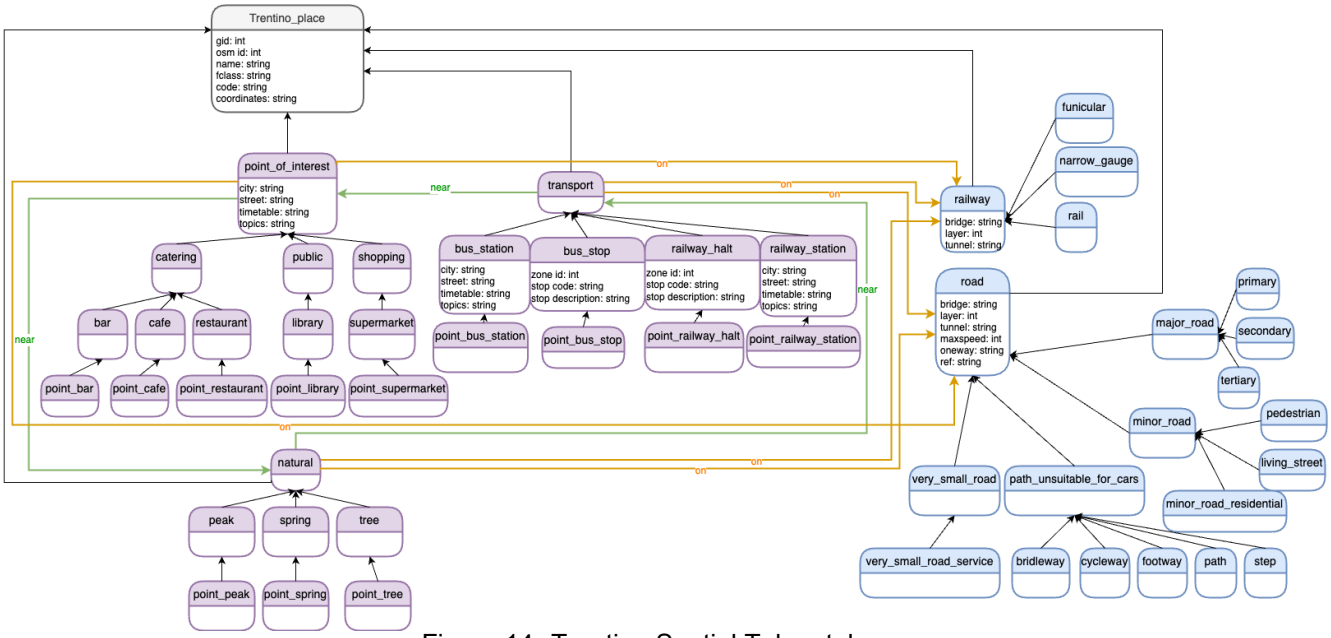


Figure 14: Trentino Spatial Teleontology.

Spatial Teleontology can be accessed and downloaded from the provided link [12], in this file each concepts are added with GIDs. Figure 15 shows the etypes with properties of the Trentino Spatial Teleontology in protege.

[12] https://drive.google.com/file/d/1uJGyxA9aUEfFzBhOmA5IPdmw21-RRC7S/view?usp=drive_link

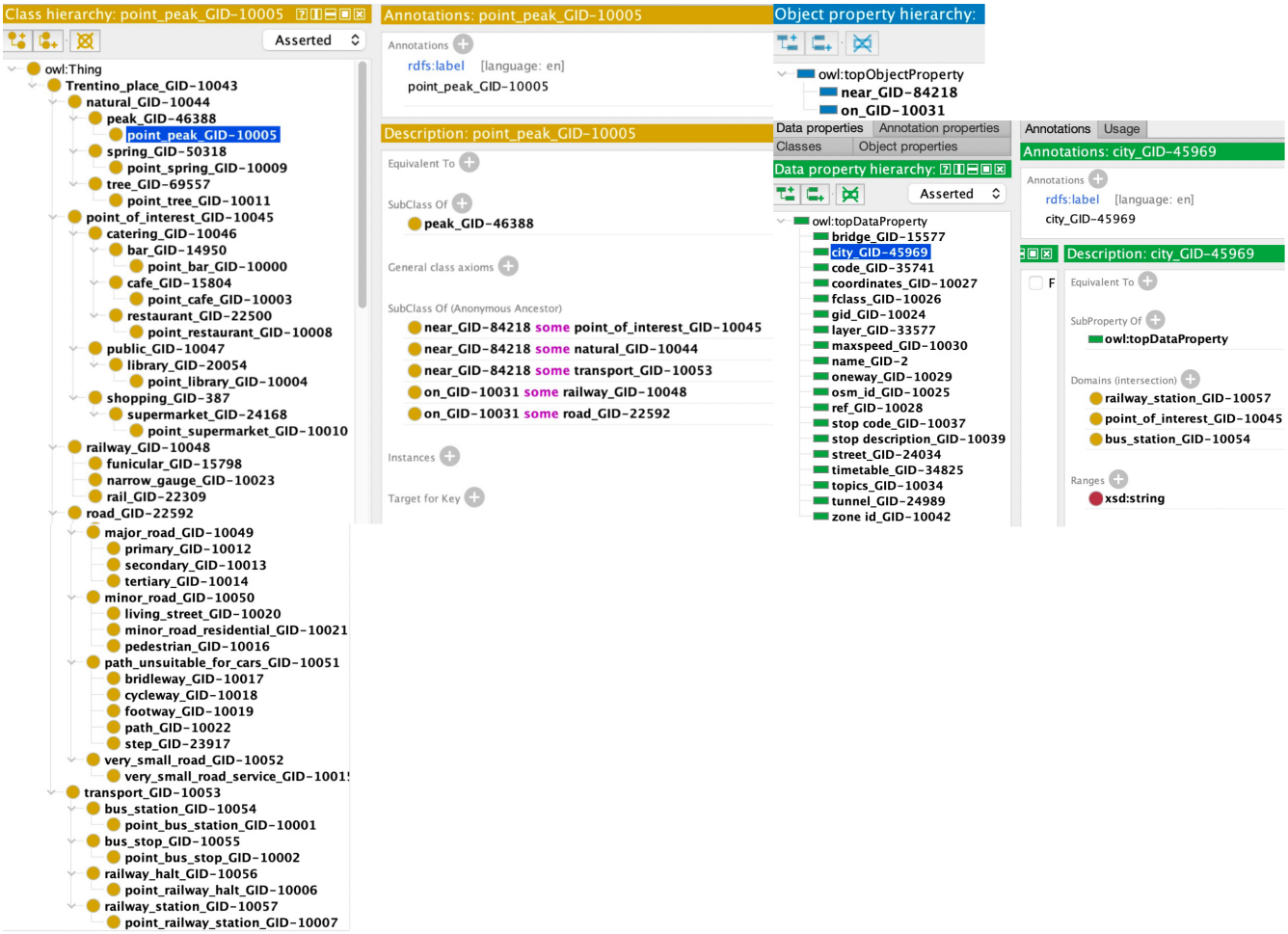


Figure 15: Trentino Spatial Teleontology in protege.

7 Data Definition

In this section, the producer aims to formalize each dataset and map each dataset to their respective schema. Meanwhile, the consumer aims to compose all datasets and merge the composed data with the final teleontology.

7.1 Dataset formatting by Consumer

Consumer formalized data to instantiate the entity types, data properties, and object properties as defined in the Trentino Spatial Teleontology. To distinguish between data files derived from different sources, we have incorporated provenance information for each file. This applies to a variety of files, including but not limited to osm_primary.txt, osm_point_bus_stop.txt, poi_point_cafe.txt, and gtfs_point_bus stop.txt.

Files are integrated if they contain instances of the same entity type. For instance, both osm_point_bus_stop.txt and gtfs_bus stop.txt contain entities of the bus stop entity type. During this integration process, the data properties within the integrated files are also amalgamated.

This was accomplished by formalizing the data properties for each entity type based on the guidelines provided by the Trentino Spatial Teleontology. The files that need to require integrate, along with the integrated files and their respective properties, are listed in Table 12.

Table 12: Integrate files from different sources

files	integrated files	properties of the integrated files
osm_point_bus_stop.txt, gtfs_bus stop.txt	osm_gtfs_point_bus_stop.txt	gid, osm_id, code, fclass, name, coordinates, stop code, stop description, zone id
osm_point_railway_halt.txt, gtfs_train stop.txt	osm_gtfs_point_railway_halt.txt	gid, osm_id, code, fclass, name, coordinates, stop code, stop description, zone id
osm_point_bar.txt, poi_bar.txt	osm_poi_point_bar.txt	gid, osm_id, code, fclass, name, coordinates, city, street, timetable, topics
osm_point_bus_station.txt, poi_bus station.txt	osm_poi_point_bus_station.txt	gid, osm_id, code, fclass, name, coordinates, city, street, timetable, topics
osm_point_cafe.txt, poi_cafe.txt	osm_poi_point_cafe.txt	gid, osm_id, code, fclass, name, coordinates, city, street, timetable, topics
osm_point_library.txt, poi_library.txt	osm_poi_point_library.txt	gid, osm_id, code, fclass, name, coordinates, city, street, timetable, topics
osm_point_railway_station.txt, poi_railway station.txt	osm_poi_point_railway_station.txt	gid, osm_id, code, fclass, name, coordinates, city, street, timetable, topics
osm_point_restaurant.txt, poi_restaurant.txt	osm_poi_point_restaurant.txt	gid, osm_id, code, fclass, name, coordinates, city, street, timetable, topics
osm_point_supermarket.txt, poi_supermarket.txt	osm_poi_point_supermarket.txt	gid, osm_id, code, fclass, name, coordinates, city, street, timetable, topics

For example, the tables osm_point_bus_stop.txt and gtfs_bus stop.txt are merged into a unified table named osm_gtfs_point_bus_stop.txt. This consolidated table encompasses entities of the bus stop entity type as delineated in the Trentino Spatial Teleontology. The data properties encapsulated in the new file include ‘gid’, ‘osm_id’, ‘code’, ‘fclass’, ‘name’, ‘coordinates’, ‘stop code’, ‘stop description’, and ‘zone id’. In this file, entities originating from osm_point_bus stop.txt have null values for ‘stop code’, ‘stop description’, and ‘zone id’. Conversely, entities from gtfs_bus stop.txt have their ‘gid’ values assigned as integer numbers that are distinct from the ‘gid’ values in osm_point_bus_stop.txt; their ‘osm_id’ is set to null; their ‘code’ and ‘fclass’ values are same as those of the entities in osm_point_bus_stop.txt; their ‘coordinates’ are formatted ‘latitude’ and ‘longitude’ values derived from gtfs_bus stop.txt, matching the ‘coordinates’ format of osm_point_bus_stop.txt (e.g., POINT(46.229416 11.29825)); their ‘name’, ‘stop code’, ‘stop description’, and ‘zone id’ values correspond to the ‘stop name’, ‘stop code’, ‘stop description’, and ‘zone id’ values from gtfs_bus stop.txt.

7.2 Entity identification by Producer

For the integrated files derived from various sources, i.e., osm_poi_point_bar.txt, osm_poi_point_bus_station.txt, osm_poi_point_cafe.txt, osm_poi_point_library.txt, osm_poi_point_railway_station.txt, osm_poi_point_restaurant.txt, osm_poi_point_supermarket.txt, osm_gtfs_point_bus_stop.txt, and osm_gtfs_point_railway_halt.txt, each file represents place instances of a specific entity type. Within each file, producer identifies identical entities by comparing their names and coordinates and eliminate any duplicates. In essence, for each

file, we have an *Entity Identifying Set* comprising of the data properties 'name', 'altitude', and 'longitude'. These are used to ascertain whether two entities within a file are identical. We operate under the assumption that if two entities have the same name and their haversine distance is less than 50 meters, they are considered to be the same entity. Consequently, only one instance of the entity is retained in the file.

7.3 Object property 'On' identification by Consumer

To recognize natural place and facility entities are 'On' which transportation ways. We created a set of new files linking natural place and facility entities with transportation way entities by *horizontal composition* as mentioned in Section 2.2.1. The method is first to construct R-tree index for the transportation ways, and for all facilities and natural places by creating new linking tables if their minimal haversine distance (the distance between two points on the earth surface based on their longitude and latitude.) is less than 50 meters. For example, as shown in Figure 16, we have a new table named osm_poi_point_bar_onWays.txt, which has the property 'gid', it is a global/unique identifier in a table for the linked place, i.e., bar identifier, and a set of transportation ways' identifiers: 'osm_funicular_gid', 'osm_narrow_gauge_gid', 'osm_rail_gid', 'osm_primary_gid', 'osm_secondary_gid', 'osm_tertiary_gid', 'osm_living_street_gid', 'osm_minor_road_residential_gid', 'osm_pedestrian_gid', 'osm_bridleway_gid', osm_cycleway_gid', 'osm_footway_gid', osm_path_gid', 'osm_step_gid', 'osm_very_small_road_service_gid'. Each linking transportation way are the nearest way and not far way 50 meters from the linked place.

gid	osm_funicular_id	osm_narrow_gauge_id	osm_rail_id	osm_primc	osm_second	osm_tertia	osm_living_stre	osm_mino	osm_ped	osm_bridlew	osm_cyclew	osm_footw	osm_pat	osm_ste	osm_very_small_road_service_id
17					515384							1244254	1244255	1244256	1244258
18					1016905										744821
19					713911										
20					746791										276497
21															
22													24832		1050986
24					580461								760197		
31				1175681				572720							

Figure 16: Part data in osm_poi_point_bar_onWays.txt

7.4 Data mapping

We uploaded the integrated data in the Final Integrated Trentino Spatial Data [13]. The final integrated data for Trentino comprises 12 txt files, 15 txt files, and 12 csv files. The txt files encompass information about natural places, facilities, and transportation ways. The csv files illustrate which transportation ways the natural places and facilities are located on.

We have generated a set of Entity Graphs (EGs), which can be acquired by linking the integrated data with the integrated Trentino Spatial Teleontology. These EGs are categorized into various types. The first category of EGs pertains to facility entities that have been integrated using files from both Trentino OSM (Trentino OpenStreetMap) resources and Trentino GTFS (Trasporti pubblici del Trentino) resource. Examples of this include entities from osm_gtfs_bus_stop.txt and osm_gtfs_point_railway_halt.txt. Figure 17 shows the

[13] https://drive.google.com/drive/folders/1QOMrkap5jDG4nAu1j5W3OsI63xuBNtUP?usp=drive_link

osm_gtfs_point_railway_halt.txt maps the point_railway_halt etype in Trentino Spatial Teleontology by karma. The second type of EGs involves facility entities that have been integrated us-

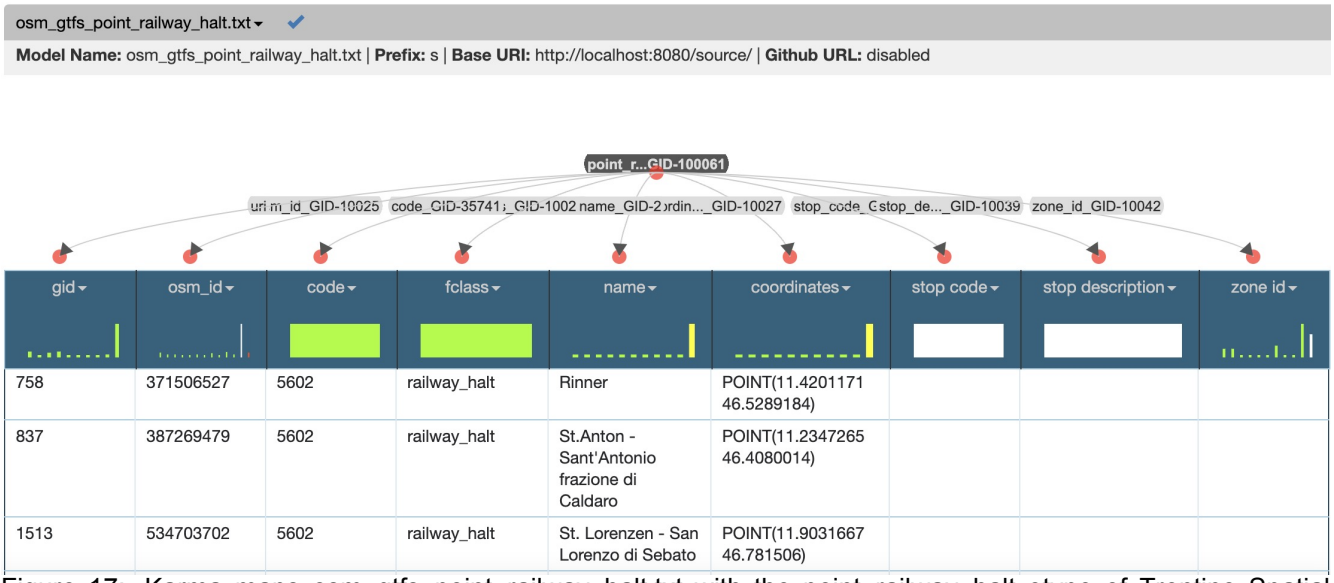


Figure 17: Karma maps osm_gtfs_point_railway_halt.txt with the point_railway_halt etype of Trentino Spatial Teleontology.

ing files from Trentino OSM and Trentino POI (Punti di interesse del Trentino) resources, such as osm_poi_bar.txt. The third type of EGs includes facilities derived solely from OSM resources that have not been integrated, for instance, osm_poi_point_tree.txt. The forth and fifth are about railway and road entities from Trentino OSM resource, e.g., osm_narrow_gauge.txt and osm_living_street.txt. The remaining types of EGs encompass the aforementioned entities linked with ways, such as osm_gtfs_point_bus_stop_onWays.csv, osm_point_tree_onWays.csv. Figure 18 shows the the mapping Trentino Spatial Teleontology with the osm_poi_point_cafe_onWays.csv by karma.

8 Outcome Exploitation

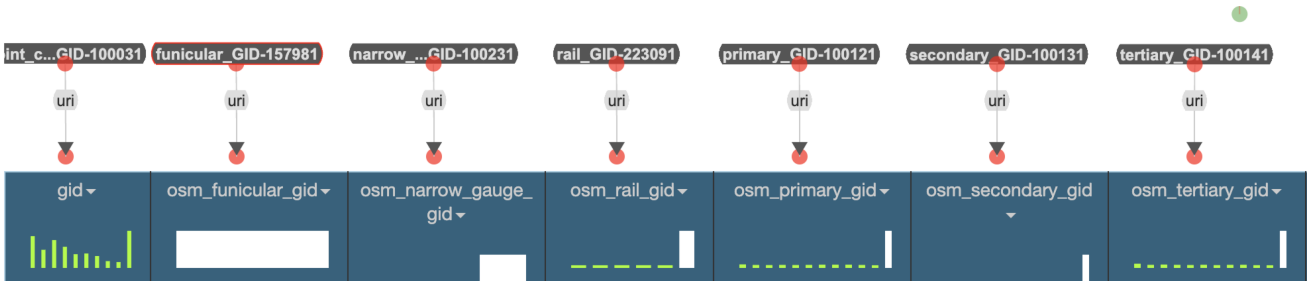
8.1 Evaluation

We evaluate the quality of our knowledge and data results by Coverage and Coverage metrics.

8.1.1 Coverage metric

- The Entity-Relationship (ER) model comprises 6 entity types, while the Trentino Spatial Teleontology encompasses a significantly larger number of entity types, totaling 52. When it comes to the coverage of entity types in the teleontology with respect to the Competency Questions (CQs), we observe a complete coverage:

$$Cov_E(CQ_E) = \frac{|CQ_E \cap ETG_E|}{CQ_E} = \frac{6}{6} = 100\% \quad (1)$$



gid	osm_funicular_gid	osm_narrow_gauge_gid	osm_rail_gid	osm_primary_gid	osm_secondary_gid	osm_tertiary_gid
26948						612760.0
208087						
88406						44447.0
208070						
11474						
208106						
1231					775421.0	
208103						
83528						1067855.0
28491				1014434.0		
472					253727.0	

Figure 18: Karma maps Trentino Spatial Teleontology with osm_poi_point_cafe_onWays.txt.

- The Entity-Relationship (ER) model consists of 14 data properties and object properties, while the Trentino Spatial Teleontology includes a total of 21 such properties. When evaluating the coverage of properties in the teleontology with respect to the Competency Questions (CQs), we observe a complete coverage:

$$Cov_P(CQ_P) = \frac{|CQ_P \cap ETG_P|}{CQ_P} = \frac{14}{14} = 100\% \tag{2}$$

- The Trentino Spatial Teleontology comprises 52 entity types, while the Trentino OSM LWontology and Trentino GTFS Knowledge contain 791 and 25 entity types respectively. When assessing the coverage of entity types in the teleontology with respect to the Reference Ontologies (RQs), we find that:

$$Cov_E(RQ_E) = \frac{|RQ_E \cap ETG_E|}{RQ_E} = \frac{52}{791 + 25} \approx 30.64\% \tag{3}$$

- The Trentino Spatial Teleontology comprises 21 data properties and object properties, while the Trentino GTFS Knowledge contains 63 such properties. The Trentino OSM LWOntology does not contain any of these properties. When evaluating the coverage of properties in the teleontology with respect to the Reference Ontologies (RQs), we find that:

$$Cov_P(RQ_P) = \frac{|RQ_P \cap ETG_P|}{RQ_P} = \frac{21}{63} \approx 33.33\% \tag{4}$$

The relatively low values of $Cov_E(RQ_E)$ and $Cov_P(RQ_P)$ can be attributed to a couple of factors. Firstly, the Trentino OSM LWontology covers a broad range of place types, while

our focus is primarily on natural locales, facilities, and transportation ways. Secondly, there is a divergence in research focus between our project and the Trentino GTFS research. Our project emphasizes geographical information, whereas the Trentino GTFS research is more concerned with bus schedule information.

8.1.2 Connectivity metric

The N is the total number of etypes in the Trentino Spatial Teleontology.

- Entity connectivity: The number of entities E(T) for each etype T in the KG, $\sum_{K=1}^{N} E(T_K)$ is 153,811.
- Object property connectivity: The number of object property values not null $Op(T)$ for each etype T in the KG, $\sum_{K=1}^{N} Op(T_K)$ is 49,828.
- Data property connectivity: The number of data property values not null $Dp(T)$ for each etype T in the KG, $\sum_{K=1}^{N} Dp(T_K)$ is 1,417,171.

8.2 KG exploitation

To evaluate the effectiveness of the Trentino Spatial Teleontology we developed, and the data we compiled in addressing Competency Questions (CQs), we employed SparQL queries via the GraphDB tool. Here, we present an example that demonstrates how to answer CQ1.

In CQ1, we consider a scenario where a person named Lily resides in Obergummer, Rovereto. The latitude and longitude of her home are 45°53′16.44″ and 11°01′59.88″, respectively. Our objective is to identify the three nearest peaks surrounding her home. We accomplish this by utilizing SparQL queries. The querying code and the results are depicted in the Figure 19.

```
PREFIX ont: <http://www.semanticweb.org/lixiaoyue/ontologies/2023/9/untitled-ontology-336#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT ?name ?code ?osm_id ?fclass ?coordinates ?distanceSquared
WHERE {
  ?peak a ont:point_peak_GID-10005 ;
        ont:name_GID-2 ?name ; ont:code_GID-35741 ?code ;
        ont:osm_id_GID-10025 ?osm_id ; ont:fclass_GID-10026 ?fclass ;
        ont:coordinates_GID-10027 ?coordinates .
  BIND(STRAFTER(STRBEFORE(?coordinates, " "), "POINT(") AS ?lon)
  BIND(STRAFTER(?coordinates, " ") AS ?lat1)
  BIND(STRBEFORE(?lat1, ")") AS ?lat)
  BIND((xsd:double(?lon) - 11.015988)*(xsd:double(?lon) - 11.015988) + (xsd:double(?lat) -
45.531644)*(xsd:double(?lat) - 45.531644) AS ?distanceSquared)}
ORDER BY ?distanceSquared
LIMIT 3
```

	name	code	osm_id	fclass	coordinates	distanceSquared
1	"Cima Campiglia"	"4111"	"9247961622"	"peak"	"POINT(11.2188455 45.787914)"	"0.10682547820625049"^^xsd:double
2	"Fratòn di Sorapache"	"4111"	"9155906740"	"peak"	"POINT(11.214568 45.7922899)"	"0.10737030158680996"^^xsd:double
3	"Punta del Vècio"	"4111"	"9155852694"	"peak"	"POINT(11.2167084 45.7932699)"	"0.10873679052696913"^^xsd:double

Figure 19: Result of the SparQL query for CQ 1

9 Conclusion and Open Issues

This document adheres to the iTelos methodology process to compose the three spatial resources for generating the Trentino spatial reference. Throughout this process, we discover some relevant open issues.

Firstly, the resources we chosen recorded the place information in different time, i.e., Punti di interesse del Trentino encompasses places as of April 2014, Trasporti pubblici del Trentino includes places from the June to Septemeber of 2023, Trentino OSM resource contains places in the June of 2023. However, it is important to note that places recorded in 2014 may have undergone changes by 2023. Second issue is about the usage of protege. The knowledge integration process in Protege is not automated. For example, when integrating the three resources' teleologies into the Trentino spatial teleology, we had to manually integrate entity types and properties. Additionally, when creating the classes and properties for schemas, it is advisable to add a label for each of them with the label value identical to the name of the entity type or property. Otherwise, there may be unexpected changes to the classes/properties. The reason for this remains unclear.

Despite the aforementioned challenges, the iTelos methodology process indeed offers a comprehensive framework for the integration of resources from a variety of sources. This approach not only enables us to efficiently achieve the project's objectives but also equips us with a suite of knowledge and data integration tools that work collaboratively.