Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

Tourism And Waste Management in Trentino

Document Data:	Reference Persons:
February 12, 2025	Maria Amalia Pelle, Gaudenzia Genoni, Yishak Tadele Nigatu

Index:

Revision History:

Revision	Date	Author	Description of Changes
0.1	October 24, 2024	Maria Amalia Pelle	Document created
1.0	October 31, 2024	Amalia, Gaudenzia, and Yishak	Phase I Compilation
2.0	November 13, 2024	Amalia, Gaudenzia, and Yishak	Phase II Compilation
3.0	November 25, 2024	Amalia, Gaudenzia, and Yishak	Phase III Compilation
4.0	December 4, 2024	Amalia, Gaudenzia, and Yishak	Phase IV Compilation
5.0	February 11, 2025	Amalia, Gaudenzia, and Yishak	Phase IV Compilation

1 Purpose Definition

The initial phase of our Knowledge Graph project focuses on clearly defining the purpose, establishing the domain of interest, and identifying personas and scenarios. These elements inform the formulation of competency questions, which address the needs and goals of the identified personas and scenarios: the competency questions will later be answered by querying the completed knowledge graph. This phase also includes the completion of the Purpose Definition Sheet, where all relevant entities and their associated properties are outlined, and the Entity-Relationship (ER) Modeling.

1.1 Informal Purpose

The purpose of our project is to offer comprehensive information regarding waste management and its relationship with tourism in the Province of Trento. The final Knowledge Graph (KG) will serve as a valuable tool for various stakeholders, including tourists, facility owners, and waste management authorities, by providing accessible information on waste disposal locations, recycling options, and the environmental impact of tourism on local waste management systems. Additionally, the KG will support researchers and experts by enabling in-depth analysis of the interactions between tourism activities and waste generation, promoting evidence-based decision-making in administration and policy development.

1.2 Domain of Interest (DoI)

The focus of this project is on waste production data and tourist infrastructure in the Province of Trento (Trento coordinates are 46°04′00″N, 11°07′00″E). The waste production data, which pertains to individual cities, covers a yearly timeline from 2014 to 2022.

1.3 Scenarios definition

S1 General waste disposal: It's a sunny winter morning at Monte Bondone, a popular skiing destination in Trentino. As visitors enjoy their skiing activities, they generate waste such as snack wrappers and drink containers, which need to be properly disposed of to maintain the area's natural beauty.

S2 Following policies guidance: Facility owners in Trentino must enforce regulations related to waste disposal to comply with local environmental laws. To achieve this, they should regularly review the waste management guidelines provided by the municipality and stay updated on any changes in legislation.

S3 Analysis on tourism waste impact: Trentino requires a thorough analysis of the impact of tourism on waste management to propose new regulations. In response to growing concerns about the increasing waste generated by tourist activities, a comprehensive study is needed to examine current waste management practices in popular tourist areas.

S4 Time Series Analysis: During the COVID-19 pandemic, Trentino faced an unprecedented tourist season with marked fluctuations in visitor numbers due to shifting travel restrictions and health protocols. These changes might have impacted waste production at popular tourist destinations.

S5 Special waste disposal: Sport tourism in Trentino attracts various travelers who bring along sporting equipment such as bicycles, climbing harnesses, and skis. However, accidents can happen even on vacation, leading to damaged gear. Unfortunately, this type of waste cannot always be disposed of easily and often requires transportation to specialized facilities for proper handling.

S6 Waste management practices:The waste management provider aims to conduct a detailed comparative analysis of waste production levels across various municipalities within the region. The primary objective is to identify and understand the most effective waste management practices employed by cities with consistently lower waste outputs.

1.4 Personas

P1 Maximilian is a 35-year-old Austrian citizen living in Innsbruck. An outdoor enthusiast, he enjoys winter sports, particularly skiing, and frequently travels to Trentino during the winter season. Passionate about environmental sustainability, Maximilian actively seeks ways to reduce his ecological footprint wherever he goes.

P2 Luciana is a 45-year-old facility owner in Trento, operating a charming bed-and-breakfast that attracts tourists year-round. She actively seeks information on current waste management regulations and best practices to ensure her establishment complies with legal requirements and promotes eco-friendly tourism.

P3 Diego is a 52-year-old policy maker in Trentino Province. Among his responsibilities, he reviews existing waste management policies to evaluate their effectiveness, identifies gaps, and proposes improvements. Diego collaborates with various stakeholders, including facility owners and environmental organizations, to ensure that the regulations he advocates are practical and beneficial for both the community and the environment.

P4 Chiara is a 28-year-old researcher at FBK, currently participating in a collaborative research project. The goal of her study is to assess how tourism in various cities of Trentino contributes to waste production. She is particularly interested in comparing fluctuations in waste generation during the COVID-19 pandemic. Chiara is passionate about sustainable tourism and aims to identify strategies for reducing waste in the region.

P5 Filippo is a 24-year-old sports enthusiast living in Verona. He thrives on outdoor activities and finds it hard to relax, even on vacation. For this reason, he has chosen Caldonazzo for a weekend getaway with his girlfriend, Giulia. Eager for adventure, he has brought along his own canoe and is excited to make the most of their time together. However, upon arriving at their destination, he realizes that he has broken his paddle during the journey and is unsure of where to dispose of it properly.

P6 Giovanni is a 41-year-old waste management analyst working for Dolomiti Ambiente. With a background in environmental science, he is passionate about finding innovative solutions to improve waste management practices. Giovanni currently aims to compare waste production across various cities in Trentino to identify successful strategies employed by those with lower waste outputs.

1.5 Competency Questions

CQ-1 (P3-S3, P4-S4, P6-S6): As a policy maker or researcher evaluating waste management, what is the total amount of a specific type of waste generated in municipalities that host tourist attractions during a given year?

CQ-2 (P1-S1, P5-S5): While enjoying winter sports, where can I find the nearest recycling bins, and what types of waste can I dispose of in these bins?

CQ-3 (P2-S2): As a facility owner, what waste disposal options are available to me in Trentino, and which types of waste do these facilities accept?

CQ-4 (P2-S2): In order to maintain compliance with local regulations, what are the waste disposal requirements specific to my bed-and-breakfast? What are the acceptable disposal methods for different waste types?

CQ-5 (P3-S3, P6-S6): What types of waste are generated by various tourism activities in Trentino, and how do these types impact overall waste management practices?

CQ-6 (P3-S3, P5-S5, P6-S6): What special waste disposal facilities are available in Trentino, including their locations, capacities, and the types of waste they manage?

CQ-7 (P3-S3, P4-S4, P6-S6): Which areas in Trentino generate the most waste from tourism activities?

CQ-8 (P3-S3, P4-S4, P6-S6): During the COVID-19 pandemic, how did visitor fluctuations impact waste production at popular tourist destinations in Trentino? Can we analyze the yearly waste production trends for specific municipalities using the available data?

1.6 Concepts identification

Table 1 below lists the entities - along with their respective properties - related to the competency questions, taking into consideration both the purpose (knowledge layer) and the available data sources (data layer).

Table 1: Scenarios, Personas, and related Competency questions

Scenarios	Personas	Competency Questions	Entities	Properties	Focus
S3, S4, S6	P3, P4, P6	CQ-1, CQ-7, CQ-8	Municipality	Municipality_ID (PK) Name Location_ID Coordinates Population_size	Common
-	P1, P2, P3, P4, P5, P6	-	Person	Person_ID (PK) Role Name Surname Date_of_birth Country Is_a_Tourist Location_ID	Common
S1, S2, S3, S4, S5, S6	P1, P2, P3, P4, P5, P6	CQ-1, CQ-2, CQ-3, CQ-4, CQ-5, CQ-6, CQ-7, CQ-8	Waste Type	Waste_Type_ID (PK) Category Recyclability Disposal_Method	Core
S1, S4	P1, P3, P4, P5, P6	CQ-1, CQ-2, CQ-5, CQ-7	Location	Location_ID (PK) Name Category Latitude Longitude Municipality_ID (FK)	Core

S2	P2	CQ-4	Tourist Facility	Facility_ID (PK) Name Type Location_ID (FK)	Core
S1, S3	P1, P3	CQ-1, CQ-5	Tourist Activity	Activity_ID (PK) Type Seasonality Location_ID (FK)	Core
S1, S2, S3, S5	P1, P2, P3, P5	CQ-2, CQ-3, CQ-6	Waste Management Facility	Waste_Facility_ID (PK) Name Location_ID (FK)	Contextual
S2, S3	P2, P3	CQ-2, CQ-4	Waste Regulations	Policy_ID (PK) Name Effective_Date Municipality_ID (FK)	Contextual
S3, S4, S6	P3, P4, P6	CQ-1, CQ-5, CQ-7, CQ-8	Waste Production	Waste_Production_ID Quantity Waste_Type_ID Municipality_ID	Contextual

1.7 ER model definition

The ER model in Figure 1 (below) was designed using the previously considered concepts of entities and properties. It represents the initial graphical version of the final structure of the Knowledge Graph.

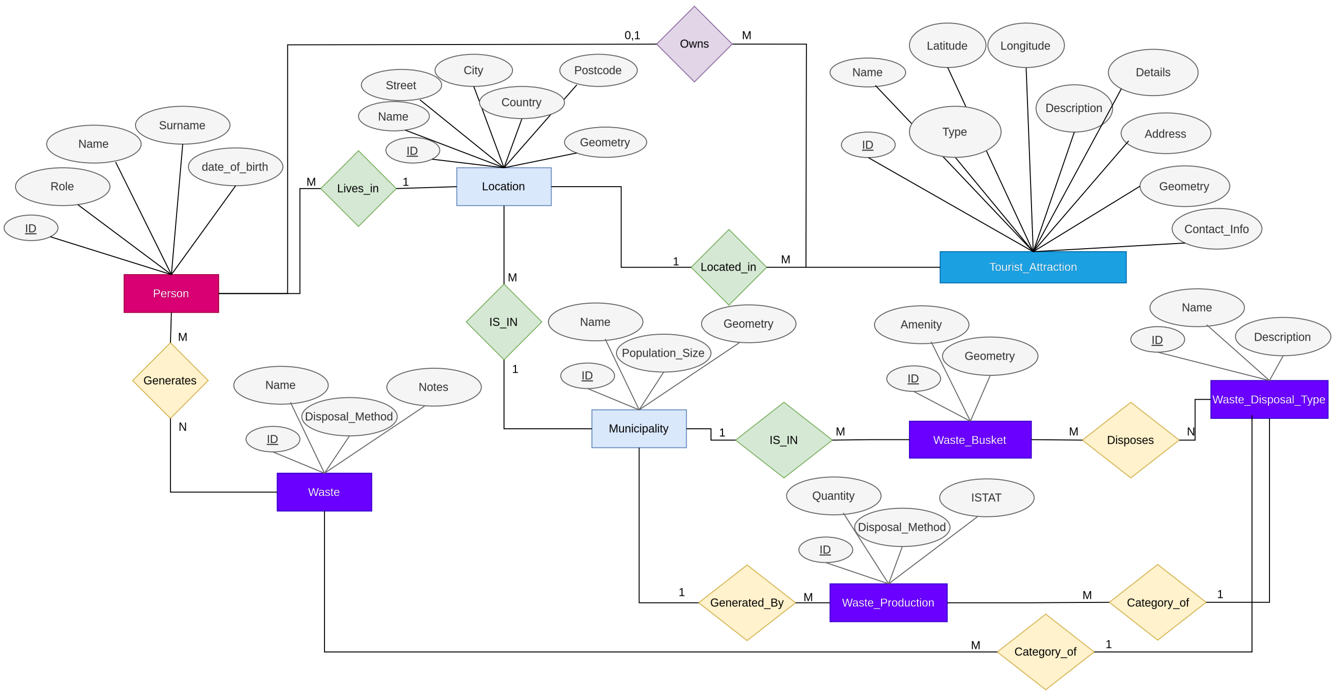


Figure 1: ER Diagram

2 Information Gathering

In this second phase of the project, we identified and gathered all the resources needed to build the graph. These resources include knowledge ontologies, language vocabularies, and data value datasets.

- Regarding the **knowledge ontologies**, we consulted schema.org for the ETypes in our ER model that directly correspond to it. However, since schema.org is a high-quality, general-domain schema, some of our ETypes—particularly those related to waste types and waste disposal methods, which are highly domain-specific and mainly defined by available datasets—were not represented: a future project could involve surveying all relevant resources in the Trentino region, with the goal of standardizing, harmonizing, and making them reusable and interoperable. Table 2 pairs our ETypes with their corresponding schema.org categories.

Table 2: Four ETypes with corresponding schema.org categories

Concept	Schema	Route	Properties
Person	Person	Thing > Person	identifier, givenName, familyName, birthDate
Location	Place	Thing > Place	identifier, name, geo
Municipality	City	Thing > Place > AdministrativeArea > City	identifier, name, geo
Tourist_Attraction	TouristAttraction	Thing > Place > TouristAttraction	identifier, name, latitude, longitude, description, address, telephone

- Regarding the **language vocabularies**, we primarily referred to the Universal Knowledge Core (UKC) and the OpenStreetMap glossary, which will be particularly useful in Phase Three of the project. Both of them are high quality sources.

- With respect to the **data value resources**, we identified seven datasets:

  1. Waste
  2. Waste Baskets
  3. Waste Disposal Types

4. Waste Baskets Disposal Types
5. Municipalities
6. Waste Production
7. Tourist Attraction

For each dataset, we dedicate three subsections to source identification, dataset collection, and cleaning/standardization procedures, as detailed in the following paragraphs.

2.1 Dataset ”Waste Baskets”

This first dataset provides geospatial information on the distribution of waste baskets, organic bins, and recycling points for various materials across the Province of Trento.

2.1.1 Source Identification

In conducting research on waste basket distribution within the Province of Trento, we found no pre-existing datasets suitable for reuse. Consequently, data acquisition was undertaken using OpenStreetMap (OSM), a widely recognized, community-contributed mapping platform. Given the open-source nature of OSM, it provides freely accessible and editable geographical data, which serves as a resource for mapping facilities and amenities. The community-driven aspect of OSM allows for dynamic updates but also introduces potential limitations in terms of data completeness and accuracy.

2.1.2 Data Collection

The data retrieval was facilitated by Overpass Turbo, a tool that allows for specific and customizable querying of the OSM database. The Overpass Turbo query used to collect data is provided below.

```
[out:json];
area["name"="Provincia di Trento"]->.a;
(
  node["amenity"="waste_basket"](area.a);
  node["amenity"="recycling"](area.a);
  node["amenity"="waste_basket"]["waste:organic"="yes"](area.a);
  node["amenity"="recycling"]["recycling:glass"="yes"](area.a);
  node["amenity"="recycling"]["recycling:plastic"="yes"](area.a);
  node["amenity"="recycling"]["recycling:paper"="yes"](area.a);
  node["amenity"="waste_disposal"](area.a);
  node["amenity"="waste_collection_point"](area.a);
);
out body;
>;
out skel qt;
```

This query yielded data on a variety of waste management facilities in the Province of Trento, including general-purpose waste baskets, bins for organic waste, and recycling points for specific materials such as glass, plastic, and paper. These data were collected in a GeoJSON file containing 3,883 points of interest and as many as 84 attributes.
Each point's precise location within the Province of Trento can be visualized on an accompanying map (please,allow a few seconds for it to load).
Given the community-driven nature of the OSM platform, we acknowledge the potential limitations regarding data completeness and accuracy; it is plausible that some areas or types of bins may be underrepresented or inconsistently documented. Nevertheless, in the absence of a more comprehensive source, OSM serves as a viable data source that provides sufficient granularity and geographical coverage for our research objectives.

2.1.3 Data Cleaning and Standardization

In this case, data cleaning involved the removal of 45 columns containing mostly NaN values or information not relevant to our project, while retaining the 39 most meaningful columns. Furthermore, the GeoJSON file was transformed into a CSV format: although the dataset is generally sparse, with many rows missing values for most columns, the available data is useful.
The most important attributes in the dataset are:

- **id**: a unique identifier for each data point.
- **amenity**: it describes the type of facility or service (e.g., waste basket, recycling point, or waste collection point).
- **recycling types**: it indicates the types of materials accepted for recycling, such as glass, plastics, metals, paper, organic waste, and special items.
- **geometry**: it provides spatial coordinates (latitude and longitude) essential for mapping and spatial analysis.
- **municipality**: it lists the municipality in which the waste basket is located (this attribute was derived from a spatial join with the ”Municipalities” dataset, presented below.).

The recycling columns were initially filled with values such as 'yes', 'no', and NaN. To standardize the data, we converted 'yes' values to True, 'no' values to False, and retained NaN values as False. Additionally, we observed that a single waste basket could be associated with multiple recycling types in a so-called ”many-to-many” relationship. As a result, we decided to create a new dataset, namely **“Waste Disposal Types”**, containing 10 waste disposal categories, as follows:

1. Organic: Includes food waste, organic waste, garden waste, and green waste.
2. Paper/Cardboard: Includes paper, cardboard, paper packaging, books, magazines, and newspapers.
3. Glass: Includes general glass waste and glass bottles.
4. Metal: Includes aluminium, cans, scrap metal, and other metal items.
5. Plastic: Includes PET, plastic, plastic bottles, plastic packaging, and beverage cartons.
6. Textiles: Includes clothes and shoes.
7. Electronic Waste: Includes batteries, electrical appliances, electrical items, small appliances, and lamps.
8. Wood: Includes wood items and garden pots.
9. Construction Waste: Includes construction-related waste.
10. Miscellaneous: Includes items like tetrapak and other non-categorized materials.

Following this categorization, which proved to be very important also in the subsequent analysis of waste types and waste production, we mapped the individual recycling types of each waste basket to these higher-level disposal categories, resulting in the creation of a new dataset, **"Waste Baskets Disposal Type"**: the dataset lists baskets alongside the types of waste they accept, with separate rows for each applicable category.

2.2 Dataset "Waste"

This dataset provides structured information on waste types and disposal methods in the Province of Trento, supporting analysis of local waste management practices.

2.2.1 Source Identification

To obtain data regarding waste production, management, and disposal in the Province of Trento, we referred to Dolomiti Ambiente, a subsidiary of the Dolomiti Energia Group responsible for environmental hygiene services and waste collection. Although Dolomiti Ambiente operates exclusively in the municipalities of Trento and Rovereto, it was selected as a representative case study, under the assumption that waste collection practices in other municipalities within the province are managed in a similar and comparable manner. Future research could expand this study to include an analysis of waste collection regulations in municipalities beyond Trento and Rovereto.

2.2.2 Data Collection

Although Dolomiti Ambiente's website does not provide a dedicated dataset on waste types and their proper disposal methods, the company offers a downloadable PDF document titled "Riciclabolario", which serves as a guide for navigating waste sorting for domestic waste. This document became the primary source for the data extraction process. From this source, an initial CSV file was created, listing all waste types alongside their corresponding disposal categories.

2.2.3 Data Cleaning and Standardization

The initial CSV file was subjected to extensive cleaning to extract relevant data into appropriate columns and standardize the structure. The cleaning process began by splitting the *waste* column into two: *waste* and *disposal_method_1*, separating the main waste category from disposal method information following a hyphen. If *disposal_method_1* contained multiple disposal methods separated by a semicolon, the data was split further into a new column, *disposal_method_2*, to ensure each disposal method was clearly identified. Parenthetical information, such as special handling instructions or waste characteristics, was extracted and moved into separate "notes" columns (*waste_notes*, *notes_1*, *notes_2*), ensuring that relevant details were retained without cluttering the primary columns. Finally, an index column was added to provide a sequential identifier for each row, facilitating easier referencing and further indexing operations as needed.
After performing the above cleaning steps, the final dataset consists of the following columns:

- **index**: A sequential identifier for each row.
- **waste**: The main waste category.
- **waste_notes**: Additional information extracted from the *waste column*, if any.
- **disposal_method_1**: The primary disposal type.
- **notes_1**: Additional notes extracted from the *disposal_method_1* column, if any.
- **disposal_method_2**: The secondary disposal type.
- **notes_2**: Additional notes extracted from the *disposal_method_2* column, if any.
- **category**: The ID of the waste category to which the disposal type belongs, based on the dataset "Waste Disposal Types".

2.3 Dataset ”Municipality”

This dataset contains the geographical boundaries of municipalities in the Province of Trento, along with the corresponding ISTAT code and population data. It can be useful for determining which municipality a user is located in.

2.3.1 Source Identification

The ”Municipality” dataset was compiled using two primary sources: population data provided by ISTAT (Italian National Institute of Statistics) and geographic coordinates obtained from OpenStreetMap (OSM). ISTAT is a reliable source for demographic data in Italy, regularly publishing population statistics, including municipal-level data. The Overpass Turbo API was used to retrieve the spatial boundaries of each municipality in the specified region.

2.3.2 Data Collection

Population data for each municipality were collected directly from the ISTAT website in CSV format. These data represent the official population counts recorded on January 1, 2024. The spatial data from OSM, specifically the municipal boundary coordinates, were retrieved using queries to the Overpass Turbo platform, similar to how the waste basket data was collected. The query was designed to capture the administrative boundaries at the municipality level:

```
[out];
area["name"="Provincia di Trento"]
["boundary"="administrative"] ["admin_level"="6"];
rel(area)["boundary"="administrative"]["admin_level"="8"];
out body;
out skel qt;
```

2.3.3 Data Cleaning and Merging

Once collected, both datasets underwent preprocessing to ensure uniformity and compatibility for merging. The ISTAT dataset had all column names translated and adjusted to ensure consistency and compatibility with the other tables. Most columns were dropped, leaving only the ISTAT code and total population. Similarly, the geographic data from Overpass Turbo were structured into a GeoJSON format to facilitate easy data merging and visualization. To create a comprehensive table that includes both population and geographic data, the ISTAT code was used as a key to merge the two datasets. This approach was preferred over using city names, as it ensures consistency—especially given that some municipalities have undergone name changes in the last five years. The GeoJSON file was enhanced by appending a ”Total

Population” tag to each municipality entry, allowing a single dataset to reflect both spatial and demographic information for each municipality.

The final dataset includes the following attributes for each municipality:

- **Name**: The name of the municipality
- **ref: ISTAT**: The ISTAT code of the municipality
- **population**: The population count as of January 1, 2024
- **geometry**: The coordinates of the polygon defining the municipality's borders

As mentioned in the previous section, we acknowledge potential limitations regarding the completeness and accuracy of the data from the OSM platform.

2.4 Dataset ”Waste_production”

This dataset contains a table with the annual waste production in tons for all the cities in the Province of Trento, covering the years from 2014 to 2022. Some data are provided in an aggregated form.

2.4.1 Source Identification

The data was sourced from the Italian Institute for Environmental Protection and Research (ISPRA), which provides comprehensive information on waste production across Italy. The datasets are publicly available on the ISPRA website.

2.4.2 Dataset Collection

These data were collected and compiled by ISPRA, offering insights into waste management trends over the specified period. The dataset were collected from their website.

2.4.3 Dataset Cleaning and Merging

To ensure consistency, the data from all years were combined into a single dataset, with each record labeled by the corresponding year. All column names and attributes were translated, and the names were chosen to be compatible with the other tables to ensure uniformity across the dataset. During this process, unnecessary columns with missing or irrelevant data were removed to improve the quality and relevance of the dataset. Additionally, a new ”category” column was added, where a numeric value describes the specific waste type, using the ”waste_disposal_type” table as a reference. The dataset was then melted to transform it into a long-format table, making it easier to analyze over time. The cleaned dataset was saved as a single CSV file presenting the following attributes:

- **Municipality**: The name of the municipality
- **ref: ISTAT**: The ISTAT code of the municipality
- **category**: waste id from the waste_disposal_type table
- **year**: Year in which data was collected

2.5 Dataset "Tourist Attraction"

Tourist attraction encompass a diverse range of categories, including natural locations, cultural landmarks, and facilities tailored to enhance visitor experiences. In this project, we considered a wide array of tourist attractions, including:

- **Natural Attractions:** Such as protected areas, lakes, rivers, beaches, peaks, viewpoints, caves, waterfalls, and springs.
- **Accommodation Facilities:** Including hotels, holiday apartments, and houses.
- **Dining and Hospitality:** Encompassing food and drink establishments such as Restaurants, Cafes and Bars.
- **Cultural Sites:** Including museums, historic sites, and other cultural attractions.
- **Entertainment and Recreation:** Such as amusement parks, recreational facilities, skiing and winter sports, and other adventure locations.

This comprehensive categorization reflects the broad appeal and variety of attractions available to tourists, from immersive natural environments to comfortable accommodations and lively entertainment options.

2.5.1 Source Identification

The data for this project was sourced from OpenStreetMap via the Overpass Turbo API. We executed multiple queries tailored to each attraction type to gather comprehensive information. The resulting data was retrieved in GeoJSON format and subsequently processed to align with our specific project requirements.

2.5.2 Data Collection

The following is a list of tourist attractions that were taken into consideration in this project:

- Protected Areas
- Lakes and Rivers
- Beaches
- Peaks and Viewpoints
- Caves
- Hotel and Accommodations
- Holiday Appartments and Houses
- Food and Drink Establishments
- Cultural Attractions
- Amusement and Recreational Facilities
- Skiing and Winter Sport Facilities
- Waterfall and Springs

The following query is executed in Overpass API but, to make the data manageable, the queries are run separately and the data is downloaded separately. For instance for *Hotels and Accommodation* the following query is executed:

```
[out:json];
area["name"="Provincia di Trento"]->.searchArea;
(
  // Hotels and Accommodations
  node["tourism"="hotel"](area.searchArea);
  node["tourism"="guest_house"](area.searchArea);
  node["tourism"="hostel"](area.searchArea);
  node["tourism"="camp_site"](area.searchArea);
  node["tourism"="caravan_site"](area.searchArea);
  node["tourism"="chalet"](area.searchArea);
  node["tourism"="alpine_hut"](area.searchArea);
  node["building"="hotel"](area.searchArea);
);
out geom;
```

While for *Food and Drink Establishment* the following query is executed and the result is saved in a separate geoJSON file.

```
[out:json];
area["name"="Provincia di Trento"]->.searchArea;
(
  // Food and Drink Establishments
  node["amenity"="restaurant"](area.searchArea);
  node["amenity"="cafe"](area.searchArea);
  node["amenity"="bar"](area.searchArea);
  node["amenity"="pub"](area.searchArea);
  node["amenity"="fast_food"](area.searchArea);
);
out geom;
```

Every Overpass query can be found in the project's GitHub repository.

2.5.3 Data Preprocessing

The GeoJSON data provides extensive details about each type of attraction, resulting in a sparse dataset with information that varies significantly across attraction types. For instance, food and drink establishments are typically privately owned and include contact details such as phone numbers and emails. In contrast, natural attractions like lakes and rivers lack ownership information, as they are not associated with individual proprietors. Therefore, we carefully selected specific features for each type of attraction to best represent them, ensuring the dataset effectively serves the purpose of our project. Table 3 provides comprehensive information on each processed attraction type.

Table 3: Summary of Generated Data Files and Their Properties

Generated File Name	Columns	Unique Values	NaN Values	Table Size
caves.csv	ID	489	0	489 x 4
	Name	242		
	Latitude	489		
	Longitude	489		
artworks.csv	ID	419	0	419 x 8
	Name	210	208	
	Latitude	419	0	
	Longitude	417	0	
	Artist Name	98	292	
	Artwork Type	11	0	
	Description	14	90	
	Website	2	-	

Generated File Name	Columns	Unique Values	NaN Values	Table Size
memorials.csv	ID Name Latitude Longitude Memorial Type Historic Type Description	504 222 504 504 14 3 80	0 244 0 0 258 0 425	504 x 7
gallery_and_museum.csv	ID Name Latitude Longitude Website Type Street City Postcode House Number	71 69 71 71 11 2 26 18 16 21	0 2 0 0 61 0 46 47 47 47	71 x 10
food_and_drink_establishments.csv	ID Latitude Longitude Name Cuisine Operator Street City Postcode House Number Website Phone Email	2293 2293 2292 1942 139 280 645 163 66 199 334 619 204	0 0 0 192 1598 2007 1185 1314 1298 1217 0 0 0	2293 x 13
holiday_apartments_and_houses.csv	ID Latitude Longitude Name Tourism Type Street City Postcode House Number Website Phone Email	244 244 244 236 1 128 33 21 80 79 89 82	0 0 0 2 45 37 40 51 66 22 49 44	244 x 12

Generated File Name	Columns	Unique Values	NaN Values	Table Size
hotels_and_accommodations.csv	ID Latitude Longitude Name Tourism Type Street City Postcode House Number Cuisine Description Operator Website Phone Email	759 759 759 688 7 303 114 56 111 8 6 37 365 326 264	0 0 0 28 1 356 388 379 365 739 753 721 0 0 0	759 x 15
lakes.csv	ID Central_Latitude Central_Longitude Name	38 31 38 38	0	38 x 4
rivers.csv	ID Central_Latitude Central_Longitude Name	138 137 138 25	0	138 x 4
peaks_and_viewpoints.csv	ID Latitude Longitude Name Description Historic Amenity Height Website	2542 2540 2539 1676 22 4 3 2 3	0 0 0 799 2516 2533 2524 2539 2052	2542 x 9
protected_areas.csv	ID Latitude Longitude Name Source Website Protection Title Leisure	3 3 3 3 3 3 2 1	0	3 x 8

Generated File Name	Columns	Unique Values	NaN Values	Table Size
skiing_and_winter_sports.csv	ID Latitude Longitude Name Sport Description	1292 1292 1291 461 1 1	0 0 0 644 1191 1291	1292 x 6
waterfall_and_spring.csv	ID Latitude Longitude Name Amenity Description	295 295 295 98 3 5	0 0 0 196 264 290	295 x 6

2.5.4 Merging Dataframes

After generating the data in CSV format for each type of attraction, we refined our ER Diagram to highlight the key features aligned with our CQs. We then consolidated this information into the Tourist Attractions table, summarizing the essential attributes as represented by the column names listed in Table 4.

Table 4: Summary of The final Tourist Attraction Table

Column name	Description
ID	Specific Identifier taken from the Overpass API
Name	Name
Type	Tourist Attraction Type. example: hotel, lake, skiing
Latitude and Longitude	Central Coordinates of the place
Description	General Description
Details	Relevant information depending on the type of Tourist Attraction. For instance, a Cafe might have cuisine information, whereas Artwork may consist of the artist's name.
Address	City, Street, Postcode and House number information
Contact	Phone, Email or Website information
Geometry	Location information. A Node, Polygon or way, according to Open Street Map Schema.

3 Language Definition

3.1 Concept identification and Dataset filtering

In this section, the first activity was dedicated to the **concept identification**. We aimed to formalize the concepts and terminologies used in the KG by leveraging two primary data sources for lexical and semantic definitions:

- The Universal Knowledge Core (UKC), a repository of domain-independent concepts.
- OpenStreetMap (OSM), used for domain-specific terminologies, particularly for datasets obtained through the Overpass Turbo tool.

We carefully selected concepts to represent the following elements in the final Language Resource table:

- **Entity Types (ETypes)**, highlighted in red,
- **Data Properties**, highlighted in orange,
- **Object Properties**, highlighted in blue,
- **Data values** (only if used as enumeration categories), highlighted in green.

As input, we primarily relied on the ER model developed in the first phase, which was revised and refined for this purpose based on the real datasets available, ensuring alignment with the actual data.
As mentioned above, the UKC was particularly useful for defining widely-used general terms, such as *name*, *category*, and *description*, while OSM proved effective for domain-specific terminology. For concepts from the UKC, we used the UKC number as the identifier. In the case of OSM, we primarily referenced the link to the relevant web page. However, when a specific web page dedicated to a single concept was unavailable, we used the link to the general web page, preceded by the name of the specific item (for example: aluminium-https://wiki.openstreetmap.org/wiki/Tag:amenity%3Drecycling).
We aimed to maintain consistency across our language resources; however, a few terms had to be newly defined due to their absence in existing vocabularies. Most notably, we defined *id* as ”a unique identifier assigned to each record in a table, ensuring that each entry can be distinctly referenced and retrieved.”, since the UKC vocabulary associated the term’s first concept with a physical object (”a card or badge used to identify the bearer”) and the second concept specifically with psychoanalytic terminology. Another interesting case is the new definition of *geometry* in the GIS context, which was absent in the UKC and was subsequently added by

us, as follows: ”In GIS terminology, geometry refers to the spatial representation of geographic features, defining their shape, location, and size using points, lines, and polygons within a coordinate system”. Furthermore, we had to define specific waste categories (such as *organic*, *paper*, *glass*, *metal*, *plastic*) and specific waste type (such as *Selective Collection (t)* and *Street Cleaning Waste for Recovery (t)*), as they belong to a highly contextual vocabulary and were either absent from the UKC or did not align with the meanings relevant to our domain.

The primary focus of the second activity, **dataset filtering**, was to ensure that all elements included in the final KG were clearly defined by formalized concepts. During this process, we discovered that certain terms in the ”Waste Baskets” dataset — specifically, *recycling:bottles*, *recycling:drugs*, *recycling:electrical_items*, *recycling:garden_pots*, *recycling:lamps*, *recycling:organic_waste*, *recycling:small_appliances*, and *recycling:tetrapak* — were no longer maintained by the OpenStreetMap community. As a result, we decided to filter them out of the final KG.

The final result of this phase is the **Language Resource Table**, which has been uploaded to our GitHub repository under the Phase 3 section.

4 Knowledge Definition

In this phase, we employed the kTelos methodology to model the knowledge teleontology. This process began with developing a bottom-up teleology, progressed to creating a top-down knowledge teleontology, and concluded with their integration into a final teleontology using a middle-out approach. The resulting teleontology will serve as the foundational schema for constructing the Knowledge Graph (KG) in the subsequent phase. All modelling tasks were carried out using Protégé, an industry-standard tool.

4.1 kTelos

4.1.1 Teleology Definition

This teleology represents the initial schema of the Knowledge Graph, constructed using a bottom-up approach based on the ER model. It defines the KG's purpose-specific elements, focusing on the representation of ETypes and their data properties derived from the available datasets.

This teleology is mostly flat and non-hierarchical (except for the Place superclass), reflecting the immediate structure of the data without introducing abstraction. Its primary purpose is to align with the project's competency questions, ensuring that the schema directly supports the intended queries. A preliminary evaluation confirmed this alignment; however, querying the graph with SPARQL in subsequent phases may necessitate revisions to both the competency questions and the schema. Figure 2 presents the teleology as generated in Protégé.

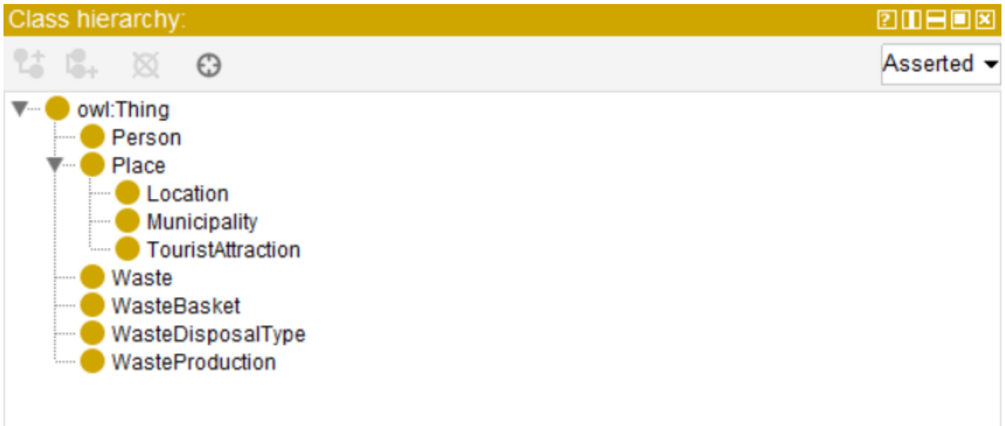


Figure 2: Teleology

4.1.2 Teleontology Definition

The knowledge teleontology represents the most abstract and comprehensive layer of the modelling process, offering a hierarchical classification of Entity Types (ETypes). Its primary objective is to enhance the project's reusability by providing a conceptually rigorous and extensive definition of the domain, facilitating future adaptations and applications in related fields.

To construct the teleontology, we employed a top-down approach, starting with developing an Enhanced Entity-Relationship (EER) model that refined the knowledge schema established during Phase Two. This process involved vertically expanding the TouristAttraction class to align it with a broader range of categories defined in schema.org, thereby incorporating more subclass relationships into the model. Although we may lack data for these specific ETypes, which will therefore not be present in our final knowledge graph, we believe these classes should still be represented in the knowledge layer, as they provide a granular and valuable level of analysis for TouristAttraction data that could be used in future projects.

Additionally, we identified and formalised superclasses for ETypes not covered by schema.org, including Waste, WasteDisposalType, WasteBasket, and WasteProduction, to ensure a comprehensive domain representation. The EER model is shown in Figure 3 below, where ETypes in red indicate elements reused from external knowledge resources (usually well-established classes, like Person and Place), while ETypes in blue represent custom classes designed specifically for this project.

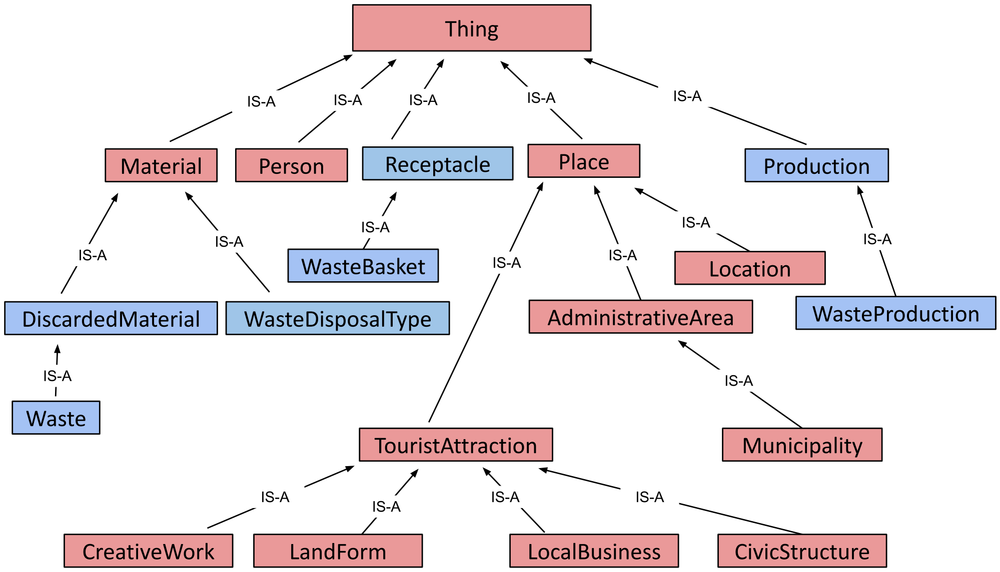


Figure 3: EER Model. ETypes in red indicate elements reused from schema.org, while ETypes in blue represent custom clases designed specifically for this project.

These advancements required extending the linguistic resources of the domain language to accommodate new terms. Once these terms were aligned with the UKC vocabulary to ensure semantic consistency, the complete teleontology was modelled in Protégé, incorporating both data properties and object properties. Figure 4 presents the teleontology as generated in Protégé.

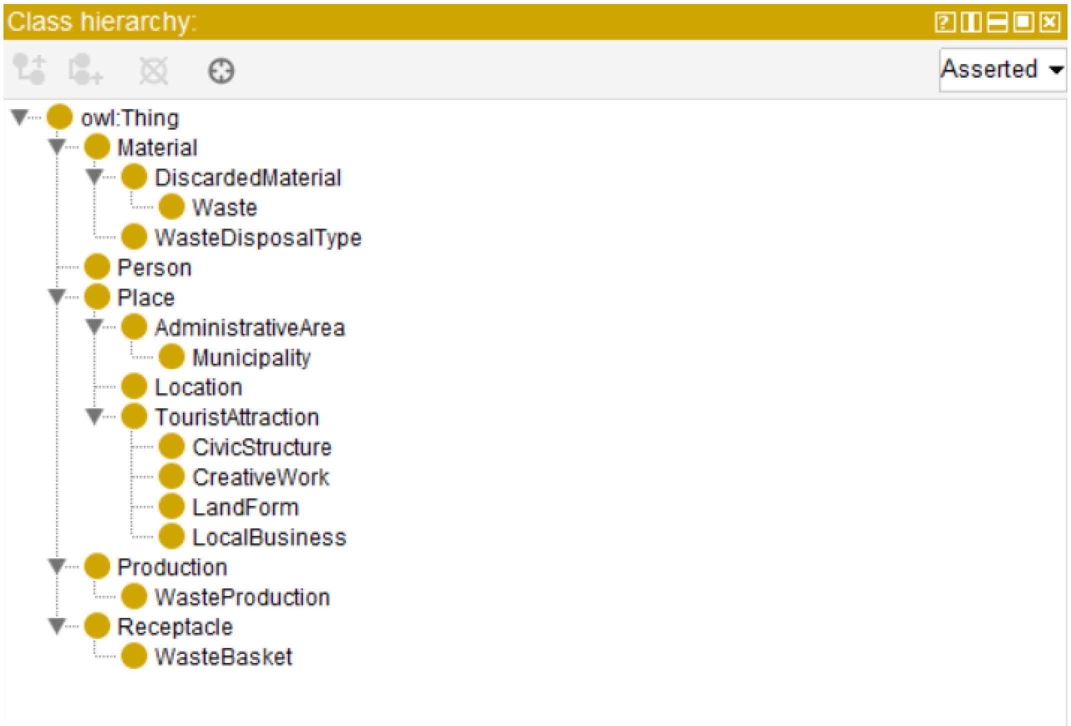


Figure 4: Teleontology

4.2 Schema alignment

The final teleontology was developed by aligning the knowledge teleontology with the OWL file derived from the ER model, following a middle-out approach. This step integrates the teleontology's abstraction with the ER model's data-driven structure to achieve a balance between conceptual rigour and operational efficiency. As this final teleontology will serve as the operational schema for generating the Knowledge Graph, it is designed to facilitate SPARQL queries and address the competency questions identified earlier.
In particular, to simplify query execution in cases where subclass distinctions are not critical, we collapsed certain hierarchies to their leaf nodes for entities such as Waste, WasteDisposalType, WasteBasket, and WasteProduction. Conversely, we preserved the hierarchical structure for ETypes like Place and TouristAttraction: although our project may not populate these classes with extensive real-life instances due to data limitations, we think that maintaining these distinctions is valuable for addressing possible granular, domain-specific competency questions and enhances the overall reusability of our work. Figure 5 presents the final teleontology as generated in Protégé. Figure 6 represents the 25 data properties in the final teleontolgy, while Figure 7 represents the 8 object properties.

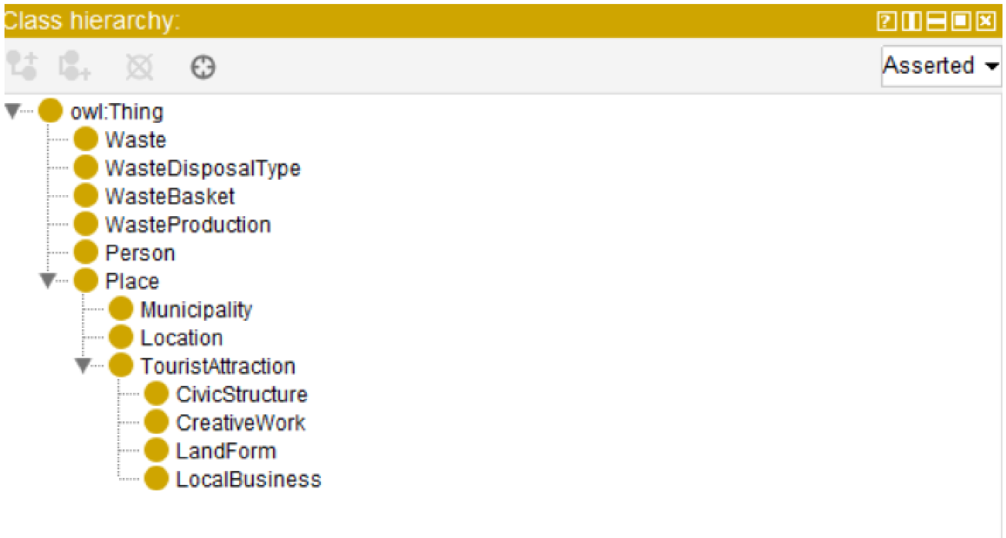


Figure 5: Aligned Teleontology

Figure 6: Data properties

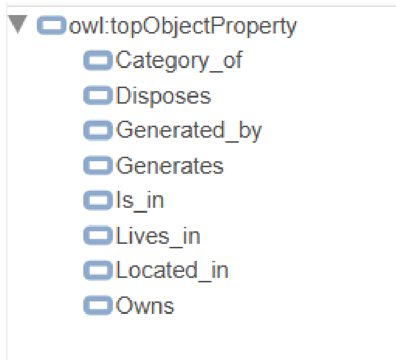


Figure 7: Object properties

5 Entity Definition

The Entity Definition phase is the final step of the iTelos methodology. Its goal is to merge the knowledge layer—represented by the final teleontology—and the data layer—represented by the collected datasets—into a unified structure: the final Knowledge Graph. In the preceding phases, we addressed heterogeneity across sources, formats, and structures. However, data value heterogeneity—which concerns differences in the values of information properties used to identify real-world entities—may still persist. Therefore, before completing the **data mapping** process, it is essential to perform **entity matching** and **entity identification** to ensure consistency and accuracy across all datasets.

5.1 Entity matching

Since all preceding phases were conducted following the iTelos middle-out approach, the final teleontology has been modeled in alignment with the available datasets, while the datasets themselves have been adjusted to reflect the modeling decisions adopted in the teleontology.
We have verified whether an entity appears in multiple datasets described by different sets of properties, ensuring that the various representations of the same entity correspond. As a result, we can confidently say that there are no conflicts at the level of data values.

5.2 Entity identification

Within each dataset, entities must be uniquely identified using an appropriate identifier. Most of our datasets were well-structured and already included unique identifiers for each entity. We only added identifiers for the Municipalities and Waste Production datasets. Additionally, we verified that the foreign keys in each dataset correctly reference the corresponding identifiers in the relevant ETypes. This way, we ensured accurate entity identification.

5.3 Entity mapping

The final entity mapping activity aims to merge the information structure defined in the teleontology with the corresponding data values from the datasets. We carried out this process with the Karma tool (Figure 8). For each EType, we linked its structure—defined in the final teleontology created with Protégé—to the data layer represented by the datasets. A unique URI was associated with each EType, preserving their identifiers, and we carefully mapped all data properties and object properties to the corresponding column attributes in the datasets.
All the mapping operations performed are documented in relative TTL files, to ensure reusability. As a final output from Karma, we generated an RDF-Turtle (TTL) file for each EType, and then we merged them all together to represent our final Knowledge Graph (Figure 9). All the materials and outputs of this phase have been uploaded to our repository.

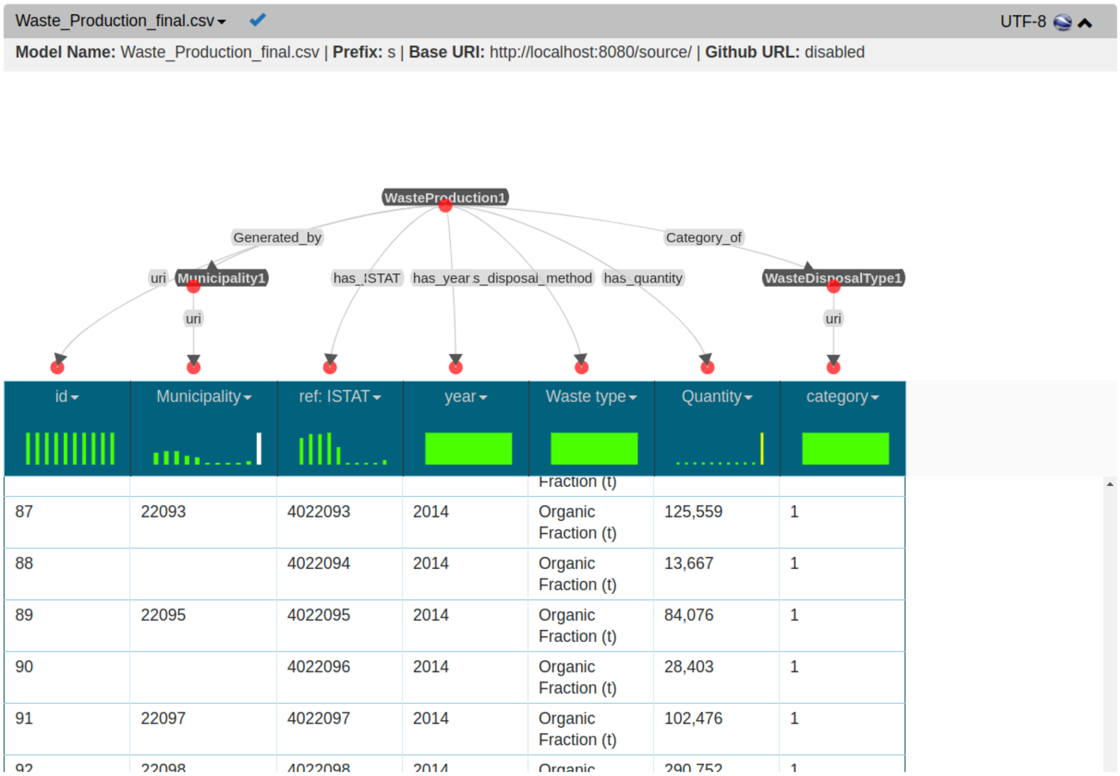


Figure 8: Example of Entity Mapping with the Karma Tool: The image illustrates the mapping of the EType "WasteProduction" to its corresponding data. An URI is assigned to the identifier, while two columns reference other ETypes via object properties, establishing relationships between different classes within the knowledge graph.

```
<http://localhost:8080/source/node/261081646> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://knowdive.disi.unitn.it/etype#WasteBasket> .
<http://localhost:8080/source/node/261081646> <http://knowdive.disi.unitn.it/etype#has_amenity> "recycling" .
<http://localhost:8080/source/22047> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://knowdive.disi.unitn.it/etype#Place> .
<http://localhost:8080/source/node/261081646> <http://knowdive.disi.unitn.it/etype#Is_in> <http://localhost:8080/source/22047> .
<http://localhost:8080/source/node/261081646> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#Thing> .
<http://localhost:8080/source/node/261081646> <http://knowdive.disi.unitn.it/etype#has_geometry> "POINT (11.4362294 46.2814163)" .
<http://localhost:8080/source/22047> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://knowdive.disi.unitn.it/etype#Municipality> .
<http://localhost:8080/source/22047> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#Thing> .
```

Figure 9: The first entity for the EType "WasteBaskets" represented in the karma ttl file that contains our final knowledge graph.

6 Evaluation

In this section, we assess our knowledge graph in terms of purpose satisfaction and reusability to determine whether the final product is both useful and comprehensive. Additionally, we analyze the results obtained by applying evaluation metrics to the knowledge layer and the data layer of the final KG. To validate its effectiveness in retrieving relevant information, we conduct queries using SPARQL.

6.1 Information statistics

- CQ_E: Number of ETypes extracted from the competency questions (CQs) = **9** (see ¶ 1.6).
- CQ_p: Number of properties extracted from the CQs = **42** (see ¶ 1.6).
- RO_E: Number of ETypes extracted from the resource ontologies (ROs) = **5** (see ¶ 2 and ¶ 4.1.2).
- RO_P: Number of properties extracted from the ROs = **17** (see ¶ 2).
- T_E: Number of ETypes of the Teleontology = **14** (see ¶ 4.2).
- T_P: Number of properties of the Teleontology = **33**, of which 8 object properties and 25 data properties (see ¶ 4.2).

6.2 Knowledge Layer Evaluation

To evaluate purpose satisfaction and reusability, we use coverage as a metric. Coverage measures the extent to which a knowledge graph encompasses a specific portion of knowledge, defined in terms of entity types and properties.

6.2.1 Purpose Satisfaction Evaluation

Coverage in the context of purpose satisfaction indicates how well the Teleontology represents the entities and properties derived from the Competency Questions.

- EType Level: $Cov_E(CQ_E) = \frac{CQ_E \cap T_E}{CQ_E} = \frac{6}{9} \approx 0.7$

  All the ETypes extracted from the competency queries are present in the final teleontology, except for *Tourist Activity*, *Waste Management Facility*, and *Waste Regulations*. Overall, the knowledge graph is appropriate for the domain.

- Property Level: $Cov_p(CQ_p) = \frac{CQ_p \cap T_p}{CQ_p} = \frac{30}{42} \approx 0.7$

  The proportion aligns well with the previous evaluation: in this case as well, the properties not included in the intersection are primarily associated with notions from the CQs that were not incorporated into the teleontology.

6.2.2 Reusability

Coverage in terms of reusability measures how well the Teleontology encompasses the entity types and properties extracted from the reference ontologies (in our case, mainly schema.org).

- Etype Level: $Cov_E(RO_E) = \frac{RO_E \cap T_E}{RO_E} = \frac{6}{6} = 1$

  All the ETypes extracted from the reference ontology are present in the final teleontology (*CivicStructure*, *CreativeWork*, *LandForm*, and *LocalBusiness* could also be included, as they are part of the final teleontology). The overall coverage of ETypes compared to the classes extracted from the reference ontology is therefore good. (This specific evaluation however does not reflect the number of ETypes that are custom-specific to our project and thus not grounded in an existing ontology).

- Property Level: $Cov_p(RO_p) = \frac{RO_p \cap T_p}{RO_p} = \frac{17}{17} = 1$

  The value aligns with and confirms the previous analysis.

6.3 Data Layer Evaluation

The evaluation of the data layer focuses on how different parts of the knowledge graph are connected and how dense the graph is. When assessing connectivity, we consider two perspectives: entity connectivity and property connectivity.

6.3.1 Entity Connectivity

Entity Connectivity (EC) measures how well the entities (ETypes) in the knowledge graph are connected through object properties. For each entity type X, the connectivity is computed as

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

where $\sum_{Y=1}^{N}(X,Y)$ is the sum of connectivity values from X to all other entity types (with each object property contributing one unit per specified range) and $OP(X)$ is the number of object properties in which X appears as a domain. The overall connectivity for the knowledge graph is then given by

$$EC(KG) = \sum_{X=1}^{N} EC(X).$$

In our ontology, the defined object properties along with their domains and ranges are as follows:

- **Category_of**: Domains — `Waste`, `WasteProduction`; Range — `WasteDisposalType`.
- **Disposes**: Domains — `WasteBasket`, `WasteDisposalType`; Ranges — `WasteBasket`, `WasteDisposalType`.

- **Generated_by**: Domain — `WasteProduction`; Range — `Municipality`.
- **Generates**: Domain — `Person`; Range — `Waste`.
- **Is_in**: Domains — `Location`, `WasteBasket`; Range — `Municipality`.
- **Lives_in**: Domain — `Person`; Range — `Location`.
- **Located_in**: Domain — `TouristAttraction`; Ranges — `Location`, `Municipality`.
- **Owns**: Domain — `Person`; Range — `TouristAttraction`.

The entity types (ETypes) that occur as domains are: `Waste`, `WasteProduction`, `WasteBasket`, `WasteDisposalType`, `Person`, `Location`, and `TouristAttraction`. Their respective counts of object properties ($OP(X)$) and connectivity row sums (each object property contributes one unit per range specified) are:

- `Waste`: $OP = 1$, row sum $= 1 \Rightarrow EC = \frac{1}{1} = 1$.
- `WasteProduction`: $OP = 2$, row sum $= 2 \Rightarrow EC = \frac{2}{2} = 1$.
- `WasteBasket`: $OP = 2$, row sum $= 3 \Rightarrow EC = \frac{3}{2} = 1.5$.
- `WasteDisposalType`: $OP = 1$, row sum $= 2 \Rightarrow EC = \frac{2}{1} = 2$.
- `Person`: $OP = 3$, row sum $= 3 \Rightarrow EC = \frac{3}{3} = 1$.
- `Location`: $OP = 1$, row sum $= 1 \Rightarrow EC = \frac{1}{1} = 1$.
- `TouristAttraction`: $OP = 1$, row sum $= 2 \Rightarrow EC = \frac{2}{1} = 2$.

Thus, the overall Entity Connectivity for the knowledge graph is:

$$EC(KG) = 1 + 1 + 1.5 + 2 + 1 + 1 + 2 = 9.5.$$

6.3.2 Property Connectivity

Property Connectivity (PC) measures the extent to which data properties connect to a given entity type (EType) within the knowledge graph. For each EType X, the property connectivity is computed as

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

where (X, X) is the number of data property connections for X (i.e., the value in the diagonal cell of the connectivity matrix for data properties) and $DP(X)$ is the number of data properties defined for X. The overall Property Connectivity for the knowledge graph is then

$$PC(KG) = \sum_X PC(X).$$

In our ontology, we determined the contributions by examining, for each EType, how many times it appears in the domain of a data property (each occurrence contributing one unit to the diagonal cell) and by counting the number of distinct data properties for that EType. For example:

- **Waste**: Data properties with `Waste` in their domain are `has_disposal_method`, `has_notes`, `has_name`, and `has_id`. Thus,
$$PC(\text{Waste}) = \frac{4}{4} = 1.$$
- **WasteProduction**: The data properties defined for `WasteProduction` are `has_ISTAT`, `has_disposal_method`, `has_quantity`, `has_year`, and `has_id`. Hence,
$$PC(\text{WasteProduction}) = \frac{5}{5} = 1.$$
- **WasteBasket**: With data properties `has_amenity`, `has_geometry`, and `has_id` having `WasteBasket` as domain, it gives
$$PC(\text{WasteBasket}) = \frac{3}{3} = 1.$$
- **WasteDisposalType**: Data properties for `WasteDisposalType` include `has_description`, `has_id`, and `has_name`. Thus,
$$PC(\text{WasteDisposalType}) = \frac{3}{3} = 1.$$
- **Person**: For `Person`, the data properties are `has_date_of_birth`, `has_id`, `has_name`, `has_role`, and `has_surname`. Hence,
$$PC(\text{Person}) = \frac{5}{5} = 1.$$
- **Location**: The data properties defined for `Location` are `has_city`, `has_country`, `has_geometry`, `has_id`, `has_name`, `has_postcode`, and `has_street`. Therefore,
$$PC(\text{Location}) = \frac{7}{7} = 1.$$
- **TouristAttraction**: For `TouristAttraction`, the data properties include `has_address`, `has_contact_info`, `has_description`, `has_details`, `has_geometry`, `has_id`,

`has_latitude`, `has_longitude`, `has_name`, and `has_type`. Thus,

$$PC(\text{TouristAttraction}) = \frac{10}{10} = 1.$$

For each of the seven ETypes considered, the Property Connectivity is 1. Consequently, the overall Property Connectivity for the knowledge graph ($PC(KG)$) is 7.

6.4 Query Execution

In this section, we present the SPARQL queries executed on the graph to answer the most interesting of the competency questions, along with their results.

CQ-1 *As a policy maker or researcher evaluating waste management, what is the total amount of a type of waste generated in municipalities that host specific types of tourist attractions during a given year?*

The query (Figure 10) retrieves data on the total quantity of **plastic waste** generated in municipalities that host tourist attractions related to **skiing** during **2022**. The results (Figure 11) are grouped by municipality and waste category and then ordered in descending order based on the total waste quantity.
Modifiable variables:

1. `TouristAttraction:has_type` ”skiing” – the type of tourist attraction
2. `FILTER(?wasteCategory = "Plastic")` – the waste category
3. `WasteProduction:has_year` ”2022” – the specific year

```
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd:   <http://www.w3.org/2001/XMLSchema#>

SELECT ?municipalityName ?wasteCategory
       (SUM(xsd:decimal(REPLACE(REPLACE(STR(?wasteQuantity), "\\.", ""), ",", "."))) AS ?totalWaste)
WHERE {
  # Get the municipality and its name.
  ?municipality rdf:type etype:Municipality ;
                etype:has_name ?municipalityName .

  # Get a tourist attraction in the municipality.
  ?attraction rdf:type etype:TouristAttraction ;
              etype:Located_in ?municipality ;
              etype:has_type ?attractionType .
  # Accept attractions with a type containing "skiing" (case-insensitive).
  FILTER(regex(?attractionType, "skiing", "i"))

  # Get waste production data for the municipality for the year 2022.
  ?wasteProduction rdf:type etype:WasteProduction ;
                   etype:Generated_by ?municipality ;
                   etype:has_quantity ?wasteQuantity ;
                   etype:has_year "2022" ;
                   etype:Category_of ?wasteDisposal .

  # Get the waste category.
  ?wasteDisposal rdf:type etype:WasteDisposalType ;
                 etype:has_name ?wasteCategory .

  # Restrict results to only the "Plastic" category.
  FILTER(?wasteCategory = "Plastic")
}
GROUP BY ?municipalityName ?wasteCategory
ORDER BY DESC(?totalWaste)
```

Figure 10: SPARQL query for CQ-1.

municipalityName	wasteCategory	totalWaste
Trento	Plastic	397,422.796
Moena	Plastic	325,676.12
Pinzolo	Plastic	272,545.152
Folgaria	Plastic	226,388.421
Predazzo	Plastic	224,604.513
Canazei	Plastic	223,104.823
Tesero	Plastic	179,786.845
Cavalese	Plastic	128,264.422
Ville di Fiemme	Plastic	111,717.502
Brentonico	Plastic	101,045.3
Lavarone	Plastic	70,737.342
Tre Ville	Plastic	66,175.577
Vallelaghi	Plastic	57,616.656
Andalo	Plastic	40,357.9
Dimaro Folgarida	Plastic	39,378.066
Pergine Valsugana	Plastic	33,910.9
Vermiglio	Plastic	33,443.045
Grigno	Plastic	31,106.212
Peio	Plastic	29,640.465
Ala	Plastic	23,281.602

Figure 11: First rows of query results for CQ-1: Total plastic waste generated in municipalities with skiing and winter sports attractions in 2022.

CQ-2 *While enjoying winter sports, where can I find the nearest recycling bins, and what types of waste can I dispose of in these bins?*

The query (Figure 12) retrieves the closest **waste baskets** to a specified location—here chosen as **Monte Bondone**, one of the most popular skiing destinations near Trento, with coordinates 45.9881°N, 11.0308°E—for each of five selected **waste categories**: Organic, Paper/Cardboard, Glass, Metal, and Plastic. The waste baskets are filtered based on their proximity to the given location, using GeoSPARQL functions to calculate the distance. The query returns the waste basket's ID, its location (coordinates), its distance from the user's position, the waste category that can be disposed of in each waste basket, and a description of the waste type. The results (Figure 13) are ordered by distance to ensure that the nearest waste baskets for each category are listed first.

```
PREFIX rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX geo:   <http://www.opengis.net/ont/geosparql#>
PREFIX geof:  <http://www.opengis.net/def/function/geosparql/>
PREFIX uom: <http://www.opengis.net/def/uom/OGC/1.0/>


SELECT ?recyclingBin ?amenity ?Location ?distance ?wasteTypeName ?wasteTypeDescription
WHERE {
  BIND("recycling" AS ?amenity)
  BIND("POINT (11.0308 45.9881)"^^geo:wktLiteral AS ?userLocation)
  VALUES ?wasteTypeName { "Organic" "Paper/Cardboard" "Glass" "Metal" "Plastic" }

  ?recyclingBin rdf:type etype:WasteBasket ;
                etype:has_amenity ?amenity ;
                etype:has_geometry ?binGeometry ;
                etype:Disposes ?wasteType ;
                etype:Is_in ?municipality .
  ?municipality etype:has_name ?municipalityName .
  ?wasteType etype:has_name ?wasteTypeName ;
             etype:has_description ?wasteTypeDescription .
     BIND(STRDT(?binGeometry, geo:wktLiteral) AS ?binGeometryLiteral)

  BIND(geof:distance(?userLocation, ?binGeometryLiteral, uom:metre) AS ?distance)
    BIND(?binGeometry AS ?Location )

}
ORDER BY ?distance

```

Figure 12: First SPARQL query for CQ-2.

recyclingBin	amenity	Location	distance	wasteTypeName	wasteTypeDescription
http://localhost:8080/source/node/904488844	recycling	POINT (11.0429505 46.0202857)	3699.2107900048068	Paper/Cardboard	Recyclable materials like paper, cardboard, newspapers, magazines, and packaging.
http://localhost:8080/source/node/2405762054	recycling	POINT (10.981728 45.980044)	3906.3811188731524	Organic	Waste that is biodegradable and can be composted, such as food scraps and organic waste.
http://localhost:8080/source/node/2405762054	recycling	POINT (10.981728 45.980044)	3906.3811188731524	Metal	Recyclable metals like aluminum, cans, and scrap metal.
http://localhost:8080/source/node/2405762054	recycling	POINT (10.981728 45.980044)	3906.3811188731524	Plastic	Recyclable plastic items, including PET, plastic bottles, packaging, and plastic bags.
http://localhost:8080/source/node/2368726799	recycling	POINT (10.982041 45.977888)	3945.0109640387327	Paper/Cardboard	Recyclable materials like paper, cardboard, newspapers, magazines, and packaging.
http://localhost:8080/source/node/2368726799	recycling	POINT (10.982041 45.977888)	3945.0109640387327	Glass	Recyclable glass items, including bottles, jars, and containers.
http://localhost:8080/source/node/2368726799	recycling	POINT (10.982041 45.977888)	3945.0109640387327	Metal	Recyclable metals like aluminum, cans, and scrap metal.
http://localhost:8080/source/node/2368726799	recycling	POINT (10.982041 45.977888)	3945.0109640387327	Plastic	Recyclable plastic items, including PET, plastic bottles, packaging, and plastic bags.
http://localhost:8080/source/node/2368726811	recycling	POINT (10.97864 45.978284)	4186.384860971696	Organic	Waste that is biodegradable and can be composted, such as food scraps and organic waste.
http://localhost:8080/source/node/2368726811	recycling	POINT (10.97864 45.978284)	4186.384860971696	Paper/Cardboard	Recyclable materials like paper, cardboard, newspapers, magazines, and packaging.
http://localhost:8080/source/node/2368726811	recycling	POINT (10.97864 45.978284)	4186.384860971696	Glass	Recyclable glass items, including bottles, jars, and containers.
http://localhost:8080/source/node/2368726811	recycling	POINT (10.97864 45.978284)	4186.384860971696	Plastic	Recyclable plastic items, including PET, plastic bottles, packaging, and plastic bags.
http://localhost:8080/source/node/2544544992	recycling	POINT (10.9766423 45.9962387)	4292.23744127991	Paper/Cardboard	Recyclable materials like paper, cardboard, newspapers, magazines, and packaging.
http://localhost:8080/source/node/2544544992	recycling	POINT (10.9766423 45.9962387)	4292.23744127991	Glass	Recyclable glass items, including bottles, jars, and containers.
http://localhost:8080/source/node/2544544992	recycling	POINT (10.9766423 45.9962387)	4292.23744127991	Metal	Recyclable metals like aluminum, cans, and scrap metal.
http://localhost:8080/source/node/2368726829	recycling	POINT (10.975432 45.993036)	4324.659086931598	Paper/Cardboard	Recyclable materials like paper, cardboard, newspapers, magazines, and packaging.
http://localhost:8080/source/node/2368726829	recycling	POINT (10.975432 45.993036)	4324.659086931598	Glass	Recyclable glass items, including bottles, jars, and containers.
http://localhost:8080/source/node/2368726829	recycling	POINT (10.975432 45.993036)	4324.659086931598	Plastic	Recyclable plastic items, including PET, plastic bottles, packaging, and plastic bags.
http://localhost:8080/source/node/6694223516	recycling	POINT (10.9730176 45.9963324)	4569.204606435706	Paper/Cardboard	Recyclable materials like paper, cardboard, newspapers, magazines, and packaging.
http://localhost:8080/source/node/6694223516	recycling	POINT (10.9730176 45.9963324)	4569.204606435706	Glass	Recyclable glass items, including bottles, jars, and containers.

Figure 13: First query results for CQ-2.

For the Competency Question number 2 we have executed a second query, since the question lends itself to a broad investigation. This second query (Figure 14) retrieves **tourist attractions** involved in **skiing** activities located in the municipality of **Folgaria**, and, for each attraction, it finds nearby **recycling facilities** (waste baskets) that accept **Organic** waste and are within a 5000 meter radius. The query computes the distance (in meters) between each attraction and the corresponding waste basket, and returns the municipality name, the attraction name and type, the geographic locations (geometries) of both the attraction and the waste basket, the waste disposal type of the facility, and the computed distance. The results (Figure 15) are ordered by increasing distance.

```
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX geo:   <http://www.opengis.net/ont/geosparql#>
PREFIX geof:  <http://www.opengis.net/def/function/geosparql/>
PREFIX uom:   <http://www.opengis.net/def/uom/OGC/1.0/>

SELECT ?municipalityName ?attractionName ?attractionType ?attractionLocation ?basketLocation ?wasteDisposalType ?distance
WHERE {
  # Retrieve tourist attractions located in Folgaria.
  ?attraction rdf:type etype:TouristAttraction ;
              etype:has_type ?attractionType ;
              etype:has_name ?attractionName ;
              etype:has_geometry ?attractionLocation ;
              etype:Located_in ?municipality .
  ?municipality etype:has_name ?municipalityName .
  FILTER(?municipalityName = "Folgaria")

  # Filter so that the attraction's type is exactly "skiing".
  FILTER(?attractionType = "skiing")

  # Optionally include additional tourist attraction attributes.
  OPTIONAL { ?attraction etype:has_description ?attractionDescription . }

  # For each attraction, find waste baskets (recycling facilities).
  ?wasteBasket rdf:type etype:WasteBasket ;
               etype:has_amenity "recycling" ;
               etype:has_geometry ?basketLocation ;
               etype:Disposes ?wasteDisposal .

  # Compute the distance (in meters) from the attraction to the waste basket.
  BIND(geof:distance(?attractionLocation, ?basketLocation, uom:metre) AS ?distance)

  # Only include waste baskets within 5000 meters.
  FILTER(?distance <= 5000)

  # Retrieve the waste disposal type name.
  ?wasteDisposal rdf:type etype:WasteDisposalType ;
                 etype:has_name ?wasteDisposalType .

  # Filter to only include waste baskets that accept Organic waste.
  FILTER(?wasteDisposalType = "Organic")
}
ORDER BY ASC(?distance)
```

Figure 14: Second SPARQL query for CQ-2.

municipalityName	attractionName	attractionType	attractionLocation	basketLocation	wasteDisposalType	distance
Folgaria	Dosso della Madonna	skiing	POINT (11.1928802 45.9179893)	POINT (11.1921738 45.9190188)	Organic	126.87392690511479
Folgaria	Costa	skiing	POINT (11.1943825 45.9180297)	POINT (11.1921738 45.9190188)	Organic	203.58099373082524
Folgaria	Dosso della Madonna	skiing	POINT (11.193116 45.9163082)	POINT (11.1921738 45.9190188)	Organic	310.02231315140125
Folgaria	Moreta	skiing	POINT (11.2032445 45.9096547)	POINT (11.1921738 45.9190188)	Organic	1349.4443473458496
Folgaria	Dosso della Madonna	skiing	POINT (11.1928802 45.9179893)	POINT (11.1701324 45.9164737)	Organic	1772.7672980662178
Folgaria	Dosso della Madonna	skiing	POINT (11.193116 45.9163082)	POINT (11.1701324 45.9164737)	Organic	1783.1600724592367
Folgaria	Dosso della Madonna	skiing	POINT (11.1928802 45.9179893)	POINT (11.168569 45.9170333)	Organic	1889.0132589109214
Folgaria	Costa	skiing	POINT (11.1943825 45.9180297)	POINT (11.1701324 45.9164737)	Organic	1889.2239874326824
Folgaria	Dosso della Madonna	skiing	POINT (11.193116 45.9163082)	POINT (11.168569 45.9170333)	Organic	1906.048694940968
Folgaria	Costa	skiing	POINT (11.1943825 45.9180297)	POINT (11.168569 45.9170333)	Organic	2005.6275882098776
Folgaria	Dosso della Madonna	skiing	POINT (11.1928802 45.9179893)	POINT (11.1665854 45.9176184)	Organic	2040.3124793272464
Folgaria	Dosso della Madonna	skiing	POINT (11.193116 45.9163082)	POINT (11.1665854 45.9176184)	Organic	2063.365395892541
Folgaria	Dosso della Madonna	skiing	POINT (11.1928802 45.9179893)	POINT (11.1660356 45.9175432)	Organic	2083.139888680702
Folgaria	Dosso della Madonna	skiing	POINT (11.193116 45.9163082)	POINT (11.1660356 45.9175432)	Organic	2105.3540910939064
Folgaria	Costa	skiing	POINT (11.1943825 45.9180297)	POINT (11.1665854 45.9176184)	Organic	2156.925003373935
Folgaria	Costa	skiing	POINT (11.1943825 45.9180297)	POINT (11.1660356 45.9175432)	Organic	2199.7590138133482
Folgaria	Moreta	skiing	POINT (11.2032445 45.9096547)	POINT (11.1701324 45.9164737)	Organic	2678.460171082592
Folgaria	Moreta	skiing	POINT (11.2032445 45.9096547)	POINT (11.168569 45.9170333)	Organic	2812.5000902981283
Folgaria	Sommo Alto	skiing	POINT (11.1995053 45.8939233)	POINT (11.1921738 45.9190188)	Organic	2846.7714014441663
Folgaria	Fondo Grande	skiing	POINT (11.1874393 45.8930376)	POINT (11.1921738 45.9190188)	Organic	2911.070614148175

Figure 15: First rows of results for the second query on CQ-2: Organic Waste Disposal Facilities Near Skiing Tourist Attractions in Folgaria.

CQ-4 *In order to maintain compliance with local regulations, what are the acceptable disposal methods for different waste types?*

The query (Figure 16) returns the names of **waste items** belonging to one specific waste category, namely **Paper/Cardboard**, chosen by the user, along with their associated **disposal methods** based on Dolomiti Ambiente specifications (Figure 17).

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

SELECT ?wasteName (GROUP_CONCAT(DISTINCT ?disposalMethod; separator=", ") AS ?disposalMethods)
WHERE {
  ?waste rdf:type etype:Waste ;
         etype:has_name ?wasteName ;
         etype:has_disposal_method ?disposalMethod ;
         etype:Category_of ?wasteCat .
  ?wasteCat rdf:type etype:WasteDisposalType ;
            etype:has_name "Paper/Cardboard" .
  FILTER(?disposalMethod != "Organic Fraction (t)")
}
GROUP BY ?wasteName
```

Figure 16: SPARQL query for CQ-4.

wasteName	disposalMethods
CALENDARIO DA MURO O DA TAVOLO IN CARTA O CARTONE	CARTA E CARTONE
CARTA	CARTA E CARTONE
CARTA IN ROTOLO "COPRI LETTINI"	CARTA E CARTONE
CARTA TISSUE E FAZZOLETTI	CARTA E CARTONE
CARTONE	CARTA E CARTONE
CONFEZIONI IN CARTONCINO PER ALIMENTI	CARTA E CARTONE
DETERSIVO	CARTA E CARTONE
ETICHETTE IN CARTA	CARTA E CARTONE
FAZZOLETTI DI CARTA	CARTA E CARTONE
FOGLI DI CARTA O CARTONE	CARTA E CARTONE
GIORNALE	CARTA E CARTONE
IMBALLAGGI IN CARTONE	CR, CARTA E CARTONE
LIBRI	CARTA E CARTONE
CARTA E CARTONE	CARTA E CARTONE
QUADERNO	CARTA E CARTONE
RIVISTA	CARTA E CARTONE
SACCHETTO DI CARTA	CARTA E CARTONE
SCATOLA PER PIZZA	CARTA E CARTONE
SHOPPER IN CARTA	CARTA E CARTONE
TOVAGLIA IN CARTA	CARTA E CARTONE

Figure 17: Query results for CQ-4: Waste items and their associated disposal methods in the 'Paper/Cardboard' category.

CQ-6 *What special waste disposal facilities are available in Trentino, including their locations, capacities, and the types of waste they manage?*

This query (Figure 18) retrieves all **waste baskets** (recycling facilities) located in the municipality of **Trento**. For each facility, it returns the municipality's name, the waste basket identifier, its **geospatial location**, and the associated waste disposal type for ten specific waste categories (Organic, Paper/Cardboard, Glass, Metal, Plastic, Textiles, Electronic Waste, Wood, Construction Waste, and Miscellaneous). The records are alphabetically ordered by the waste disposal type.

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

SELECT ?municipalityName ?wasteBasket ?basketLocation ?wasteDisposalType
WHERE {
  # Retrieve the municipality and its name.
  ?municipality rdf:type etype:Municipality ;
                etype:has_name ?municipalityName .
  # Only include municipalities named "Trento".
  FILTER(?municipalityName = "Trento")

  # Get waste baskets located in the municipality.
  ?wasteBasket rdf:type etype:WasteBasket ;
               etype:Is_in ?municipality ;
               etype:has_geometry ?basketLocation ;
               etype:Disposes ?wasteDisposal .

  # Get the waste disposal type (the type of waste managed).
  ?wasteDisposal rdf:type etype:WasteDisposalType ;
                 etype:has_name ?wasteDisposalType .

  # Filter to only include specified waste categories.
  FILTER(?wasteDisposalType IN (
      "Organic", "Paper/Cardboard", "Glass", "Metal", "Plastic",
      "Textiles", "Electronic Waste", "Wood", "Construction Waste", "Miscellaneous"
  ))
}
ORDER BY ?municipalityName ?wasteDisposalType

```

Figure 18: SPARQL query for CQ-6.

municipalityName	wasteBasket	basketLocation	wasteDisposalType
Trento	http://localhost:8080/source/node/6400959249	POINT (11.123234 46.090864)	Construction Waste
Trento	http://localhost:8080/source/node/9010796607	POINT (11.1222552 46.0494316)	Construction Waste
Trento	http://localhost:8080/source/node/434209691	POINT (11.1459495 46.0484302)	Electronic Waste
Trento	http://localhost:8080/source/node/434209691	POINT (11.1459495 46.0484302)	Glass
Trento	http://localhost:8080/source/node/560073337	POINT (11.1259863 46.0703468)	Glass
Trento	http://localhost:8080/source/node/672005529	POINT (11.1390106 46.0422464)	Glass
Trento	http://localhost:8080/source/node/678059806	POINT (11.1628721 46.0716864)	Glass
Trento	http://localhost:8080/source/node/678084020	POINT (11.17673 46.0718219)	Glass
Trento	http://localhost:8080/source/node/764073211	POINT (11.1018759 46.0615945)	Glass
Trento	http://localhost:8080/source/node/814239873	POINT (11.1243672 46.0713448)	Glass
Trento	http://localhost:8080/source/node/1000099577	POINT (11.1241181 46.0723374)	Glass
Trento	http://localhost:8080/source/node/1333553419	POINT (11.1407777 46.0399316)	Glass
Trento	http://localhost:8080/source/node/1333553457	POINT (11.139667 46.0389526)	Glass
Trento	http://localhost:8080/source/node/1364971670	POINT (11.0587539 46.041357)	Glass
Trento	http://localhost:8080/source/node/1749611847	POINT (11.1216299 46.0770724)	Glass
Trento	http://localhost:8080/source/node/1825508119	POINT (11.0477934 46.0330065)	Glass
Trento	http://localhost:8080/source/node/1827497717	POINT (11.0488988 46.0324836)	Glass
Trento	http://localhost:8080/source/node/1937993531	POINT (11.1490711 46.0665354)	Glass
Trento	http://localhost:8080/source/node/2805921698	POINT (11.1240289 46.06744)	Glass
Trento	http://localhost:8080/source/node/2805923903	POINT (11.1238983 46.0693416)	Glass

Figure 19: First rows of results for CQ-6: Waste baskets in Trento, ordered by waste disposal type.

CQ-8 *During the COVID-19 pandemic, how did visitor fluctuations impact waste production at popular tourist destinations in Trentino? Can we analyze the yearly waste production trends for specific municipalities using the available data?*

The query (Figure 20) first selects the **top municipalities** based on the number of tourist attractions connected to them. For each municipality and for each waste disposal type among Glass, Organic, Paper, and Plastic, it aggregates the waste production for 2018, 2020, and 2022. It then calculates **differences in waste production** between 2020–2018 and 2022–2020 to assess fluctuations before, during, and after the **COVID-19 pandemic**. The final results (Figure 21) are ordered by the number of tourist attractions (in descending order), followed by municipality name and waste category.

```sparql
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd:   <http://www.w3.org/2001/XMLSchema#>

SELECT ?municipalityName ?attractionCount ?wasteCategory
       ((?w2020 - ?w2018) AS ?diff_2020_2018) ((?w2022 - ?w2020) AS ?diff_2022_2020)
WHERE {
  { # Top 10 municipalities by attractions
       SELECT ?municipality ?municipalityName (COUNT(?attraction) AS ?attractionCount) {
      ?municipality a etype:Municipality ; etype:has_name ?municipalityName .
      ?attraction a etype:TouristAttraction ; etype:Located_in ?municipality .
    }
    GROUP BY ?municipality ?municipalityName
    ORDER BY DESC(?attractionCount) LIMIT 10
  }

  # Unified waste data collection for all years
  OPTIONAL {
    SELECT ?municipality ?wasteCategory
        (SUM(xsd:decimal(REPLACE(REPLACE(STR(?q2018), "\\.", ""), ",", "."))) AS ?w2018)
        (SUM(xsd:decimal(REPLACE(REPLACE(STR(?q2020), "\\.", ""), ",", "."))) AS ?w2020)
        (SUM(xsd:decimal(REPLACE(REPLACE(STR(?q2022), "\\.", ""), ",", "."))) AS ?w2022)
    WHERE {
      { ?w a etype:WasteProduction ; etype:Generated_by ?municipality ; etype:has_year "2018" ;
          etype:has_quantity ?q2018 ; etype:Category_of/etype:has_name ?wasteCategory . }
      UNION
      { ?w a etype:WasteProduction ; etype:Generated_by ?municipality ; etype:has_year "2020" ;
          etype:has_quantity ?q2020 ; etype:Category_of/etype:has_name ?wasteCategory . }
      UNION
      { ?w a etype:WasteProduction ; etype:Generated_by ?municipality ; etype:has_year "2022" ;
          etype:has_quantity ?q2022 ; etype:Category_of/etype:has_name ?wasteCategory . }
      FILTER(?wasteCategory IN ("Organic", "Paper/Cardboard", "Glass", "Plastic"))
    }
    GROUP BY ?municipality ?wasteCategory
  }
  FILTER(BOUND(?w2018) && BOUND(?w2020)) # Ensure required years exist
}
ORDER BY DESC(?attractionCount) ?municipalityName ?wasteCategory
```

Figure 20: SPARQL query for CQ-8.

municipalityName	attractionCount	wasteCategory	diff_2020_2018	diff_2022_2020
Moena	2,374	Organic	-139.15	145.329
Moena	2,374	Paper/Cardboard	7.027	-0.313
Moena	2,374	Plastic	211.931	-165.807
Trento	1,883	Glass	190.16	-4,306.26
Trento	1,883	Organic	-2,394.289	434.209
Trento	1,883	Paper/Cardboard	-1,178.241	47.499
Trento	1,883	Plastic	-3,778.034	-23.241
Canazei	1,804	Glass	-317.058	292.276
Canazei	1,804	Organic	-121.639	-5.633
Canazei	1,804	Paper/Cardboard	5.929	-15.414
Canazei	1,804	Plastic	243.832	-210.352
Tesero	1,471	Glass	11.142	30.187
Tesero	1,471	Organic	-75.546	71.887
Tesero	1,471	Paper/Cardboard	-9.936	37.102
Tesero	1,471	Plastic	-6.303	29.549
Pinzolo	1,420	Glass	-30.811	-1.551
Pinzolo	1,420	Organic	-149.956	33.972
Pinzolo	1,420	Paper/Cardboard	-5.983	-16.268
Pinzolo	1,420	Plastic	-6.355	0.775
Folgaria	1,410	Glass	23.4	29.36

Figure 21: Results for CQ-8: Yearly waste production differences (2020–2018 and 2022–2020) for some of the most popular tourist municipalities in Trentino, categorized by four waste disposal types: Glass, Organic, Paper, and Plastic.

7 Metadata Definition

Metadata is data that describes other data, providing information about its characteristics, structure, and content. Structured and informative metadata are essential for enabling data reusability and the distribution of resources through data catalogs.
We have structured the metadata into three categories, each represented as a table in our GitHub repository:

- People Metadata
- Project Metadata
- Dataset Metadata

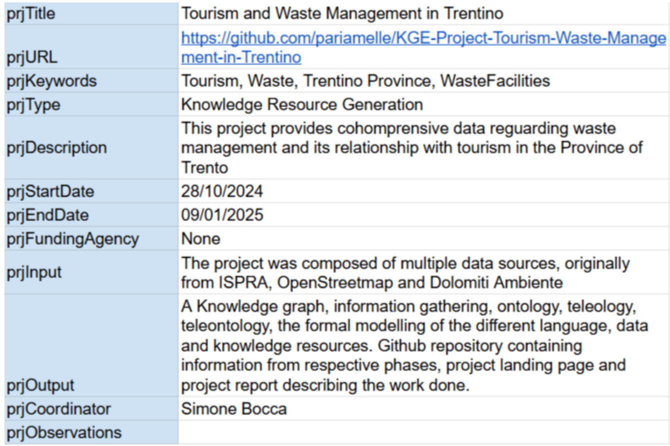
The following sections present these tables along with additional metadata collected at different project stages.

7.1 Project Metadata Description

The following metadata describe the project's authors, inputs, and outcomes. These metadata are included in the People Metadata and Project Metadata tables:

comIdentifier	firstName	lastName	email	nationality	gender	affiliation	personalWebpage
245653	Gaudenzia	Genoni	gaudenzia.genoni@studenti.unitn.it	Italian	female	University of Trento	https://github.com/Ggenoni
247815	Yishak Tadele	Nigatu	yishaktadele.nigatu@studenti.unitn.it	Ethiopian	male	University of Trento	https://github.com/isaaclucky
249468	Maria Amalia	Pelle	mariaamalia.pelle@studenti.unitn.it	Italian	female	University of Trento	https://github.com/pariamelle

Figure 22: People Metadata Table



prjTitle	Tourism and Waste Management in Trentino
prjURL	https://github.com/pariamelle/KGE-Project-Tourism-Waste-Management-in-Trentino
prjKeywords	Tourism, Waste, Trentino Province, WasteFacilities
prjType	Knowledge Resource Generation
prjDescription	This project provides cohomprensive data reguarding waste management and its relationship with tourism in the Province of Trento
prjStartDate	28/10/2024
prjEndDate	09/01/2025
prjFundingAgency	None
prjInput	The project was composed of multiple data sources, originally from ISPRA, OpenStreetmap and Dolomiti Ambiente
prjOutput	A Knowledge graph, information gathering, ontology, teleology, teleontology, the formal modelling of the different language, data and knowledge resources. Github repository containing information from respective phases, project landing page and project report describing the work done.
prjCoordinator	Simone Bocca
prjObservations	

Figure 23: Project Metadata Table

7.2 Language Resources Metadata

The following metadata describe the language resources, including information on the terms that compose the language resource and the sources from which the terms were collected:

- **Resource License**: Open Database License (ODbL)
- **Resource Name**: KGE24 - TWM Project - Language resource.xlsx
- **Resource URL**: Project GitHub Repository
- **Resource Keywords**: Waste management, tourism facilities, waste, places
- **Resource Publisher**: Gaudenzia Genoni, Yishak Tadele Nigatu, Maria Amalia Pelle
- **Resource Creator**: Gaudenzia Genoni, Yishak Tadele Nigatu, Maria Amalia Pelle
- **Resource Language**: English
- **Resource Item Source**: UKC, wiki.openstreetmap, KGE2024
- **Resource Size**: 134 concepts
- **Resource Publication Date**: 21/01/2024
- **Resource Description**: Description of the language resources used in the Tourism and Waste Management in Trentino project
- **Resource Version**: 1.0
- **Resource Domain**: Territory, society, administration

7.3 Knowledge Resources Metadata

The following table presents metadata describing the Etypes. Since no specific metadata were available for each Etype, we decided to use the last website update as the last publication date and applied the same approach for the Etype version.

EtypLicence	EtypURL	EtypCreator	EtypOwner	EtypLang	EtypName	EtypPubTime	EtypDesc	EtypVersion
W3C CLA	https://schema.org/Person	Schema.org	Schema.org	english	Person	22/11/2024	A person (alive, dead, undead, or fictional).	V28.1
W3C CLA	https://schema.org/Place	Schema.org	Schema.org	english	Place	22/11/2024	Entities that have a somewhat fixed, physical extension.	V28.1
W3C CLA	https://schema.org/City	Schema.org	Schema.org	english	City	22/11/2024	A city or town.	V28.1
W3C CLA	https://schema.org/TouristAttraction	Schema.org	Schema.org	english	TouristAttraction	22/11/2024	A tourist attraction. In principle any Thing can be a TouristAttraction, from a Mountain and LandmarksOrHistoricalBuildings to a LocalBusiness.	V28.1

Figure 24: Knowledge Metadata Table

7.4 Data Resources Metadata

The following metadata describe the dataset and its sources. These metadata are included in the Dataset Metadata:

DatLicense	CC BY 4.0	CC BY 4.0	ODbL	ODC-BY	ODC-BY	ODbL	CC BY 3.0 IT
DatURL	https://github.com/pariamelle/KGE-Project-Tourism-Waste-Management-in-Trentino/blob/main/Phase%205%20-%20Data%20Definition/processed_data/Waste/Waste_Production.csv	http://github.com/pariamelle/KGE-Project-Tourism-Waste-Management-in-Trentino/blob/main/Phase%205%20-%20Data%20Definition/processed_data/Waste/waste.csv	https://github.com/pariamelle/KGE-Project-Tourism-Waste-Management-in-Trentino/blob/main/Phase%205%20-%20Data%20Definition/processed_data/Waste/waste_baskets.csv	https://github.com/pariamelle/KGE-Project-Tourism-Waste-Management-in-Trentino/blob/main/Phase%205%20-%20Data%20Definition/processed_data/Waste/waste_baskets_disposal_type.csv	https://github.com/pariamelle/KGE-Project-Tourism-Waste-Management-in-Trentino/blob/main/Phase%205%20-%20Data%20Definition/processed_data/Waste/waste_disposal_type.csv	https://github.com/pariamelle/KGE-Project-Tourism-Waste-Management-in-Trentino/tree/main/Phase%205%20-%20Data%20Definition/processed_data/toursit_attraction	https://github.com/pariamelle/KGE-Project-Tourism-Waste-Management-in-Trentino/blob/main/Phase%205%20-%20Data%20Definition/processed_data/Cleaned_Municipality.geojson
DatKeyword	Waste Production	Waste	Waste Basket	Waste Basket Disposal Type	Waste Disposal Type	Tourist Attraction	Municipality
DatPublisher	Gaudenzia Genoni, Yishak Tadele Nigatu, Maria Amalia Pelle	Gaudenzia Genoni, Yishak Tadele Nigatu, Maria Amalia Pelle	Gaudenzia Genoni, Yishak Tadele Nigatu, Maria Amalia Pelle	Gaudenzia Genoni, Yishak Tadele Nigatu, Maria Amalia Pelle	Gaudenzia Genoni, Yishak Tadele Nigatu, Maria Amalia Pelle	Gaudenzia Genoni, Yishak Tadele Nigatu, Maria Amalia Pelle	Gaudenzia Genoni, Yishak Tadele Nigatu, Maria Amalia Pelle
DatCreator	ISPRA	Dolomiti Ambiente	OpenStreetMap comunity	Gaudenzia Genoni	Gaudenzia Genoni	OpenStreetMap comunity	ISTAT, OpenStreetMap comunity
DatOwner	Italian National Center of Waste and Circular Economy	Dolomiti Ambiente	OpenStreetMap	Gaudenzia Genoni	Gaudenzia Genoni	OpenStreetMap	ISTAT, OpenStreetMap
DatLanguage	English	English	English	English	English	English	English
DatSize	949 KB	18.4 KB	1.05 MB	54.5 KB	989 Bytes	4.25 MB	10.6 MB
DatName	Waste_production.csv	waste.csv	waste_basket.csv	waste_basket_disposal_type.csv	waste_disposal_type.csv	final_processed_data.csv	Cleaned_Municipality.geojson
DatPublicationTimestamp	11/02/2025	11/02/2025	11/02/2025	11/02/2025	11/02/2025	11/02/2025	11/02/2025
DatDescription	annual waste production in tons for all the cities in the Province of Trento, covering the years from 2014 to 2022	waste types and disposal methods in the Province of Trento	geospatial information on the distribution of waste baskets, organic bins, and recycling points for various materials across the Province of Trento	baskets alongside the types of waste they accept, with separate rows for each applicable category	waste disposal categories	tourist attraction encompass a diverse range of categories, including natural locations, cultural landmarks, and facilities tailored to enhance visitor experiences	geographical boundaries of municipalities in the Province of Trento, along with the corresponding ISTAT code and opulation data
DatVersion	7	2	4	2	1	3	5
DatDomain	Society	Administration	Territory	Administration	Administration	Territory	Territory
DatFileFormat	csv	csv	csv	csv	csv	csv	geojson

Figure 25: Dataset Metadata Table

8 Open Issues

Overall, we are satisfied with the process of building our knowledge graph and its associated phases, both at the knowledge layer and at the data layer. The evaluation metrics show good results and indicate that the final outcome of the project is satisfactory. Moreover, the queries successfully answer the majority of the competency questions (except those related to specific regulations for facility owners due to missing data), providing insights that could support future analyses on waste management and tourism in the Province of Trento.

However, some issues remain open. At the knowledge layer, we recognize that further exploration is needed to identify more comprehensive resource ontologies related to waste management operations. Thus far, we have primarily relied on `schema.org`, but the reusability and interoperability of the project could be improved by integrating additional existing standards.

At the data layer, we occasionally encountered problems of missing or incomplete information. For example, we were unable to collect detailed regulations for waste management in the Province of Trento and had to rely on specifications provided by Dolomiti Ambiente. Additionally, the data extracted from OpenStreetMap may be incomplete due to the absence of a strict validation process for community-contributed content. The lack of publicly available, granular data has, in some cases, prevented us from mapping all ETypes to the corresponding data; nonetheless, we chose to retain these classes in the final teleontology to preserve the most precise possible definition of our domain and to ensure future reusability.

Finally, a potential improvement for this project could be the inclusion of personal data under the Person EType. Integrating such data—particularly information about the user's current location—would enable real-time queries and make the knowledge graph more dynamic and responsive to users' needs, which would certainly enhance its potential for application development.