Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

# KGE 2024 - HealthRoute Trentino

**Project Report**

| Document Data: | Reference Persons: |
| --- | --- |
| February 6, 2026 | Alice Massacci |

# Index:

# Revision History:

| Revision | Date | Author | Description of Changes |
|---|---|---|---|
| 0.1 | October 16, 2024 | Alice Massacci | Document created |
| 1.0 | October 19, 2024 | Alice Massacci | Informal Purpose & DoI |
| 1.1 | October 26, 2024 | Alice Massacci | Activity 1: Personas & Scenarios definition |
| 1.2 | October 27, 2024 | Alice Massacci | Activity 2: Competency Questions definition |
| 1.3 | October 27, 2024 | Alice Massacci | Activity 3: Concepts Identification |
| 1.4 | October 30, 2024 | Alice Massacci | Purpose Definition completed |
| 2.0 | November 1, 2024 | Alice Massacci | Information Gathering - Data Sources |
| 2.1 | November 2, 2024 | Alice Massacci | Information Gathering - Data Sources and Resources Collection |
| 2.2 | November 3, 2024 | Alice Massacci | Information Gathering - Data Sources and Resources Collection |
| 2.3 | November 10, 2024 | Alice Massacci | Information Gathering - Resource Cleaning |
| 2.4 | November 11, 2024 | Alice Massacci | Information Gathering - Resource Cleaning |
| 2.5 | November 13, 2024 | Alice Massacci | Information Gathering - Resource Cleaning |
| 2.6 | November 14, 2024 | Alice Massacci | Information Gathering - Resource Cleaning |
| 2.7 | November 15, 2024 | Alice Massacci | Information Gathering completed |
| 3.0 | November 17, 2024 | Alice Massacci | Language Definition - Concept Identification |
| 3.0 | November 17, 2024 | Alice Massacci | Language Definition - Concept Identification |
| 3.1 | November 23, 2024 | Alice Massacci | Language Definition - Concept Identification |
| 3.2 | November 24, 2024 | Alice Massacci | Language Definition - Concept Identification |
| 3.3 | November 30, 2024 | Alice Massacci | Language Definition - Dataset Filtering |
| 4.0 | December 01, 2024 | Alice Massacci | Knowledge Definition |
| 4.1 | December 08, 2024 | Alice Massacci | Knowledge Definition - Protégé |
| 4.2 | December 10, 2024 | Alice Massacci | Knowledge Definition - Protégé |
| 5.0 | December 14, 2024 | Alice Massacci | Entity Definition - Entity Matching, Entity Identification |
| 5.1 | December 15, 2024 | Alice Massacci | Entity Definition - Entity Mapping |
| 5.2 | December 16, 2024 | Alice Massacci | Entity Definition - Entity Mapping |
| 5.3 | December 20, 2024 | Alice Massacci | Entity Definition - Entity Mapping completed |
| 6.0 | December 27, 2024 | Alice Massacci | Knowledge Layer Evaluation |
| 6.1 | December 30, 2024 | Alice Massacci | Data Layer Evaluation |
| 6.2 | January 4, 2025 | Alice Massacci | Data Layer Evaluation |
| 6.3 | January 5, 2025 | Alice Massacci | Query execution |
| 6.4 | January 6, 2025 | Alice Massacci | Query execution |
| 6.4 | January 25, 2025 | Alice Massacci | Query execution |
| 6.5 | January 26, 2025 | Alice Massacci | Query execution |
| 6.6 | January 31, 2025 | Alice Massacci | Query execution |
| 6.7 | February 1, 2025 | Alice Massacci | Query execution |
| 6.8 | February 2, 2025 | Alice Massacci | Query execution |
| 6.9 | January 8, 2026 | Alice Massacci | Query execution |
| 6.10 | January 9, 2026 | Alice Massacci | Query execution, Metadata Definition |
| 7.0 | January 10, 2026 | Alice Massacci | Metadata Definition |
| 7.1 | January 11, 2026 | Alice Massacci | Metadata Definition |
| 7.2 | January 17, 2026 | Alice Massacci | Open Issues |
| 7.3 | January 18, 2026 | Alice Massacci | Query execution |

# 1 Introduction

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

- Section 2: Definition of the project's purpose and its domain of interest.
- Section 3: Definition of the Information Gathering iTelos phase.
- Section 4: Description of the Language Definition iTelos phase.
- Section 5: Description of the Knowledge Definition iTelos phase.
- Section 6: Description of the Entity Definition iTelos phase.
- Section 7: Evaluation of the KG.
- Section 8: Description of the Metadata Definition.
- Section 9: Open Issues.

# 2 Purpose Definition

## 2.1 Informal Purpose

The **HealthRoute Trentino** project aims to develop a Knowledge Graph (KG) that provides seamless access to up-to-date information about Healthcare Facilities across the Trentino Province, along with their integration into the public transportation network. The KG will empower users—whether healthcare professionals or patients—to easily locate medical services (e.g., hospitals, pharmacies, medical centers) and discover the most efficient transit routes to access them. The system aims to streamline healthcare access and improve overall coordination between health services and transportation options.

The informal purpose can be stated as: *"A KG that helps users quickly find healthcare facilities/services in Trentino and determine how to reach them efficiently using public transportation, based on their location and transport data."*

The KG should be capable of answering complex queries like "Which hospital can I reach by bus within 30 minutes from my current location?" or "Which pharmacies are accessible via public transport in the evening?" The KG will integrate static transportation data from different sources with healthcare facilities and synthetic user data, and will serve as a foundation for web or smartphone applications that assist users in planning trips to healthcare facilities based on real-time schedules, transportation options, and user location.

## 2.2 Domain of Interest (DoI)

The DoI for this project is bounded in space and time:

- Space: autonomous province of Trentino.
- Time: To ensure practical applicability, the project will utilize current public transportation data from the year 2024.

These boundaries ensure that the project operates within a realistic and usable framework for public transportation information.

## 2.3 Purpose Definition – Activities

Once the informal purpose and DoI are defined, the iTelos methodology moves to formalization of the initial purpose statement into a purpose-specific Entity-Relationship (ER) model. Purpose formulation involves four key Activities:

- Personas & Scenarios definition
- Competency Questions definition
- Concepts Identification
- ER modeling

### 2.3.1 Activity 1: Personas & Scenarios definition

To guide the design and development of the KG, a set of personas—fictional representations of different end-user types with specific needs and objectives—and scenarios have been defined. These scenarios cover a variety of transportation and healthcare-related challenges, ensuring coverage of rush hour demands, working days and holidays. By framing the project through these user-centered examples, the practical applications of the HealthRoute Trentino KG are demonstrated.

## Scenarios

**Working Day (Monday to Friday) – Morning (S1)**
On a typical weekday morning in Trento, between 7 a.m. and 12 p.m., the city is bustling with activity as residents commute to work and students head to school. The public transport system runs at full capacity. Buses, trams, and trains operate frequently to accommodate the surge in passengers, with most routes servicing both urban and suburban areas. Healthcare workers–doctors, nurses, and administrative staff–rely heavily on public transport to reach hospitals, clinics, and medical centers on time to start their shifts. Patients, especially those with scheduled appointments, also need reliable transportation.

**Nighttime Travel (S2)**
It's late at night in Trento, between 12 a.m. and 6 a.m. The city is quiet, with most residents at home and businesses closed, except for emergency services like hospitals. Public transportation services are limited, with only a few night buses running. Passengers needing to travel during these hours, such as night-shift workers, may

experience longer waiting times and fewer route options.

**Weekend Reduced Service (S3)**
On weekends, especially Sundays, Trento slows down as businesses close. Public transportation runs on reduced schedules, with fewer buses and trains in operation. Hospitals and essential services remain open.

**Heavy Rain (S4)**
It's raining heavily, and streets are slick with water, making outdoor travel less appealing. Fewer people are walking outside, opting instead for the warmth and shelter of public transportation. Buses are more crowded than usual. The weather conditions also slow down traffic. Passengers face longer waiting times and a slower commute.

**Strike Day (S5)**
On strike days, both buses and trains can be delayed or cancelled altogether, creating significant disruptions for commuters. People who rely solely on public transportation face difficulties, as they might be unable to get to work, miss important meetings, or face long delays.

**Traffic Disruption Due to Construction Work (S6)**
Ongoing construction work on a major road in Trento has caused significant delays throughout the city. Public buses are forced to take alternative routes, resulting in longer travel times for passengers who might miss connections or arrive late at their destinations. The detours also add extra pressure on already busy streets.

## Personas

**Stefania (P1)**: Stefania is a 36-year-old social worker living in Borgo Valsugana. She frequently visits nursing homes and hospitals as part of her job. Stefania also has to manage her own health, picking up prescriptions for thyroid medication and visiting her endocrinologist in Trento.

**Riccardo (P2)**: Riccardo is a 45-year-old IT consultant from Pergine Valsugana. After suffering a minor heart attack a year ago, he became committed to improving his health through regular cardiovascular check-ups and a strict fitness routine. He prefers using public transportation to reduce stress and improve his health further, often combining his trips with walks in the local parks. He also suffers from seasonal allergies and regularly visits the local pharmacy during spring and summer to buy antihistamines and nasal sprays. Riccardo often schedules his trips to the pharmacy for early morning to avoid the intense heat in the summer.

**Elena (P3)**: Elena is a 38-year-old teacher living in Rovereto. She has asthma and needs regular check-ups with a pulmonologist. When the weather is favorable, Elena uses her scooter to travel to her appointments. However, during colder months or when her asthma flares up, she relies on public transportation for her commutes. Her appointments are often scheduled for the early morning, during the busy weekday rush hours.

**Chiara (P4)**: Chiara is a 29-year-old nurse working at a hospital in Trento. She is often on rotating shifts, including nights and weekends. Her day shifts run from 7 am to 3 pm, while night shifts run from 11 pm to 7 am. In addition to her regular duties, Chiara is part of a professional development program that requires her to travel to different healthcare facilities across the region. She mentors new nurses and attends training sessions and seminars to stay updated on the latest medical practices. Chiara often relies on public transport for her commutes.

**Edoardo (P5)**: Edoardo is a 50-year-old man who lives in Caldonazzo and works in Trento. Due to his job as a restaurant manager, Edoardo often works late into the night. He has chronic insomnia and sometimes needs to visit the pharmacy after work to get over-the-counter sleep aids. Since he doesn't drive, Edoardo depends on late-night bus services and 24-hour pharmacies to manage his condition without having to take time off work.

### 2.3.2 Activity 2: Competency Questions (CQs) definition

- **CQ-1 (P1-S5)**: Stefania has a busy Monday morning with back-to-back appointments at two nursing homes, but a public transport strike has been announced.
  a) Which train/bus lines does she rely on for her appointments?
  b) Are they affected by the strike?

- **CQ-2 (P1-S1)**: Stefania has a morning visit scheduled at a nursing home in Trento at 11:00 a.m.
  a) What is the most efficient public transportation route for her to arrive at the nursing home on time?
  b) Is there a pharmacy within 500 meters of the final bus stop?
  c) If she leaves home at 9:30 a.m., will she have sufficient time to stop by the pharmacy to buy her thyroid medications before proceeding to her appointment?

- **CQ-3 (P2-S1)**: I'ts Wednesday, Riccardo has a cardiovascular check-up scheduled for 9:00 a.m. at Ospedale Santa Chiara and plans to pick up some antihistamines beforehand.
  a) How many pharmacies are located within a 5-10 minute walk from each bus stop on his route to the hospital, and which one would be the most convenient to stop at?
  b) If he leaves home at 8:30 a.m., will he arrive on time for his appointment?

- **CQ-4 (P2-S3)**: After enjoying a walk in Parco delle Albere on a sunny Sunday afternoon, Riccardo begins experiencing allergy symptoms and needs to buy antihistamines.
  a) Is there a nearby pharmacy open on Sundays?
  b) Given the reduced weekend public transport service, what is the fastest route he can take to reach the pharmacy and then continue home?

- **CQ-5 (P3-S6)**: Elena has a morning appointment with her pulmonologist at Ospedale Santa Chiara in Trento. Due to ongoing construction on the train line between Rovereto and Trento, she must rely on replacement bus services.
  a) What is the best route for her to take?
  b) Which is the travel time using the replacement bus?

- **CQ-6 (P4-S6)**: Chiara finishes her shift at Ospedale S. Camillo at 3:00 p.m. and needs to travel to Centro Medico di Rovereto for a training session at 4:30 p.m., but there is ongoing construction causing delays on the usual bus route. What alternative bus routes can she use to reach the training center on time despite the disruption?

- **CQ-7 (P4-S4)**: Chiara is finishing her night shift at the hospital at 7:00 a.m. on a rainy weekday.
  a) What public transportation options are available for her to get home after her shift?
  b) If the bus arrives but is overcrowded due to the heavy rain, how long will she have to wait for the next available bus?

- **CQ-8 (P5-S2)**: Edoardo finishes his late shift at the restaurant at 1:00 a.m. and needs to stop by a 24-hour pharmacy to pick up sleep aids. Given the limited late-night bus service, what is the quickest way for him to reach the nearest open pharmacy and then head home to Caldonazzo?

### 2.3.3 Activity 3: Concepts Identification

Considering the scenarios and personas involved in the CQs, the following entities and their properties can be identified:

| Scenarios | Personas | CQs | Entities | Properties | Focus |
|---|---|---|---|---|---|
| 1–6 | 1–5 | 1–8 | End_User | id, name, type | Contextual |
| 1–6 | 1–5 | 1–8 | Trip | id, start_time, end_time, start_point, end_point, route | Contextual |
| 1–6 | 1–5 | 1–8 | Route | id, type, length, schedule | Core |
| 1–6 | 1–5 | 1–8 | Position | id, address, latitude, longitude | Contextual |
| 1–6 | 1,2 | 2b,3a | Stop | id, name, coordinates, arrival_time, departure_time | Core |
| 1–4 | 1,2,4,5 | 1a,2,3,4,7,8 | Weekly_Schedule | id, schedule | Core |
| 5,6 | 1,3,4 | 1b,5,6 | Schedule_Exception | id, date, exception_type, affected_routes | Core |
| 5,6 | 1,3,4 | 1b,5,6 | Event | id, type, date | Core |
| 1–6 | 1–5 | 1–8 | Health_Facility | id, name, coordinates, type, access_schedule | Core |
| 1,5,6 | 1–7 | 1,2,3,5c,6 | Appointment | id, user, place, time | Contextual |

Table 1: Purpose Formalization sheet

### 2.3.4 Activity 4: ER modeling

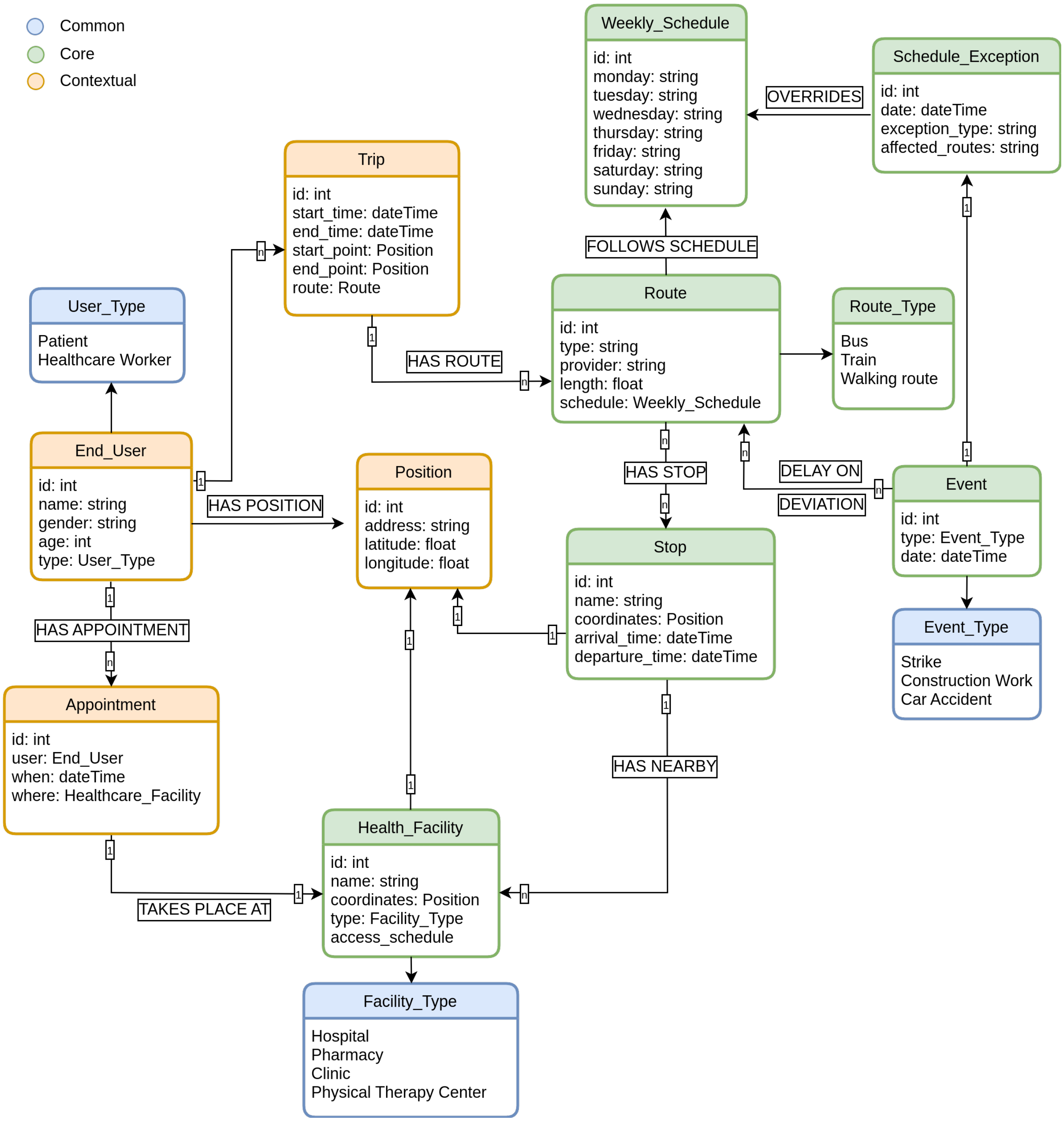



Figure 1: ER model

Each EType in the ER model has specific attributes that provide information relevant to answering CQs and fulfilling the KG's purpose.

The **End_User** EType captures individual user data and categorizes users into distinct **User_Type** groups—patients and healthcare professionals. **End_Users** are linked to the **Appointment** EType, which is key to aligning travel routes with users' healthcare schedules, supporting the KG's goal of matching users with timely travel options to and from their appointments.

Transportation-related ETypes—**Trip**, **Route**, and **Stop**—model the transportation network. A **Trip** defines a single journey taken by an **End_User** from a start point to a destination, and is served by one or more **Routes** that define the specific transit paths and schedules. This design supports queries like finding the nearest healthcare facility within a set travel time. The **Stop** EType defines precise locations along a route where users can board or exit. Each **Stop** has coordinates, scheduled arrival and departure times, and supports queries about where users can access transport, travel times, and nearby healthcare facilities.

The **Weekly_Schedule** EType defines the recurring weekly timetable for transportation services. By associating each route with its **Weekly_Schedule**, the KG can provide users with detailed timing and frequency information. Real-world schedules are subject to disruptions, and **Schedule_Exception** captures temporary changes to a regular schedule, such as delays or cancellations, caused by external occurrences (**Event** EType) such as traffic accidents, road closures, strikes. This structure allows the model to answer queries about potential disruptions.

Finally, healthcare service locations are managed through the **Health_Facility** EType, which catalogs healthcare access points—hospitals, clinics, and pharmacies—with data properties (attributes) for coordinates, facility type, and open hours.

Object properties (relationships) between ETypes have also been established. Key relationships include **HAS ROUTE** (linking a **Trip** to its **Route/s**), **FOLLOWS SCHEDULE** (associating **Route** with a **Weekly_Schedule**), and **HAS STOP** (connecting a **Route** to its **Stop** locations). Each relationship has defined cardinalities to specify how many instances of the two entities are involved. For instance, the **HAS STOP** relationship between **Route** and **Stop** typically has a many-to-many cardinality, as each route has multiple stops, and each stop may serve multiple routes. The **HAS APPOINTMENT** relationship, instead, exhibits a one-to-many cardinality, whereby an **End_User** can have multiple appointments, while each **Appointment** is associated with a single user. Defining these cardinalities ensures that the model accurately represents the complexity of real-world entities and can effectively processes user queries.

# 3 Information Gathering

## 3.1 Data Sources and Resources Collection

This section contextualizes each data source's role in the project.

The primary source of public transportation data for this project was the Trentino Transporti website. Through its Open Data section, Trentino Trasporti provides curated datasets covering both urban and extra-urban transportation following the General Transit Feed Specification (GTFS) standards. This adherence to GTFS ensures that data on bus stops, routes, and timetables is consistent and well-structured, supporting easy processing and meeting the iTelos methodology's requirements for interoperability. The latest update to these datasets was on September 09, 2024, covering the period from September 09, 2024 to June 12, 2025. This ensures that the KG is built upon the most current information available.
Public transportation datasets collected by project teams from previous academic years and stored within the LiveData catalogs were carefully reviewed to assess their relevance and potential for reuse. However, data was limited to the years 2022-2023, and could not fully meet the current project's needs for up-to-date transit information.

Data on Trentino region's pharmacies and para-pharmacies was initially sought from the 2022 Trentino Healthcare Connectivity Project. The datasets, sourced from OPENdata Trentino (dataset identifiers: *apss_farmacie-pat* and *apss_parafarmacie-pat*), covered the periods from January 1, 2021, to December 31, 2021, and from January 1, 2017, to December 31, 2017, respectively. Given the outdated nature of this data, it was deemed unsuitable for the current project's objectives. OPENdata Trentino references the Ministero della Salute's Open Data system as the original source of these datasets. Therefore, an attempt was made to access the Ministero della Salute's Open Data directly in the hope to acquire more up-to-date information. A couple of datasets were identified that met the project's requirements:

- FRM_FARMA provides a list of pharmacies across the entire national territory, covering branches, dispensaries, and seasonal dispensaries. Specifically, it provides detailed administrative data for each pharmacy (e.g. Ministerial Identification Code, name, VAT number, local health authority (ASL) pharmacy code) and precise location details, including the full address (street, postal code, district, municipality, province, and region). For further details, a dictionary of the dataset's fields is available for reference. The dataset is provided in JSON format and is updated weekly. At the time of retrieval for this project, the latest update was on November 6, 2024.

- FRM_PFARMA provides a complete list of businesses, other than pharmacies, authorized to sell medications to the public, along with administrative and location data. For more information about the dataset, a dictionary is available for reference. The dataset is provided in JSON format and is updated daily. At the time of retrieval for this project, the latest update was on November 6, 2024.

The FRM_FARMA and FRM_PFARMA datasets were both deemed suitable for the project, as they provide not only precise addresses and geographical coordinates (latitude and longitude), but also include the key fields "data_inizio_validità" (start date of validity) and "data_fine_validità" (end date of validity). These fields enable filtering to include only currently active businesses, ensuring that the KG can support users in identifying nearby pharmacies and para-pharmacies based on up-to-date operational status.

OPENdata Trentino provides a dataset of hospitals within the municipality of Trento, available in both CSV and GeoJSON formats. The dataset was last updated on October 18, 2023, and contains precise location coordinates, institution names, facility types, and full addresses. Given that hospitals are typically stable institutions with minimal turnover, it is reasonable to assume that the listed facilities remain operational and relevant for the current project's objectives.

The 2022 Trentino Healthcare Connectivity Project has collected and processed a dataset on Trentino's public and accredited healthcare facilities, sourced from OPENdata Trentino under the identifier *apss_strutture-sanitarie-dell-azienda-sanitaria-e-convenzionate*. The dataset covers a timeframe from December 1, 2017, to December 31, 2017. Given the age of this data, its relevance and accuracy for the current project requires careful consideration. Although healthcare facilities tend to have low turnover and often remain operational over long periods, significant updates or structural changes may have occurred since 2017. Therefore, while this dataset remains available for reference, efforts were made to identify up-to-date resources.
An alternative, more suitable source was identified in the *Aziende Ospedaliere, Aziende Ospedaliere Universitarie e IRCCS pubblici* dataset available through the Ministero della Salute's Open Data system. This dataset provides a list of healthcare facilities—including hospital trusts, university hospital trusts, and public IRCCS (Istituti di Ricovero e Cura a Carattere Scientifico)—actively operating across Italy as of January 1, 2024. For detailed information on the dataset, please refer to the accompanying data dictionary.

Data on healthcare facilities were ultimately sourced from the Humanitarian Data Exchange (HDX) platform, specifically from the Italy Healthsites dataset within the Geodata Datasets section. This dataset is part of the Healthsites.io initiative, which aims to build an open data commons of health facility data with OpenStreetMap. The dataset is available in GeoJSON format and includes records on hospitals, clinics, pharmacies, medical offices, and dental practices throughout Italy. It provides key attributes such as facility type, location, and opening hours. Designed to support humanitarian efforts and improve accessibility to healthcare infrastructure data, this resource is regularly updated and validated through contributions from local and international health organizations. At the time of retrieval for this project, the latest update was on February 8, 2024.

GeoJSON data was download from the openpolis/geojson-italy GitHub repository which contains geo-referenced limits for all municipalities in Italy. Files are upgraded periodically, and refer to the latest administrative subdivisions, as published by ISTAT. Each municipality's boundary is represented as a polygon or multipolygon, defining the precise shape and size of its geographic area.

## 3.2 Resource Cleaning and Formatting

Once the relevant datasets were identified, the next step was to ensure their suitability for the construction of the KG by cleaning and formatting the data. This process involved several key steps, such as removing duplicate records, outdated data points or irrelevant attributes for the purpose, standardizing field names and values, and ensuring consistent formatting across all datasets. Data anonymization was not required, as all the data used was sourced from publicly accessible platforms, and it is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

Each dataset was prepared to address the CQs defined during Purpose Formalization, ensuring they serve the information needs of scenarios. For reusability and compliance with iTelos standards, all resources have been formatted according to the CSV standard.

In accordance with the FAIR principles, access to both the original and processed datasets, as well as the Python scripts used for data processing, is provided through the project GitHub repository.

- ***Public Transportation Data:*** Urban and extra-urban transportation data from Trentino Trasporti Open Data was well-organized and fully compliant with the General Transit Feed Specification (GTFS) standard. The data was already aligned with best practices for transit data interoperability. As a result, no preprocessing or data manipulation was necessary. This RMarkdown notebook offers an overview of the files included in the dataset, providing insight into their structure and content.

  Train service data was unstructured (i.e. timetables in PDF format). Manual extraction was necessary to convert the data into a usable and machine-readable format for analysis. This process involved reviewing the timetables, identifying relevant information such as train schedules and station stops, and manually entering the data into a structured format suitable for further processing. Google Maps service was used to geolocate the train lines of interest. Due to the time-intensive nature of the manual extraction process, the final train service dataset has not yet been completed as of the writing of this report. Upon completion, it will be made available in the GitHub repository, along with the accompanying processing code.

  *Update Note: The manual extraction and processing of the train service data have been successfully completed. The final dataset is available in the GitHub repository.*

- ***FRM_FARMA and FRM_PFARMA:*** The data cleaning process involved filtering the records to retain only those related to the 'PROV. AUTON. TRENTO' region. After isolating these entries, irrelevant fields such as 'cod_farmacia_asl' and 'p_iva' were removed to streamline the dataset. Address information was standardized by splitting the 'indirizzo' field at the first comma, separating the street name and house number and assigning each to distinct fields. If no house number was provided, 'None' was used to indicate the absence of data. The 'indirizzo' field displayed inconsistent casing, with some entries in uppercase and others in lowercase. For consistency and ease of use, this field was standardized to title case. Coordinates were standardized by replacing commas with periods.

  Each pharmacy and para-pharmacy in the dataset is uniquely identified by an ID, yet it is common to find multiple records with the same ID, sometimes repeated dozens of times. These duplicates only vary in their validity period, defined by the 'data_inizio_validità' and 'data_fine_validità' fields. It appears that, rather than updating an existing record when a pharmacy renews its validity period, a new record is added to the

dataset. Outdated entries were filtered out, leaving a single valid record for each ID.
'None' was used for missing data in the dataset.

- ***Aziende Ospedaliere, Aziende Ospedaliere Universitarie e IRCCS pubblici:*** Upon processing the dataset, it became clear that no entries corresponded to the 'Prov. Auton. Trento' region. The dataset dictionary does include a region code (i.e., 042) specifically assigned to the Province of Trento, suggesting that data for this region was intended to be included but ultimately missing. The complete absence of relevant records for the targeted region makes the dataset unusable for the project's objectives.

- ***Italy Healthsites:*** To narrow down the GeoJSON data to healthcare facilities specifically located within Trentino, a spatial filter was applied. The filter relied on a separate GeoJSON boundary file of Trentino-Alto Adige, and retained only the entries that fell within the specified polygonal geographic boundaries. Before applying the filter, both the healthcare facilities data and boundary geometry were aligned to the same coordinate reference system (CRS). Further filtering was applied to isolate municipalities within the 'Provincia Autonoma di Trento'. A list of Trentino municipalities, sourced from Wikipedia, guided this process, allowing the dataset to retain only the relevant municipalities for targeted analysis. To streamline the dataset, several non-essential fields were dropped.

  The dataset includes records for hospitals, pharmacies, clinics, and dental offices. Many pharmacy entries overlap with those found in the Ministero della Salute datasets, FRM_FARMA and FRM_PFARMA. Since these datasets are updated weekly and feature a 'data_fine_validità' field, they are reliable for maintaining an up-to-date list of active businesses, making them preferable as primary sources. However, the geographic coordinates (latitude and longitude) in FRM_FARMA and FRM_PFARMA are sometimes imprecise.
  For instance, in the FRM_FARMA dataset, the coordinates for *Farmacia Grandi* at *Via Alessandro Manzoni 7A, Trento*, are given as 46.0748132409383, 11.1251346567747 (marked by the red pin in Figure 2). Meanwhile, the coordinates from the Healthsites dataset are 46.073819833990825, 11.125191578971851 (marked by the green pin). As shown in Figure 2, the green pin closely matches the pharmacy's actual location, whereas the red pin is a relatively rough approximation. This pattern holds for many pharmacies in the dataset, with similar discrepancies observed across many entries. Given the recurring issue, it was decided to prioritize the Healthsites dataset's location data for spatial accuracy, while still relying on FRM_FARMA and FRM_PFARMA for information on the active status and operational details of each pharmacy. This combined approach aimed to combine the precise geographical positioning from Healthsites with the regularly updated and reliable business data from the Ministero della Salute sources, ultimately improving the dataset's overall quality for downstream applications.

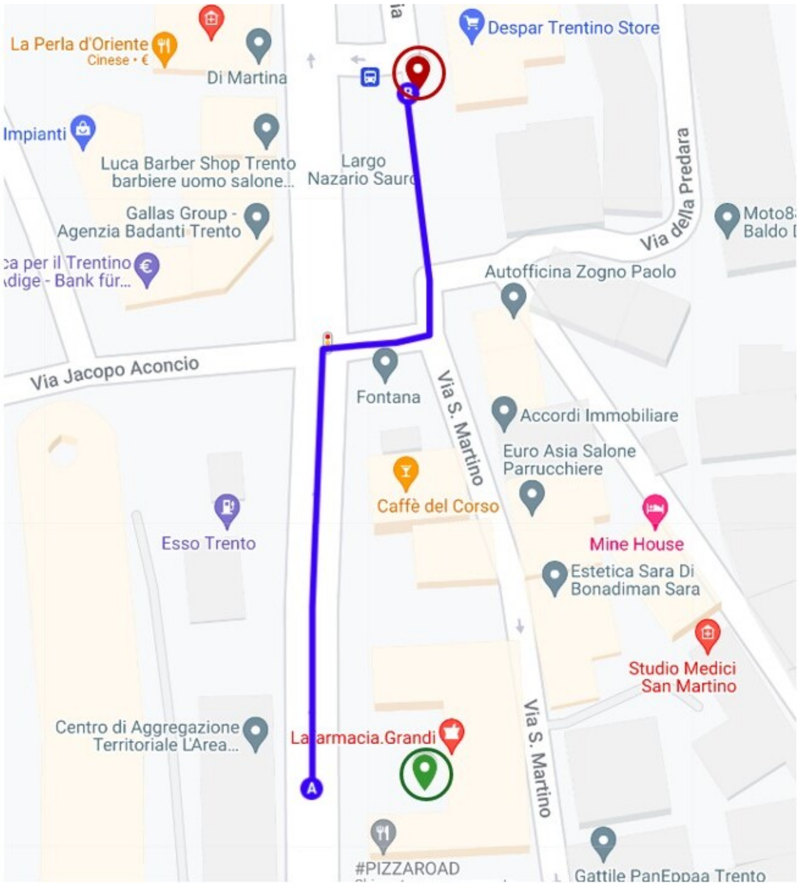



Figure 2: Discrepancies in spatial coordinates between Healthsites and Ministero della Salute datasets

The challenge, however, lies in integrating the two sources. Matching by pharmacy name proves unreliable due to inconsistencies: in the Ministero della Salute datasets, the pharmacy name is theoretically derivable from the 'descrizione_farmacia' field, intended to include the *ragione sociale* according to the data dictionary. However, in practice, this field varies widely—sometimes it contains only the pharmacy name, sometimes only the owner's name, and occasionally both. Merging by address is also impractical, as the Healthsites dataset often lacks address data, and address formats differ across the datasets (e.g., *Piazza Dante Alighieri* and *Piazza D. Alighieri*).

To address the inconsistencies, a geospatial matching approach was implemented, based on proximity rather than exact matches on name or address. Coordinate columns or WKT geometry were transformed into spatial point data (any non-Point geometries like polygons were converted to centroids), and a 50-meter buffer zone was defined around each point in the Healthsites dataset. Any point from the Ministero della Salute datasets falling within the search radius was then treated as a potential match, suggesting likely correspondence to the same pharmacy. A KD-Tree Nearest Neighbor Search was used to find points in close proximity (*cKDTree* from *scipy.spatial*).

Following the initial identification of potential matches based on spatial proximity, additional criteria were applied to refine these matches. When pharmacy names were available in both datasets, a fuzzy matching algorithm (using the *FuzzyWuzzy* Python library) assessed name similarity. Where address information was present, fuzzy matching was also applied to the address fields. Finally, a combined metric incorporating spatial distance, name similarity, and address similarity, was used to evaluate the quality of each match.

During processing, it became clear that the spatial coordinates in the Ministero della Salute datasets were more imprecise than anticipated. As a result, the buffer radius had to be increased from the initial 50 meters to 10 kilometers to ensure potential matches were captured. Ultimately, the maximum distance observed between two pharmacies identified as a match reached 6.54 km! (Figure 3).

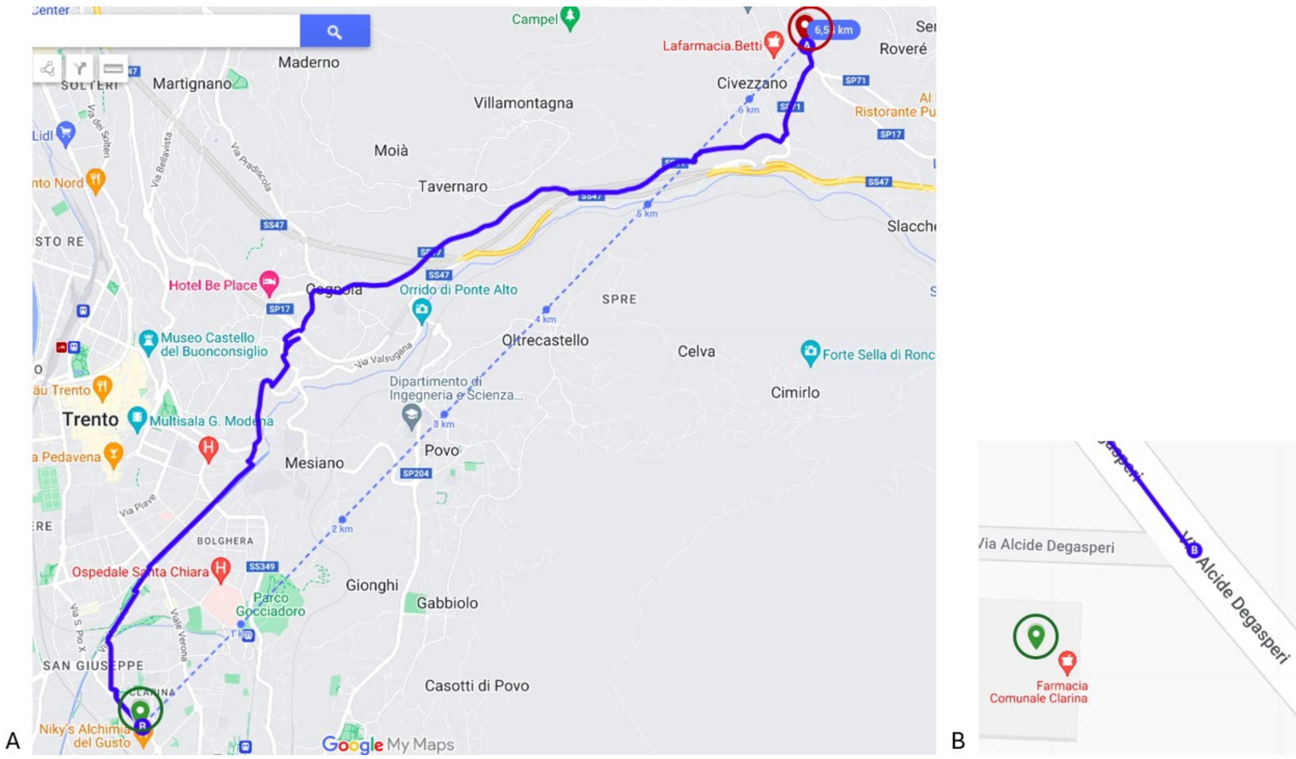



Figure 3: A) The red pin marks the Ministero della Salute coordinates for "Comunale N.4 Clarina" at Via De Gasperi 112, 38123, Trento, while the green pin marks the Healthsites coordinates for "Farmacia Clarina" at Via Alcide De Gasperi 112, 38123, Trento. Despite the differences in name and address, there is high confidence that these entries refer to the same pharmacy. B) The green pin aligns with the pharmacy's location on Google Maps, whereas the red pin, positioned 6.54 km away in a different postal code (38045), reflects the high imprecision in the Ministero della Salute dataset.

Out of 94 Healthsites records that included a full address, 74 matched entries in the Ministero della Salute datasets. All matches were manually reviewed.

In conclusion, the final integrated dataset comprises entries from FRM_FARMA and FRM_PFARMA, with Healthsites' more accurate spatial data replacing the original coordinates wherever possible. Additionally, opening hours from the Healthsites dataset were included in the entries when available.

Following initial processing, records for hospitals, clinics, and dental offices from the Healthsites dataset required no further refinement.

Route transfer entities were not directly available in the original GTFS and auxiliary datasets and were therefore derived programmatically. First, all bus and train stops, trips, and stop times were merged into a unified transportation graph. A directed multigraph was constructed in which nodes represent public transport stops and directed edges represent consecutive stop pairs along a trip, weighted by travel time derived from scheduled arrival and departure times. Second, potential walking-based transfers between nearby stops were generated based on spatial proximity. All stops with valid geographic coordinates were pairwise compared, and for each pair of stops within a predefined walking radius (300 meters), bidirectional walking edges were added to the graph. The walking time was estimated from geodesic distance using an average walking speed. These walking edges represent potential physical transfers between routes at different but nearby stops. For each walking edge, the associated routes serving the origin and destination stops were identified. Each route-to-route combination was materialized as a RouteTransfer entity, explicitly linking the arrival stop on the first route, the departure stop on the second route, and the corresponding walking distance.

# 4 Language Definition

The Language Definition phase of the iTelos methodology addresses the challenges of language heterogeneity by establishing the formal concepts required to represent the information in the Knowledge Graph. This phase is pivotal in creating a consistent and unambiguous domain language tailored to the KG's purpose, facilitating integration of heterogeneous data sources and supporting robust query capabilities.
The Language Definition phase comprised two key activities:

- Concept Identification
- Dataset Filtering

## 4.1 Activity 1: Concept Identification

This activity involved identifying all relevant concepts to be used to represent the information in the final KG. These concepts included ETypes, data properties, and object properties, which had been outlined during the Purpose Formalization phase and documented in the Purpose Formalization Sheet and ER model.
The identified concepts were then aligned with the Universal Knowledge Core (UKC). If a concept already existed in the UKC, it was directly adopted for its broad reusability and established formalization. If the concept was absent from the UKC, it was searched for in other ontologies or vocabularies (e.g., Schema.org, General Transit Feed Specification). When existing concepts were either too general or absent, new concepts were defined to address the project's specific context. For instance, *Weekly Schedule* and *Schedule Exception* were introduced to address the periodicity of public transit schedules and deviations from them, respectively, filling gaps not covered by existing definitions in the UKC.
The formalized concept definitions resulting from the UKC alignment process are presented in Tables 2, 3, and 4, with each entry uniquely identified and accompanied by a precise label and gloss.

**Note to the Reader:** In Table 3, references to specific concepts from the General Transit Feed Specification (GTFS) are provided in the *Language_Resources.html* file available on the Project's GitHub repository. In this document, only a general link to the GTFS Reference is provided, as LaTeX does not support URLs with text highlights.

### Language concepts for ETypes

| Concept ID | Language Concept Word | Gloss |
|---|---|---|
| UKC-49456 | end user | The ultimate user for which something is intended. |
| KGE24-4-0014* | start position | A defined beginning location from which an event, activity, or process is initiated. |
| KGE24-4-0015* | end position | A defined final location at which an event, activity, or process concludes. |
| UKC-20121 | health facility | Building where medicine is practiced. |
| KGE24-4-0001* | health appointment | A scheduled meeting with a healthcare provider for consultation, treatment, or check-up. |
| UKC-27518 | position | The spatial property of a place where or way in which something is situated. |
| UKC-45485 | route | An established line of travel or access. |
| UKC-45117 | bus route | The route regularly followed by a passenger bus. |
| KGE24-4-0002* | train route | The route regularly followed by a passenger train. |
| UKC-1484 | trip | A journey for some purpose (usually including the return). |
| UKC-5348 | stop | A brief stay in the course of a journey. |
| UKC-45118 | bus stop | A place on a bus route where buses stop to discharge and take on passengers. |
| KGE24-4-0003* | train stop | A place on a train route where trains stop to discharge and take on passengers. |
| gtfs:StopTime | stop time | Describes a stop time as part of a trip. |
| KGE24-4-0004* | weekly schedule | An ordered timetable outlining activities or operations for each day of the week. |
| KGE24-4-0005* | schedule exception | An instance that does not conform to a schedule. |
| KGE24-4-0006* | service rule | A regulation that defines public transport schedules. |
| KGE24-4-0017* | transport option | A transport service available at a specific public transport stop. It represents the possibility for a passenger to access a route or trip from that stop. |
| KGE24-4-0021* | route transfer | A transfer opportunity between two public transport routes. It encodes the possibility for a passenger to switch from one route to another, including the walking path and associated metadata. |
| UKC-17277 | pharmacy | A retail shop where medicine and other articles are sold. |
| UKC-18969 | hospital | A health facility where patients receive treatment. |
| UKC-16098 | clinic | A healthcare facility for outpatient care. |
| KGE24-4-0007* | physical therapy center | A facility where patients receive treatment to restore movement, reduce pain, and improve physical function. |
| KGE24-4-0008* | dental clinic | A facility where dental care services, such as checkups, cleanings, and treatments, are provided by dentists and oral health professionals. |

| | | |
|---|---|---|
| schema:OpeningHours | opening hours | The general opening hours for a business. |
| UKC-51562 | patient | A person who requires medical care. |
| KGE24-4-0009* | healthcare professional | A trained individual licensed to provide medical, nursing, or therapeutic care to patients. |
| UKC-56 | event | Something that happens at a given place and time. |

Table 2: In this table is presented how ETypes have been formalized and assigned to a UKC concept. Concept IDs marked with a " * " are new concepts specifically created for this project.

### Language concepts for ETypes attributes (Data Properties)

| **Data Property Name** | **Concept ID** | **Language Concept Word** | **Gloss** |
|---|---|---|---|
| **End User** | | | |
| has_user_id | UKC-38423 | identifier | A symbol that establishes the identity of the one bearing it. |
| has_user_first_name | UKC-33531 | first name, given name, forename | The name that precedes the surname. |
| has_user_type | UKC-31362 | type | A subdivision of a particular kind of thing. |
| has_user_health_appointment | KGE24-4-010* | health appointment | A scheduled meeting with a healthcare provider for consultation, treatment, or check-up. |
| **Health Facility** | | | |
| has_health_facility_legal_name | Schema.org | legalName | The official name of the organization, e.g. the registered company name. |
| has_health_facility_type | UKC-31362 | type | A subdivision of a particular kind of thing. |
| **Health Appointment** | | | |
| has_health_appointment_id | UKC-38423 | identifier | A symbol that establishes the identity of the one bearing it. |
| has_health_appointment_place | UKC-66454 | place | Proper or appropriate position or location. |
| has_health_appointment_date | UKC-73013 | date | The specified day of the month. |
| has_health_appointment_time | UKC-38632 | time | An instance or single occasion for some event. |
| **Position** | | | |
| has_position_addr_street | UKC-45008 | street address | The address where a person or organization can be found. |
| has_position_addr_housenumber | KGE24-4-0011* | house number | A number identifying a building on a street, used in postal addresses. |
| has_position_addr_postcode | UKC-33635 | zip code, zip, postcode, postal code | A code of letters and digits added to a postal address to aid in the sorting of mail. |
| has_position_municipality | UKC-45537 | municipality | An urban district having corporate status and powers of self-government. |

| | | | |
|---|---|---|---|
| has_position_latitude | UKC-45423 | latitude | The angular distance between an imaginary line around a heavenly body parallel to its equator and the equator itself. |
| has_position_longitude | UKC-45429 | longitude | The angular distance between a point on any meridian and the prime meridian at Greenwich. |
| **Route** | | | |
| has_route_id | General Transit Feed Specification Reference | route_id | Identifies a route. |
| has_route_type | General Transit Feed Specification Reference | route_type | Indicates the type of transportation used on a route. Valid options are: 0 - Tram, 1 - Subway, etc. |
| has_route_short_name | General Transit Feed Specification Reference | route_short_name | Short name of a route. Often a short, abstract identifier. |
| has_route_long_name | General Transit Feed Specification Reference | route_long_name | Full name of a route. This name is generally more descriptive than the route_short_name. |
| **Trip** | | | |
| has_trip_id | General Transit Feed Specification Reference | trip_id | Identifies a trip. |
| has_trip_headsign | General Transit Feed Specification Reference | trip_headsign | Text that appears on signage identifying the trip's destination to riders. |
| has_trip_direction | General Transit Feed Specification Reference | trip_direction | Indicates the direction of travel for a trip. Valid options are: 0 - Travel in one direction (e.g. outbound travel). 1 - Travel in the opposite direction (e.g. inbound travel). |
| **Stop Time** | | | |
| has_stop_time_id | UKC-38423 | identifier | A symbol that establishes the identity of the one bearing it. |
| has_stop_time_arrival_time | UKC-73119 | arrival_time | The time at which a public conveyance is scheduled to arrive at a given destination. |
| has_stop_time_departure_time | UKC-73120 | departure_time | The time at which a public conveyance is scheduled to depart from a given point of origin. |
| has_stop_time_stop_sequence | General Transit Feed Specification Reference | stop_sequence | Order of stops for a particular trip. The values must increase along the trip but do not need to be consecutive. |
| **Stop** | | | |
| has_stop_id | General Transit Feed Specification Reference | stop_id | Identifies the serviced stop. A stop may be serviced multiple times in the same trip, and multiple trips and routes may service the same stop. |
| has_stop_name | General Transit Feed Specification Reference | stop_name | Name of the location. The stop_name should match the agency's rider-facing name for the location as printed on a timetable, published online, or represented on signage. |
| has_stop_lat | General Transit Feed Specification Reference | stop_lat | Latitude of the location. |

| | | | |
|---|---|---|---|
| has_stop_lon | General Transit Feed Specification Reference | stop_lon | Longitude of the location. |
| **Weekly Schedule** | | | |
| has_weekly_schedule_id | UKC-38423 | identifier | A symbol that establishes the identity of the one bearing it. |
| has_weekly_schedule_start_date | General Transit Feed Specification Reference | start_date | Start service day for the service interval. |
| has_weekly_schedule_end_date | General Transit Feed Specification Reference | end_date | End service day for the service interval. This service day is included in the interval. |
| has_weekly_schedule | UKC-34204 | schedule | An ordered list of times at which things are planned to occur. |
| **Schedule Exception** | | | |
| has_schedule_exception_id | UKC-38423 | identifier | A symbol that establishes the identity of the one bearing it. |
| has_schedule_exception_type | UKC-31362 | type | A subdivision of a particular kind of thing. |
| has_schedule_exception_date | UKC-73013 | date | The specified day of the month. |
| **Opening Hours** | | | |
| has_opening_hours_id | UKC-38423 | identifier | A symbol that establishes the identity of the one bearing it. |
| has_opening_hours_weekday | UKC-73013 | | The specified day of the week. |
| has_opening_hours_start_time | UKC-38632 | time | An instance or single occasion for some event. |
| has_opening_hours_end_time | UKC-38632 | time | An instance or single occasion for some event. |

Table 3: In this table is presented how ETypes properties (first column) have been formalized and assigned to a UKC concept (second and third columns). The last column provides a description of what the concept means. Concept IDs marked with a " * " are new concepts specifically created for this project.

### Language concepts for ETypes relations (Object Properties)

| Object Property | Relationship | Language Concept Word | Concept ID | Gloss |
|---|---|---|---|---|
| has_appointment | End User - Appointment | schedule | UKC-95928 | Plan for an activity or event. |
| has_position | End User - Position; Health Facility - Position; Stop - Position | located, has position | UKC-85982 | Situated in a particular spot or position. |
| plans | End User - Trip | plan | UKC-96049 | Make plans for something. |
| takes_place_at | Appointment - Health Facility | takes place at | KGE24-4-0012* | The specific location where a medical consultation, check-up, or treatment occurs. |
| has_route | Trip - Route | follow, travel along | UKC-102467 | Travel along a certain course. |
| has_stop | Route - Stop | include | UKC-105576 | Have as a part, be made up out of. |
| has_nearby | Stop - Health Facility | near, close | UKC-76467 | Not far distant in time or space or degree or circumstances. |

| | | | | |
|---|---|---|---|---|
| follows_schedule | Route - Weekly Schedule | follow schedule | KGE24-4-0013* | Adheres to a predefined timetable. |
| overrides | Schedule Exception - Weekly Schedule | override | UKC-1088 | The act of nullifying; making null and void; counteracting or overriding the effect or force of something. |
| causes | Event - Schedule Exception | cause, do, make | UKC-100755 | Give rise to; cause to happen or occur, not always intentionally. |
| delays | Event - Route | delay, detain, hold up | UKC-94853 | Cause to be slowed down or delayed. |
| deviates | Event - Route | deviate | UKC-102843 | Cause to turn away from a previous or expected course. |
| has_start_point | End User - Position | has start point | KGE24-4-0014* | A defined beginning location from which an event, activity, or process is initiated. |
| has_end_point | End User - Position | has end point | KGE24-4-0015* | A defined final location at which an event, activity, or process concludes. |
| has | Stop - Stop Time | have, hold | UKC-103527 | |

Table 4: In this table is presented how "generic" words (first column) regarding ETypes relationships have been formalized and assigned to a UKC concept (third and fourth columns). The last column provides a description of what the concept means. Concept IDs marked with a " * " are new concepts specifically created for this project.

## 4.2 Activity 2: Dataset Filtering

In Activity 2, the focus shifted to aligning the collected data resources with the formalized language concepts identified in Activity 1. The objective was to remove any irrelevant, redundant, or misaligned data points that did not meet the established definitions and ensure that all data entries conformed to the newly defined domain language. Any data elements that were not explicitly defined in the language resources were excluded. This step also involved renaming data columns to better align with their corresponding ETypes, improving both clarity and consistency in how the data was represented. This renaming process helped ensure that the data was more intuitively organized and aligned with the formalized domain language, making it easier to integrate and query.

# 5 Knowledge Definition

The Knowledge Definition phase followed a structured approach articulated in the three phases of the kTelos process. Starting from the information resources previously collected for the KGE project along with the produced language resources and the formalized user purpose, the objective was to produce the final KG Teleontology and the aligned datasets. This process aimed at unifying the representation of the information, improving the interoperability and reusability of the final KG.

## 5.1 Top-Down

This section describes the *Top-Down* knowledge definition phase of the kTelos process. The focus was on reusing Lightweight Ontologies that are already aligned with the **Universal Knowledge Core (UKC)** to establish a high-level view of the key entities involved in the project. By leveraging pre-existing ontologies, the modeling process was accelerated, and semantic interoperability was enhanced. The primary source of ontologies was the **Datascientia LiveKnowledge catalog**.

Specifically, entities from the **GTFS ontology** were utilized due to their relevance to the project's domain. However, only the essential entities required for the teleology were selected, ensuring that the knowledge model remained purpose-specific and streamlined. For instance, while the GTFS ontology includes entities such as *Shape*, *Fare Rule*, and *Payment Method*, these were deemed irrelevant to the project's objectives and were therefore excluded, along with their associated data properties. Instead, key entities such as *Trip*, *Stop*, *Stop Time*, *Route* and *Service rule* were retained, as they directly align with and support the project's goals. Data properties not relevant for the purpose were removed as well (e.g., *TextColor* from the class *Route*).

Additionally, the **Trentino OSM LWOntology** was employed to incorporate entities pertaining to specific healthcare-related points of interest. The selected entities include:

- Thing > osm_place > point_of_interest > health > clinic > *point_clinic*
- Thing > osm_place > point_of_interest > health > dentist > *point_dentist*
- Thing > osm_place > point_of_interest > health > hospital > *point_hospital*
- Thing > osm_place > point_of_interest > health > doctor > *point_doctor*
- Thing > osm_place > point_of_interest > health > pharmacy > *point_pharmacy*

While the OSM ontology also included entities related to transportation, its emphasis on the concept of *"place"* made it less suitable for the project's requirements. Instead, more appropriate and domain-specific alternatives were sourced from the GTFS ontology, resulting in the exclusion of transportation-related entities from the OSM ontology.

Additionally, concepts from **Schema.org** were incorporated to fill any structural gaps left by the GTFS and OSM ontology. Specifically, the chosen entities include:

- Thing > Intangible > StructuredValue > *GeoCoordinates*
- Thing > Person > *Patient*

This selective reuse of entities from multiple ontological sources facilitated the definition of a high-level view of the entities relevant to the project. The specific choices made are detailed in Table 5.

| **Source Ontology** | **Entity** |
|---|---|
| GTFS ontology | Route |
| GTFS ontology | Stop |
| GTFS ontology | Stop Time |
| GTFS ontology | Trip |
| GTFS ontology | Service Rule |
| GTFS ontology | Calendar Date Rule |
| GTFS ontology | Calendar Rule |
| OSM ontology | point_clinic |
| OSM ontology | point_dentist |
| OSM ontology | point_hospital |
| OSM ontology | point_doctor |
| OSM ontology | point_pharmacy |
| Schema.org | GeoCoordinates |
| Schema.org | Patient |

Table 5: Ontology Source and Entities

Each ontology contributes distinctly to its respective subcontext within the project. Specifically, the GTFS ontology is employed for transportation-related entities, while the OSM ontology is utilized for healthcare-related points of interest.
The Protégé result is accessible in the GitHub repository as an OWL RDF/XML file.

## 5.2 Bottom-Up

This section describes the *Bottom-Up* knowledge definition phase of the kTelos process, which was dedicated on building a Teleology tailored to the project's specific requirements. These requirements are captured through the CQs and modeled in the ER diagram established during the Purpose Definition phase of the iTelos methodology. Thus, the primary purpose of this Teleology is to align with the project's competency questions, ensuring that the schema directly supports the intended queries.

In this phase, the ER diagram was formalized in Protégé as a formal schema by modeling ETypes (classes), object properties, and data properties to produce a final OWL file fitting the project purpose and data. First, a parent EType named **"Entity_GID-1"** was created to serve as a common superclass for all the entity types to be modeled. For each EType identified in the ER model, a corresponding subclass of **"Entity_GID-1"** was created, positioned as a sibling to other ETypes, with additional hierarchical relationships added where applicable. Each subclass was named according to the corresponding EType name found in the Language Resource Sheet, and its IRI (Internationalized Resource Identifier) was updated to follow the format http://knowdive.disi.unitn.it/etype#<EType_name>. Each EType was assigned a label and a gloss from the Language Resource Sheet to ensure semantic clarity and

provide descriptive context. Additionally, a boolean custom annotation named *"isEtype"* was added to each EType and set to True, explicitly identifying it as an entity type within the teleology. The **Event** EType, initially envisioned in the ER diagram, was not formalized in the schema due to lack of direct representation within the datasets. Next, data properties were modeled by selecting terms representing entity attributes and defining their corresponding domain entity types (i.e., the EType it belongs to) and data types (e.g., boolean, string, float, etc). To enhance clarity and improve comprehension, a consistent naming convention for data properties was established. Each data property name follows the format *has_[EntityName]_[AttributeName]*, allowing for an intuitive and immediate understanding of the EType to which the attribute belongs. This systematic approach simplifies the identification and interpretation of properties within the knowledge teleology. For instance, the property *"departure time"* of a **Trip** is named *has_trip_departure_time*, clearly indicating that *"departure time"* is an attribute of the **Trip** EType.
Finally, object properties were modeled by identifying terms denoting relationships between entity types and specifying their domain and range.
The Teleology as generated in Protégé is accessible in the GitHub repository as an OWL RDF/XML file.

## 5.3 Middle-Out

This section describes the *Middle-Out* knowledge definition phase of the kTelos process, which focuses on transforming the Teleology—built from project-specific needs in the previous Bottom-Up phase—into a Teleontology. This is achieved by semantically aligning the Teleology with the standardized Lightweight Ontologies that provide grounding to what exist in the real world. Specifically, each concept in the Teleology, one at a time, was added as a child (using a IS-A relation) to its related general concept in the Lightweight Ontology and annotated using the *"SubClass Of"* relationship. The *"Equivalent To"* description was filled when a concept in the Teleology fully matched one in the reference ontology.
The final Teleontology, serving as the operational schema for generating the Knowledge Graph, was designed to facilitate efficient execution of SPARQL queries and address the previously identified competency questions. To simplify query execution while ensuring alignment with project objectives, certain hierarchies were collapsed to their leaf nodes (e.g., *Thing > Intangible > StructuredValue > GeoCoordinates > Location*), as additional subclass distinctions were deemed unnecessary.
Rather than modeling separate entities (**Clinic**, **DentalClinic**, **PhysicalTherapyCenter**, **Hospital**, **Pharmacy**), as suggested by the OSM schema, a unified **HealthCareFacility** entity was used. This approach minimizes redundancy, as all healthcare facilities share the same core attributes (e.g., location, name). The facility type was stored as an attribute. The same rationale was applied to **Patient** and **HealthcareProfessional**, which were unified into the **EndUser** EType and the type stored as an attribute. As before, keeping these entities separate did not provide any additional benefits for the intended use of the KG.
The Teleontology is accessible in the GitHub repository as an OWL RDF/XML file.

Once the Teleontology was produced, the Knowledge Annotation process ensured that every concept in the Teleontology was clearly identified, uniquely referenced, and semantically precise. Each concept was checked against the Language Resource Sheet. If the concept existed, it was updated in Protégé with the appropriate unique Global Identifier (GID) using the format `conceptName_GID-{GID}`. If a concept was not found in the Language Resource sheet, but came from another reference standard (e.g., GTFS, schema.org), the format for the new concept followed the pattern `source:conceptName`. This approach ensures that concepts from external standards are clearly marked with their source and are easy to trace back to their origin. Overall, the Knowledge Annotation process guarantees semantic clarity and removes ambiguity in the representation of entity types, object properties and data properties. For instance, the term "follow" is used to denote two distinct relationships within the knowledge Teleontology. The object property `follow_GID-102467` describes the relationship between a **Trip** and a **Route**, where the trip adheres to the predefined path of the route. On the other hand, `follow_schedule_KGE24-4-0013` represents the link between a **Route** and a **Weekly Schedule**, where the route follows a recurring time-based schedule. By assigning unique GIDs to each instance of "follow", the Teleontology clearly distinguishes these two relationships, supporting precise alignment with the project's domain requirements and ensuring semantic consistency throughout the Knowledge Graph.

### 5.4 Dataset Alignment Activity

At the data layer, the datasets collected, cleaned and formatted during the Information Gathering phase of the iTelos methodology were aligned with the knowledge Teleontology. This alignment involved restructuring and updating the datasets to conform to the ETypes, object properties, and data properties defined in the Teleontology. Specifically, the data types within the datasets (e.g., integers, strings, dates) were adjusted as needed to align with the corresponding properties defined in the Teleontology. As an example, in the original dataset file `calendar.csv`, route availability during the week was represented using separate columns, each corresponding to a weekday and containing a boolean value indicating availability. However, the `has_weekly_schedule_GID-3420` data property for the `WeeklySchedule_KGE24-4-0004` EType is defined as a seven-character string, with each character being either 0 or 1, encoding the same information in a more compact format. To align the `calendar.csv` dataset with this representation, the individual weekday columns were merged into a single column following the specified compact encoding.

## 6 Entity Definition

The Entity Definition phase was the final step in the iTelos methodology, where the focus was on combining the knowledge layer (represented by the Teleontology) with the data layer (represented by the data resources that have been cleaned and aligned in the earlier phases). The result of this phase was the creation of the final Knowledge Graph.

The Entity Definition phase was structured in three different activities:

- Entity Matching
- Entity Identification
- Entity Mapping

### 6.1 Entity Matching

Entity Matching refers to the challenge of identifying and reconciling different representations of the same real-world entity across multiple datasets. In this project, this challenge was effectively addressed during the Information Gathering phase, where a geospatial matching approach was implemented to match entities from the Ministero della Salute datasets with those from the Healthsites dataset. For more details, refer to the Resource Cleaning and Formatting section 3.2. As a result, by the start of the Entity Definition phase, no discrepancies in the representation of real-world entities—inconsistencies at the value level—remained to be addressed.

## 6.2 Entity Identification

With entity matching conflicts resolved, the next step focused on the formal identification of entities within the datasets. Some dataset already had unique identifiers. For those lacking a unique identifier, an Identifying Set—a collection of properties (attributes) uniquely defining an entity type—was used instead. To ensure clarity and make it easy to associate each identifier with its corresponding entity type, a Uniform Resource Identifier (URI) of the type URN (Uniform Resource Name) was defined for each entity type (Table 6). For datasets with pre-existing unique identifiers, the URN followed the format name-uniqueIdentifier (e.g., calendar-urb-000000002202409092025061 2). For datasets relying on an Identifying Set, instead of concatenating property values, the URN was structured as name-number (e.g., pharm-1, pharm-2), where the number was assigned sequentially, starting from 1.

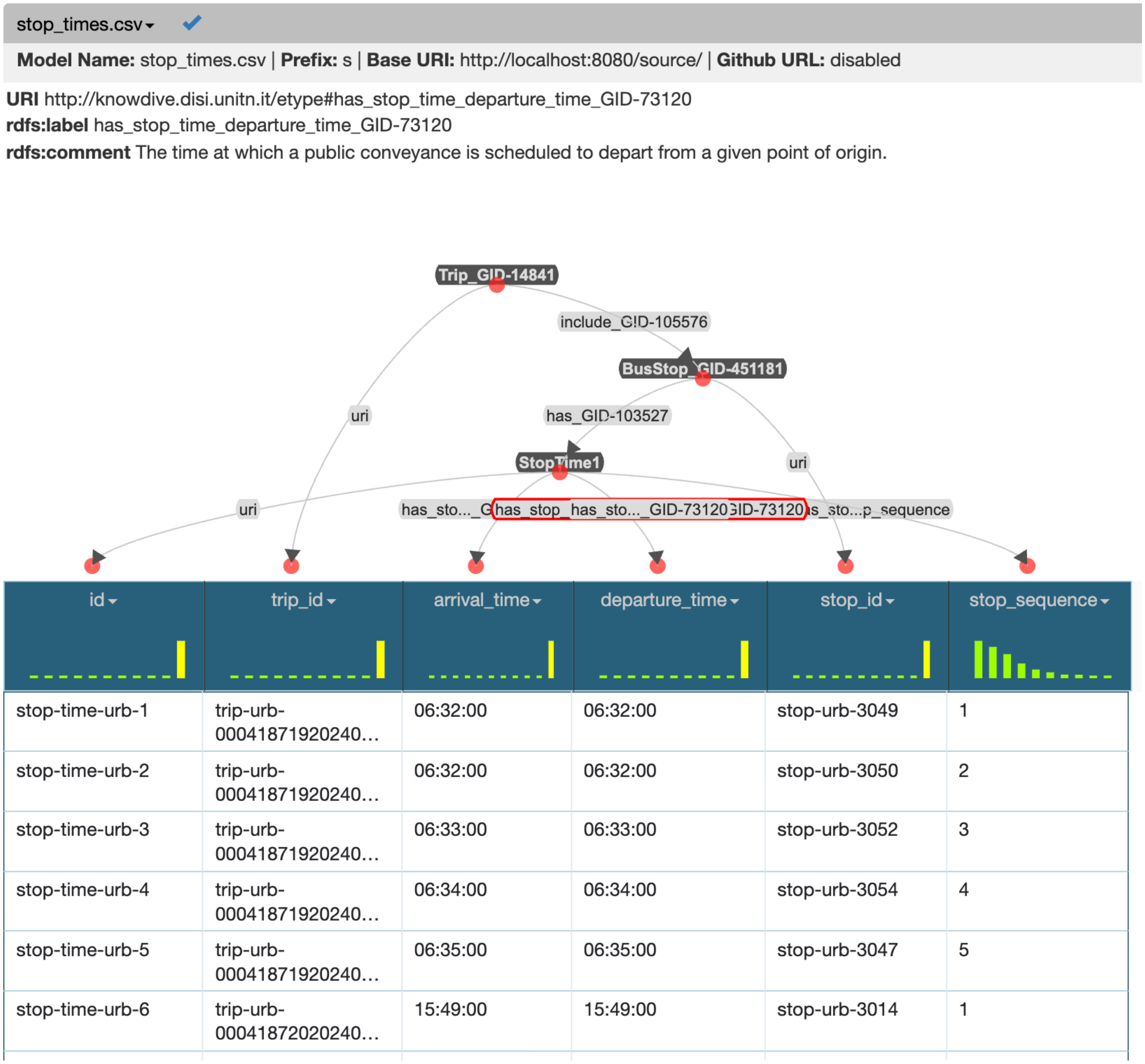
| **dataset** | **real-world object** | **URN** | **attributes** |
|---|---|---|---|
| dati_salute_gov_it_farmacie_trentino.csv | pharmacy | pharm-{n} | legal_name, latitude, longitude |
| osm_health_points_trentino.csv | osm health point | osm-health-point-{n} | legal_name, latitude, longitude |
| routes.csv | route, urban | route-urb-{route_id} | route_id |
| routes.csv | route, extraurban | route-ext-{route_id} | route_id |
| calendar_dates.csv | calendar date, urban | calendar-date-urb-{n} | service_id, date |
| calendar_dates.csv | calendar date, extraurban | calendar-date-ext-{n} | service_id, date |
| calendar.csv | calendar, urban | calendar-urb-{service_id} | service_id |
| calendar.csv | calendar, extraurban | calendar-ext-{service_id} | service_id |
| stops.csv | stop, urban | stop-urb-{stop_code} | stop_code |
| stops.csv | stop, extraurban | stop-ext-{stop_code} | stop_code |
| stop_times.csv | stop time, urban | stop-time-urb-{n} | trip_id, arrival_time, stop_id |
| stop_times.csv | stop time, extraurban | stop-time-ext-{n} | trip_id, arrival_time, stop_id |
| trips.csv | trip, urban | trip-urb-{trip_id} | trip_id |
| trips.csv | trip, extraurban | trip-ext-{trip_id} | trip_id |

Table 6: Uniform Resource Name (URN) structure assigned to each entity type across different datasets. For datasets with unique identifiers, the URN incorporates the existing identifier, while for datasets relying on an Identifying Set, a sequential number {n} is used. The attributes column specifies the properties used to define the Identifying Set when a unique identifier is not available.

## 6.3 Entity Mapping

The final step in the entity definition process involved mapping the conceptual representations of entities, as defined in the Teleontology, to their corresponding data values within the datasets. To achieve this, the Karma Data Integration Tool was employed. Datasets were imported alongside the final language-aligned schema, URI-s were defined for each dataset and a semantic type was assigned to each column. When object properties established relationships between entities, the URI for the related class was identified, and an outgoing link was created from the domain entity to the range entity, effectively representing the relationship.
As an example, in the mapping shown in Fig. 4, the **StopTime** EType has its own URI and its data properties directly map to the respective columns in the dataset (**arrival_time**, **departure_time**, **stop_sequence**). The **stop_id** column contains URIs referring to instances from *stops.csv*, so its semantic type is set accordingly. As a result, an outgoing link is established from the **BusStop** class to the **StopTime** class, with the appropriate object property added to represent their relationship. Beyond **StopTime** and **BusStop**, the mapping also includes the

Trip **EType**. The **Trip** entity is assigned a URI and linked to the **BusStop** entity through the **include** object property, indicating that a trip includes a specific bus stop.




| id | trip_id | arrival_time | departure_time | stop_id | stop_sequence |
|---|---|---|---|---|---|
| stop-time-urb-1 | trip-urb-00041871920240... | 06:32:00 | 06:32:00 | stop-urb-3049 | 1 |
| stop-time-urb-2 | trip-urb-00041871920240... | 06:32:00 | 06:32:00 | stop-urb-3050 | 2 |
| stop-time-urb-3 | trip-urb-00041871920240... | 06:33:00 | 06:33:00 | stop-urb-3052 | 3 |
| stop-time-urb-4 | trip-urb-00041871920240... | 06:34:00 | 06:34:00 | stop-urb-3054 | 4 |
| stop-time-urb-5 | trip-urb-00041871920240... | 06:35:00 | 06:35:00 | stop-urb-3047 | 5 |
| stop-time-urb-6 | trip-urb-00041872020240... | 15:49:00 | 15:49:00 | stop-urb-3014 | 1 |

Figure 4: Example of entity mapping in Karma.

The Entity Mapping activity generated a RDF-Turtle (ttl) file for each dataset, documenting all mapping operations performed. These files ensure transparency, reusability, and reproducibility of the mapping process. Moreover, future updates or changes to the datasets or Teleontology can be easily handled by revisiting the mapping model. All RDF-Turtle files and mapping model files are accessible in the GitHub repository.

# 7 Evaluation

The iTelos methodology provides an evaluation framework to ensure that the KG is both fit for its intended purpose (primary objective) and generalizable for broader reuse (secondary objective). The evaluation criteria, described below, consider both the Knowledge Layer and the Data Layer.

## 7.1 Knowledge Layer Evaluation

iTelos defines a set of metrics to evaluate the KG, with *Coverage* being one of the most valuable. Coverage quantifies how well a portion of knowledge is represented within the KG. It helps assess both the primary and secondary evaluation objectives, by measuring the extent to which the Teleontology supports the entities and properties derived from Competency Questions (CQs) and Reference Ontologies, respectively.
Table 7 summarizes the total count of ETypes, object properties, and data properties used to calculate coverage.

| **Entity/Property Type** | **Instances Count** |
|---|---|
| ETypes | 21 |
| Object Properties | 16 |
| Data Properties | 39 |

Table 7: Final Teleontology Evaluation Summary.

Table 8 summarizes the number of entities for each EType.

| **EType** | **Number of Entities** |
|---|---|
| StopTime | 165227 |
| Trip_GID-1484 | 6914 |
| ScheduleException_KGE24-4-0005 | 4973 |
| TransportOption_KGE24-4-0017 | 4868 |
| BusStop_GID-45118 | 3966 |
| HealthFacility_GID-20121 | 177 |
| Position_GID-27518 | 173 |
| BusRoute_GID-45117 | 167 |
| WeeklySchedule_KGE24-4-0004 | 146 |
| schema:OpeningHours | 108 |
| TrainStop_KGE24-4-0003 | 61 |
| RouteTransfer_KGE24-4-0021 | 34 |
| EndUser_GID-49456 | 9 |
| StartPosition_KGE24-4-0014 | 7 |
| EndPosition_KGE24-4-0015 | 7 |
| HealthAppointment | 6 |
| TrainRoute_KGE24-4-0002 | 2 |

Table 8: Number of entities for each EType in the KG.

### 7.1.1 Teleontology vs CQs

Entity/Property coverage in the context of the Teleontology and its alignment with a set of CQs is calculated by dividing the number of entities/properties that appear in both the Teleontology and the CQs by the total number of entities/properties extracted from the CQs.

Referring to Figure 1, the ER model consisted of a total of 10 ETypes. Overall, the final HealthRoute Trentino Teleontology defines more ETypes, data properties, and object properties compared to the CQs. This reflects the evolving nature of the project, as initial design choices and knowledge requirements were refined over time.
Below is a table with the final evaluation, considering the ETypes, object properties, and data properties coverage.

| | **ETypes** | **$Cov_E$** | **Object Properties** | **$Cov_{OP}$** | **Data Properties** | **$Cov_{DP}$** |
|---|---|---|---|---|---|---|
| **Total identified from CQs** | 10 | | 9 | | 25 | |
| **Total defined for the project** | 21 | 90% | 16 | 89% | 39 | 92% |

Table 9: Teleontology vs Competency Questions Coverage

Coverage is not perfect because the **Event** EType and its related properties, initially envisioned in the CQs and the ER diagram, were not formalized in the final Teleontology schema due to lack of direct representation within the datasets.

#### 7.1.2 Teleontology vs Reference Ontologies

Referring to the ontologies listed in Table 5 and the final values reported in the Teleontology Evaluation Summary (Table 7), Table 10 presents the final coverage evaluation of the Teleontology with respect to the set of Reference Ontologies.

| | **ETypes** | $Cov_E$ | **Object Properties** | $Cov_{OP}$ | **Data Properties** | $Cov_{DP}$ |
|---|---|---|---|---|---|---|
| **GTFS ontology** | | | | | | |
| Total in the ontology | 30 | | 19 | | 22 | |
| Total reused in the project | 7 | 23% | 0 | 0% | 13 | 59% |
| **OSM ontology** | | | | | | |
| Total in the ontology | 791 | | 0 | | 0 | |
| Total reused in the project | 0 | 0% | 0 | 0% | 0 | 0% |
| **Schema.org** | | | | | | |
| Total in the ontology | – | | – | | – | |
| Total reused in the project | – | – | – | – | 1 | – |

Table 10: Teleontology vs Reference Ontologies Coverage

Regarding Schema.org, the missing values in the table are due to the ontology's large size: it contains too many ETypes and properties. Therefore, only the select types and data properties used in the project are included in the table.

### 7.2 Data Layer Evaluation

Evaluation at the Data Layer assesses the connectivity of the KG, meaning how well the data within the graph is structured and interlinked to enable effective query execution. Specifically, the connectivity of a KG can be evaluated over two dimensions:

1. **Entity Connectivity (EC)** which measures how well entities are connected to one another through object properties.
2. **Property Connectivity (PC)** which measures how effectively each entity is connected to its attributes, ensuring that entities are properly described with relevant property values.

Both dimensions are computed using a *connectivity matrix*, available in the GitHub repository.

### 7.2.1 Entity Connectivity

Formally, Entity Connectivity for an entity type $X$, denoted as $EC(X)$, is defined as:

$$EC(X) = \frac{\sum_{Y \neq X}(X, Y)}{OP(X)} \tag{1}$$

where $(X, Y)$ represents the value of the off-diagonal cell of the connectivity matrix, and $OP(X)$ denotes the number of object properties defined for entity type $X$.

**7.2.1.1 Example.** Consider the entity type **HealthFacility_GID-20121**. From the connectivity matrix, entities of this type are connected to entities of type **Position_GID-27518** and **OpeningHours** through object properties. The off-diagonal matrix cell values are:

$$\text{HealthFacility} \rightarrow \text{Position} = 1, \quad \text{HealthFacility} \rightarrow \text{OpeningHours} = 0.11$$

The sum of the off-diagonal values for this row is therefore:

$$\sum_{Y \neq X}(X, Y) = 1 + 0.11 = 1.11$$

Given that the entity type **HealthFacility_GID-20121** is associated with two object properties, i.e., $OP(\text{HealthFacility}) = 2$, the Entity Connectivity is computed as:

$$EC(\text{HealthFacility}) = \frac{1.11}{2} = 0.555$$

This value represents the Entity Connectivity of the **HealthFacility_GID-20121** entity type.

### 7.2.2 Property Connectivity

Formally, Property Connectivity for an entity type $X$, denoted as $PC(X)$, is defined as:

$$PC(X) = \frac{(X, X)}{DP(X)} \tag{2}$$

where $(X, X)$ represents the value of the diagonal cell of the connectivity matrix (i.e., the completeness of data properties for that entity type), and $DP(X)$ denotes the number of data properties defined for entity type $X$.

**7.2.2.1 Example.** Consider the entity type **HealthFacility_GID-20121**. From the connectivity matrix, the diagonal value for this entity type is:

$$(X, X) = 1 \mid 1$$

which indicates that there are two data properties, both non-null. Then, the Property Connectivity is computed as:

$$PC(\text{HealthFacility}) = \frac{1+1}{2} = 1$$

This value indicates that all data properties for **HealthFacility_GID-20121** are present and non-null, i.e., entities of this type are fully described with their attributes.

The table below summarizes the Entity Connectivity (EC) and Property Connectivity (PC) for each entity type in the KG. The `TOTAL (KG)` row sums all $EC(X)$ and $PC(X)$ contributions, providing an overall measure of connectivity across the KG.

| EType | $\sum(X,Y)$ | OP(X) | EC(X) | (X,X) | DP(X) | PC(X) |
|---|---|---|---|---|---|---|
| EndUser_GID-49456 | 2.67 | 3 | 0.89 | 1 | 2 | 1 |
| StartPosition_KGE24-4-0014 | 1.495 | 3 | 0.498 | 1 | 6 | 1 |
| EndPosition_KGE24-4-0015 | 1.565 | 3 | 0.52 | 1 | 6 | 1 |
| HealthFacility_GID-20121 | 1.11 | 2 | 0.56 | 1 | 2 | 1 |
| HealthAppointment | 1 | 1 | 1 | 1 | 3 | 1 |
| Position_GID-27518 | 0.87 | 1 | 0.87 | 1 | 6 | 1 |
| BusRoute_GID-45117 | 0.73 | 1 | 0.73 | 1 | 3 | 1 |
| TrainRoute_KGE24-4-0002 | 0.5 | 1 | 0.5 | 1 | 3 | 1 |
| Trip_GID-1484 | 4 | 6 | 0.67 | 1 | 2 | 1 |
| BusStop_GID-45118 | 1 | 1 | 1 | 1 | 3 | 1 |
| TrainStop_KGE24-4-0003 | 1 | 1 | 1 | 1 | 3 | 1 |
| StopTime | – | – | – | 1 | 3 | 1 |
| WeeklySchedule_KGE24-4-0004 | – | – | – | 1 | 3 | 1 |
| ScheduleException_KGE24-4-0005 | 1 | 1 | 1 | 1 | 2 | 1 |
| TransportOption_KGE24-4-0017 | 1 | 2 | 0.5 | 1 | 1 | 1 |
| RouteTransfer_KGE24-4-0021 | – | – | – | 1 | 5 | 1 |
| OpeningHours | – | – | – | 1 | 3 | 1 |
| TOTAL (KG) | | | 11.71 | | | 17 |
| AVERAGE (KG) | | | 0.90 | | | 1 |

## 7.3 Query execution:

To conclude the evaluation phase, the competency questions were translated into SPARQL queries in order to assess the ability of the Knowledge Graph to effectively satisfy the project's intended objectives.
All the SPARQL queries used are available in the GitHub repository.

#### 7.3.1 CQ-1 (P1-S5)

*Stefania has a busy Monday morning with back-to-back appointments at two nursing homes, but a public transport strike has been announced.*
*a) Which train/bus lines does she rely on for her appointments?*
*b) Are they affected by the strike?*

CQ-1 (P1-S5) cannot be answered based on the current KG. The KG does not contain information about public transport strikes or strike-related service disruptions, because no related data are available; therefore, such events are not modeled in the KG schema (see Section 5.2).

#### 7.3.2 CQ-2 (P1–S1)

*Stefania has a morning visit scheduled at a nursing home in Trento at 11:00 a.m.*
*a) What is the most efficient public transportation route for her to arrive at the nursing home on time?*
*b) Is there a pharmacy within 500 meters of the final bus stop?*
*c) If she leaves home at 9:30 a.m., will she have sufficient time to stop by the pharmacy to buy her thyroid medications before proceeding to her appointment?*

The proposed solution exploits the fact that the HealthRoute Trentino KG supports multi-leg journeys through route transfers. In the KG, transfers are modeled as entities (`etype:has_route_transfer_KGE24-4-0022`) that connect two different routes by specifying: (i) the arrival stop on the first route ($B$), (ii) the departure stop on the second route ($C$), and (iii) the walking distance required to move between these two stops. This representation enables the modeling of realistic public transport scenarios in which reaching the destination requires changing lines.

##### 7.3.2.1 Pseudo-workflow for CQ-2 (P1-S1) – a)

*What is the most efficient public transportation route for her to arrive at the nursing home on time?*

1. **Initialize temporal context** Set the reference date and time for the query (Monday morning, 9:00 a.m.) and derive the corresponding weekday to ensure that only trips operating on that day are considered.

2. **Identify the optimal access stop ($A$)** Starting from Stefania's home address (Via Temanza), check all nearby public transport stops and select the stop with the minimum walking distance that is served by a valid trip.

3. **Select a feasible first trip (Trip 1) departing from stop ($A$)** From the chosen access stop ($A$), retrieve trips that:

   - operate on the selected weekday and within the valid schedule period,
   - depart after Stefania's starting time,
   - o allow sufficient time to walk from the start location to stop $A$.

   For each valid trip, identify a downstream arrival stop ($B$).

4. **Retrieve allowed transfers ($B \to C$)** For each arrival stop $B$, retrieve all route transfers associated with the route of Trip 1. For each candidate transfer, the walking distance between $B$ and $C$ is compared with the available time window derived from the arrival time at $B$ and the departure time from $C$. Transfers that cannot be completed within this time window are discarded. From stop $C$, retrieve second-leg trips on the second route.

   At this stage, all feasibility and efficiency constraints are applied only to the first leg of the journey (Trip 1) and to the transfer. Attempting to optimize the second leg (Trip 2) within the same query resulted in very poor performance, as the combination of schedule validation and aggregation for minimizing the walking distance to the destination for Trip 2 produced an excessively large join space. For this reason, the optimization process was deliberately split: first-leg feasibility and transfer validation are handled separately from second-leg optimization.

   Accordingly, a SPARQL `INSERT` query is used to store in a dedicated subgraph all feasible combinations of:

   - access stop $A$,
   - first-leg trip (Trip 1) and arrival stop $B$,
   - route transfer ($B \to C$),
   - associated second route (Trip 2 candidates, not yet optimized).

   This intermediate representation enables a subsequent query to efficiently optimize the second leg and select the best overall route.

5. **Select a feasible second trip (Trip 2) departing from $C$** Query the subgraph to retrieve trips on the transferred route that:

   - operate on the same day,
   - depart from stop $C$ after arrival at $C$.

   Identify all downstream stops ($D$) served by these trips.

6. **Optimize destination stop ($D$)** For each candidate destination stop ($D$), check the walking distance to the nursing home address (Via Piave) and retain only the stop that minimizes walking distance to the nursing home.

7. **Rank** Rank the journeys by earliest arrival time at stop $D$ and return only the top-ranked result.

The query result (Fig. 5) identifies a feasible multi-leg journey on **Monday, 4 November 2024**, starting from Stefania's home in **Via Temanza (Borgo Valsugana)** at **09:00** and reaching the nursing home in **Via Piave (Trento)** shortly before **11:00**. The journey begins with a regional train trip on the **Valsugana line** (Trip 1), departing from stop `stop-train-1879` at **09:25** and arriving at `stop stop-train-2264` at **10:29**. A short walking transfer of **0.09 km** (approximately **1 minute**) connects the train arrival stop to the departure stop of an urban bus. The second leg (Trip 2) is served by urban bus **4**, departing at **10:45** and arriving at stop `stop-urb-404` at **10:49**. From the final stop, Stefania completes the journey with a **7-minute walk** (0.58 km) to the destination. This journey is selected as the optimal solution because it satisfies all temporal, walking, and transfer constraints and yields the earliest arrival at the destination.

Filter query results Compact view Hide row numbers Showing results from 0 to 1 of 1. Query took 13s, moments ago.

| | datetime | weekday_na... | start_point | end_point | trip1 | route1 | A_stop | B_stop | A_stop_dep... | B_stop_arriv... | route_transfer | Bstop_to_Cs... | trip2 | route2 | route_name2 | C_stop | D_stop | C_stop_dep... | D_stop_arriv... | Dstop_to_en... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | "2024-11-04T09:00:00"^^xsd:dateTime | "Monday" | src:ViaTemanza-11-38122-BorgoValsugana | src:ViaPiave-78-38122-Trento | src:trip-train-16110 | src:route-train-valsugana | src:stop-train-1879 | src:stop-train-2264 | "09:25:00" | "10:29:00" | src:transfer-9 | "(0.09 km, 1 min)" | src:trip-urb-0004201602024090920250612 | src:route-urb-539 | "4" | src:stop-urb-247 | src:stop-urb-404 | "10:45:00" | "10:49:00" | "(0.58 km, 7 min)" |

Figure 5: Query results for CQ-2(a).
Query code: CQ-2(P1-S1)-a.sparql, Requires P1-S5_Triples.brf to run

#### 7.3.2.2 Pseudo-workflow for CQ-2 (P1-S1) – b)

*Is there a pharmacy within 500 meters of the final bus stop?*

1. **Identify the reference transport stop**
   Use the final bus stop obtained from CQ-2(a) (`stop-urb-404`) as the spatial reference point for the proximity analysis.
2. **Retrieve facilities connected to the stop via walking access**
   Retrieve all health facilities that are associated with the reference stop through a transport option.
3. **Filter by facility type (pharmacy)**
   Restrict the retrieved health facilities to those whose facility type is pharmacy.
4. **Apply spatial proximity constraint**
   Retain only those pharmacies whose walking distance from the final bus stop is less than 500 meters.
5. **Rank by walking accessibility**
   Order the pharmacies by increasing walking time.

The query results (Fig. 6) confirm that there are pharmacies located within 500 meters of the final bus stop (`stop-urb-404`). In particular, two pharmacies satisfy the proximity constraint. **Farmacia Dall'Armi**, located at **Piazza del Duomo 10 in Trento**, is reachable with a walking distance of **0.45 km**, corresponding to an estimated walking time of **5 minutes**. **Farmacia Santoni**, located at **Piazza Mario Pasi 20 in Trento**, is also within the required range, at a distance of **0.48 km** and a walking time of **6 minutes**.

Filter query results Compact view Hide row numbers Showing results from 0 to 2 of 2. Query took 0.1s, moments ago.

| | stop | health_facility | legal_name | position | access_distance |
|---|---|---|---|---|---|
| 1 | src:stop-urb-404 | src:pharm-18 | "Farmacia Dall'Armi" | src:PiazzaDelDuomo-10-38122-Trento | "(0.45 km, 5 min)" |
| 2 | src:stop-urb-404 | src:pharm-20 | "Farmacia Santoni" | src:PiazzaMarioPasi-20-38122-Trento | "(0.48 km, 6 min)" |

Figure 6: Query results for CQ-2(b).
Query code: CQ-2(P1-S1)-b.sparql

#### 7.3.2.3 CQ-2 (P1-S1) – c)

*If she leaves home at 9:30 a.m., will she have sufficient time to stop by the pharmacy to buy her thyroid medications before proceeding to her appointment?*

To answer CQ-2(c), the journey identified in part (a) and the pharmacy proximity analysis from part (b) are jointly considered under the updated departure condition.

When Stefania's departure time from home is set to **9:30 a.m.**, the optimal journey identified by the query changes and results in a later arrival at the destination. In this scenario, she boards a Valsugana train trip (`trip-train-16916`) departing from *Borgo Valsugana Est* at **09:55** and arriving in *Trento* at **10:59**. After a short transfer of approximately **1 minute**, she continues with an urban bus on route 400, departing from *Piazza Dante – Stazione FS* at **11:09** and arriving at *Piazza di Fiera* at **11:14**. From this final stop, the nursing home is reached with an additional **5-minute walk** (0.38 km), leading to an estimated arrival time of approximately **11:19 a.m.**. Since the appointment at the nursing home is scheduled for 11:00 a.m., this itinerary already results in a late arrival. Consequently, there is **no available time margin** to stop at a pharmacy before the appointment.

Filter query results  Compact view ☐ Hide row numbers ☐  Showing results from 0 to 1 of 1. Query took 13s, moments ago.

| | datetime | weekday_na… | start_point | end_point | trip1 | route1 | A_stop | A_stop_name | B_stop | B_stop_name | A_stop_dep… | B_stop_arriv… | route_transf… | Bstop_to_C… | trip2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | "2024-11-04T09:30:00" ^^xsd:dateTime | "Monday" | src: ViaTemanza-11-38122-BorgoValsugana | src:ViaPiave-78-38122-Trento | src:trip-train-16916 | src:route-train-valsugana | src:stop-train-1879 | "borgo valsugana est" | src:stop-train-2264 | "trento" | "09:55:00" | "10:59:00" | src:transfer-8 | "(0.09 km, 1 min)" | src:trip-url 00041885 2409092 612 |

Figure 7: SPARQL query for CQ-2(c).
Query code: CQ-2(P1-S1)-c.sparql

Filter query results  Compact view ☐ Hide row numbers ☐  Showing results from 0 to 1 of 1. Query took 13s, moments ago.

| stop_name | A_stop_dep… | B_stop_arriv… | route_transf… | Bstop_to_C… | trip2 | route2 | route_name2 | C_stop | C_stop_name | D_stop | D_stop_name | C_stop_dep… | D_stop_arriv… | Dstop_to_en… |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 " | "09:55:00" | "10:59:00" | src:transfer-8 | "(0.09 km, 1 min)" | src:trip-urb-00041885820240909202506 12 | src:route-urb-400 | "5" | src:stop-urb-247 | "Piazza Dante "Stazione Fs" | src:stop-urb-165 | "Piazza di Fiera" | "11:09:00" | "11:14:00" | "(0.38 km, 5 min)" |

Figure 8: Query results for CQ-2(c).
Query code: CQ-2(P1-S1)-c.sparql, Requires P1-S5_Triples.brf to run

#### 7.3.3 CQ-3 (P2–S1)

*I'ts Wednesday, Riccardo has a cardiovascular check-up scheduled for 9:00 a.m. at Ospedale Santa Chiara and plans to pick up some antihistamines beforehand.*
*a) How many pharmacies are located within a 5-10 minute walk from each bus stop on his route to the hospital, and which one would be the most convenient to stop at? b) If he leaves home at 8:30 a.m., will he arrive on time for his appointment?*

On **Wednesday, 6 November 2024**, starting at **08:30**, the most suitable public transport option for Riccardo to reach **Ospedale Santa Chiara** is the one illustrated in Fig. 9. From **Via Giuseppe Grazioli**, Riccardo reaches the boarding stop **S. Francesco – Porta Nuova** with a walking distance of approximately **220 meters (3 minutes)**. The selected bus route (**line 1**) departs from this stop at **08:39**. After travelling **four stops**, Riccardo gets off at **Bolghera "S. Antonio"** at **08:44**. From there, the hospital can be reached with an additional walk of around **270 meters (3 minutes)**. Overall, this journey ensures arrival at the hospital well in advance of the scheduled 9:00 a.m. cardiovascular check-up, leaving sufficient time for a brief stop at a nearby pharmacy along the route.

**Pseudo-workflow for CQ-3 (P2-S1)**

Filter query results | Compact view | Hide row numbers | Showing results from 0 to 1 of 1. Query took 1.8s, minutes ago.

| | datetime | weekday_... | start_point | end_point | trip | route | bus_name | time_until... | start_poi... | A_stop | A_stop_n... | B_stop | B_stop_n... | A_stop_d... | B_stop_ar... | stop_count | Bstop_to_... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | "2024-11-06T08:30:00"^^xsd:dateTime | "Wednesday" | src:ViaGiuseppeGrazioli-3-38122-Trento | src:LargoMedaglieDOro-9-38122-Trento | src:trip-urb-00042007920240909 2025061 2 | src:route-urb-536 | "1" | "0:09" | "(0.22 km, 3 min)" | src:stop-urb-410 | "S.Francesco Porta Nuova" | src:stop-urb-284 | "Bolghera "S.Antonio" | "08:39:00" | "08:44:00" | "4"^^xsd:integer | "(0.27 km, 3 min)" |

Figure 9: Optimal public transport journey for CQ-3 (P2-S1).
Query code: CQ-3(P2-S1)-optimal-public-transport-journey.sparql, Requires P2Trips.brf to run

1. **Fix the reference journey**
   The previously computed optimal public transport journey to Ospedale Santa Chiara is selected by fixing the corresponding trip identifier. The boarding stop (**S. Francesco – Porta Nuova**) and the alighting stop (**Bolghera "S. Antonio"**) are used as reference points. Their stop sequence numbers along the trip are retrieved in order to delimit the portion of the route relevant to Riccardo's journey.

2. **Extract intermediate bus stops along the route**
   All bus stops belonging to the selected trip are retrieved. Only those stops whose sequence number falls between the start and end stops (inclusive) are retained. This step defines the complete set of bus stops encountered during the journey.

3. **Associate nearby health facilities to each stop**
   For each bus stop on the route, health facilities that can be accessed on foot from that stop are identified.

4. **Filter pharmacies within walking distance constraints**
   From the set of nearby health facilities, only those classified as pharmacies are retained. A further filter is applied to keep only pharmacies reachable within a 5-minute walking distance.

5. **Count pharmacies per bus stop**
   For each bus stop on the route, the number of accessible pharmacies satisfying the walking-time constraint is computed. This produces a per-stop count of pharmacies available (Fig. 10).

6. **Select the most convenient pharmacy**
   Convenience is defined as minimum walking time. Therefore, pharmacies are ordered by increasing walking-time distance from their associated bus stop, and the pharmacy with the smallest access time is selected as the most convenient option for Riccardo (Fig. 11).

Filter query results | Compact view | Hide row numbers | Showing results from 0 to 5 of 5. Query took 0.1s, moments ago.

| | stop_sequence | stop | num_pharmacies |
|---|---|---|---|
| 1 | "15" | src:stop-urb-410 | "3"^^xsd:integer |
| 2 | "16" | src:stop-urb-277 | "1"^^xsd:integer |
| 3 | "17" | src:stop-urb-167 | "2"^^xsd:integer |
| 4 | "18" | src:stop-urb-352 | "3"^^xsd:integer |
| 5 | "19" | src:stop-urb-284 | "3"^^xsd:integer |

Figure 10: Number of accessible pharmacies per bus stop along the route.
Query code: CQ-3(P2-S1)-a.sparql

Filter query results Compact view Hide row numbers Showing results from 0 to 1 of 1. Query took 0.1s, moments ago.

| | trip | route | stop | stop_sequence | health_facility | legal_name | position | access_distance |
|---|---|---|---|---|---|---|---|---|
| 1 | src:trip-urb-0004200792024090920250612 | src:route-urb-536 | src:stop-urb-284 | "19" | src:pharm-27 | "Farmacia Bolghera" | src:LargoMedaglieDOro-8-38122-Trento | "(0.12 km, 1 min)" |

Figure 11: Selection of the most convenient pharmacy based on walking time.
Query code: CQ-3(P2-S1)-b.sparql

#### 7.3.4 CQ-4 (P2–S3)

*After enjoying a walk in Parco delle Albere on a sunny Sunday afternoon, Riccardo begins experiencing allergy symptoms and needs to buy antihistamines.*
*a) Is there a nearby pharmacy open on Sundays?*
*b) Given the reduced weekend public transport service, what is the fastest route he can take to reach the pharmacy and then continue home?*

##### 7.3.4.1 Pseudo-workflow for CQ-4 (P2-S3) – a)

*Is there a nearby pharmacy open on Sundays?*

1. **Define the scenario parameters**
   Fix the user to Riccardo, the current location to the area near Parco delle Albere, and the time context to Sunday at 16:30.
2. **Retrieve nearby health facilities**
   Identify nearby health facilities and restrict the set of health facilities to those classified as pharmacies.
3. **Apply opening-hour constraints**
   For each pharmacy, verify that it has an opening interval on Sunday and that the interval includes the query time (16:30).
4. **Rank pharmacies by convenience**
   Order pharmacies by increasing total walking time to identify the most accessible option.

For **Sunday** at **16:30**, the query returns no results, indicating that no nearby pharmacy can be confirmed as open at that time based on the available KG data. This outcome does not necessarily reflect real-world availability, as opening-hour information is provided only for a minority of pharmacies in the dataset.
To validate the correctness of the query logic, the same query was executed by changing the day of the week from **Sunday** to **Saturday**, while keeping all other parameters unchanged. In this case, the query successfully returned one nearby pharmacy that is open at the specified time (Fig. 12).

##### 7.3.4.2 CQ-4 (P2-S3) – b)

*Given the reduced weekend public transport service, what is the fastest route he can take to reach the pharmacy and then continue home?*

Filter query results Compact view Hide row numbers Showing results from 0 to 1 of 1. Query took 0.7s, moments ago.

| | start_point | health_facility | legal_name | position | access_minutes | day | interval_start | interval_end |
|---|---|---|---|---|---|---|---|---|
| 1 | src: ViaRobertodaSanseverino-95-38122-Trento | src:pharm-90 | "Farmacia Alla Madonna" | src:ViaGiannantonioManci-42-38122-Trento | "15 min" | "Saturday" | "15:00" | "19:00" |

Figure 12: Query results for CQ-4(a) when executed on Saturday.
Query code: CQ-4(P2–S3)-a.sparql

The logic adopted to compute the fastest route is the same as the one previously used for CQ-3 (P2-S1). In particular, the approach minimizes walking access and egress times, verifies schedule validity with respect to the selected day, and selects the earliest feasible departure after the query time.

For this scenario, the journey is intentionally decomposed into two consecutive segments: from Riccardo's current position to the pharmacy, and from the pharmacy back to his home. Although **Farmacia Alla Madonna** was previously identified as closed on Sundays according to the available opening-hour data, it is nonetheless used here as an intermediate destination. This choice is deliberate and serves the purpose of evaluating public transport accessibility, independently of pharmacy opening-hour completeness, which has already been discussed as a data limitation.

To further stress-test the availability of public transport under constrained conditions, the date **8 December 2024** is selected. This date is not only a **Sunday**, but also a **public holiday**, and therefore represents a worst-case scenario in terms of reduced service frequency. By doing so, the analysis focuses on identifying the fastest feasible routes under minimal transport availability.

On **Sunday, 8 December 2024**, starting at **16:30**, the fastest public transport option for Riccardo to reach **Farmacia Alla Madonna** from the area near Parco delle Albere is provided by **bus line 19**. Riccardo first walks approximately **190 meters (2 minutes)** from **Via Roberto da Sanseverino** to the boarding stop **Lavoro e Scienza "BUC"**. The bus departs from this stop at **16:36**, and after travelling **seven stops**, Riccardo alights at **Piazza Dante "Palazzo della Regione"** at **16:48**. From there, the pharmacy can be reached with an additional walk of about **250 meters (3 minutes)** (Fig. 13).

Filter query results Compact view Hide row numbers Showing results from 0 to 1 of 1. Query took 2m 16s, moments ago.

| | DateTime | WeekDay | A | C | AC_Trip | AC_Route | BusNumber | A_stop | A_stop_na... | A_to_Asto... | C_stop | C_stop_na... | C_to_Csto... | Departure... | ArrivalTime | StopCount |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | "2024-12-08T16:30:00" ^^xsd:dateTime | "Sunday" | src: ViaRobertodaSanseverino-95-38122-Trento | "Farmacia Alla Madonna" | src:trip-urb-0004218972024090920250612 | src:route-urb-623 | "19" | src:stop-urb-3100 | "Lavoro E Scienza "Buc" | "(0.19 km, 2 min)" | src:stop-urb-202 | "Piazza Dante "Pal. Regione" | "(0.25 km, 3 min)" | "16:36:00" | "16:48:00" | "7"^^xsd:integer |

Figure 13: Fastest public transport route to Farmacia Alla Madonna on Sunday, 8 December 2024.
Query code: CQ-4(P2–S3)-b01.sparql

After leaving **Farmacia Alla Madonna**, Riccardo continues his journey home, starting from **17:00**. The fastest available public transport option under weekend and public-holiday conditions is provided by **bus line A**. Riccardo reaches the boarding stop **Torre Verde – Galasso** with a short walk of approximately **100 meters (1 minute)**. The bus departs at **17:33** and arrives at **S. Francesco – Porta Nuova** at **17:37**, after travelling **two stops**. From there, Riccardo can reach **Via Giuseppe Grazioli** with an additional walk of around **220 meters (3 minutes)** (Fig. 14).

Filter query results Compact view Hide row numbers Showing results from 0 to 1 of 1. Query took 4.2s, moments ago.

| | DateTime | WeekDay | C | B | CB_Trip | CB_Route | BusNumber | C_stop | C_stop_na... | C_to_Csto... | B_stop | B_stop_na... | B_to_Bsto... | Departure... | ArrivalTime | StopCount |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | "2024-12-08T17:00:00"^^xsd:dateTime | "Sunday" | "Farmacia Alla Madonna" | src:ViaGiuseppeGrazioli-3-38122-Trento | src:trip-urb-0004205472024090920250612 | src:route-urb-614 | "A" | src:stop-urb-443 | "Torre Verde Galasso" | "(0.10 km, 1 min)" | src:stop-urb-410 | "S.Francesco Porta Nuova" | "(0.22 km, 3 min)" | "17:33:00" | "17:37:00" | "2"^^xsd:integer |

Figure 14: Fastest public transport route from Farmacia Alla Madonna to home.
Query code: CQ-4(P2–S3)-b02.sparql

#### 7.3.5 CQ-5 (P3–S6)

*Elena has a morning appointment with her pulmonologist at Ospedale Santa Chiara in Trento. Due to ongoing construction on the train line between Pergine Valsugana and Trento, she must rely on replacement bus services. a) What is the best route for her to take? b) Which is the travel time using the replacement bus?*

Elena's usual journey from **Pergine Valsugana** to **Trento** for her morning appointment at **Ospedale Santa Chiara** would rely on the **Valsugana regional railway line**. There are **three feasible train connections** on **Monday, 4 November 2024**, boarding at **Levico** and terminating at **Trento** (Fig. 15). Due to construction, these options are unavailable.

Filter query results Compact view Hide row numbers Showing results from 0 to 3 of 3. Query took 13s, moments ago.

| | datetime | weekday_name | start_point | end_point | trip1 | route1 | A_stop | A_stop_name | B_stop | B_stop_name | A_stop_departu... | B_stop_arrival_t... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | "2024-11-04T07:30:00"^^xsd:dateTime | "Monday" | "ViaStazione-1-38057-PergineValsugana" | src:LargoMedaglieDOro-9-38122-Trento | src:trip-train-16924 | src:route-train-valsugana | src:stop-train-1009 | "levico" | src:stop-train-2264 | "trento" | "08:00:00" | "08:47:00" |
| 2 | "2024-11-04T07:30:00"^^xsd:dateTime | "Monday" | "ViaStazione-1-38057-PergineValsugana" | src:LargoMedaglieDOro-9-38122-Trento | src:trip-train-16922 | src:route-train-valsugana | src:stop-train-1009 | "levico" | src:stop-train-2264 | "trento" | "07:40:00" | "08:20:00" |
| 3 | "2024-11-04T07:30:00"^^xsd:dateTime | "Monday" | "ViaStazione-1-38057-PergineValsugana" | src:LargoMedaglieDOro-9-38122-Trento | src:trip-train-16108 | src:route-train-valsugana | src:stop-train-1009 | "levico" | src:stop-train-2264 | "trento" | "08:42:00" | "09:29:00" |

Figure 15: Feasible train connections from Levico to Trento on Monday, 4 November 2024.
Query code: CQ-5(P3–S6)-feasible-train-connections.sparql, Requires P3-S6_Triples.brf to run

##### 7.3.5.1 CQ-5 (P3-S6) – ab)

*What is the best route for her to take? Which is the travel time using the replacement bus?*

The query logic follows the same **two-leg journey modeling approach** adopted in earlier scenarios. Candidate journeys are retrieved from a **pre-materialized subgraph** containing feasible combinations of first-leg trips, transfer points, and second-leg trips. To reflect the disruption scenario, the query **explicitly filters for substitute bus services**, excluding regular train trips. Feasibility is ensured by verifying that both legs operate on the selected date, that the replacement bus departs after Elena's intended start time, and that the second leg departs after the arrival of the first one, within a bounded transfer window that also accounts for the required walking time. Finally, the query computes the total travel time and selects the journey with the earliest arrival at the destination.

The optimal solution consists of a **replacement bus from Levico to Trento**, followed by an **urban bus connection**, and a short walking segment to the hospital. Elena first boards the **replacement bus service (TT922)** at **Levico**, departing at **07:41** and arriving at **Trento railway station** at **08:20**. From there, she performs a short transfer, involving a **1-minute walk**, to reach the nearby urban bus stop at **Piazza Dante (Stazione FS)**. She then continues her journey on **urban bus route 404**, departing at **08:22** and arriving at **Verona–Fogazzaro** at **08:32**. The final stop is located approximately **5 minutes** on foot from **Ospedale Santa Chiara**, resulting in a total travel time of **56 minutes** (Fig. 16, Fig. 17).

Filter query results | Compact view ☐ | Hide row numbers ☐ | Showing results from 0 to 1 of 1. Query took 17s, moments ago.

| | datetime | weekday_na... | start_point | end_point | trip1 | route1 | A_stop | A_stop_name | B_stop | B_stop_name | A_stop_dep... | B_stop_arriv... | route_transf... | Bstop_to_C... | trip2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | "2024-11-04T07:30:00" ^^xsd:dateTime | "Monday" | "ViaStazione-1-38057-PergineValsugana" | src:LargoMedaglieDOro-9-38122-Trento | src:trip-train-substitute-bus-TT922 | src:route-train-valsugana | src:stop-train-1009 | "levico" | src:stop-train-2264 | "trento" | "07:41:00" | "08:20:00" | src:transfer-3 | "(0.09 km, 1 min)" | src:trip-url 00041971 2409092C 612 |

Figure 16: Replacement bus segment from Levico to Trento.
Query code: CQ-5(P3–S6)-ab.sparql, Requires P3-S6_Triples.brf to run

Filter query results | Compact view ☐ | Hide row numbers ☐ | Showing results from 0 to 1 of 1. Query took 17s, moments ago.

| | stop_dep... | B_stop_arriv... | route_transf... | Bstop_to_C... | trip2 | route2 | route_name2 | C_stop | C_stop_name | D_stop | D_stop_name | C_stop_dep... | D_stop_arriv... | Dstop_to_en... | travel_time |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 00" | "08:20:00" | src:transfer-3 | "(0.09 km, 1 min)" | src:trip-urb-00041971220240909202506 12 | src:route-urb-404 | "8" | src:stop-urb-247 | "Piazza Dante "Stazione Fs" | src:stop-urb-457 | "Verona Fogazzaro" | "08:22:00" | "08:32:00" | "(0.38 km, 5 min)" | "0:56 min" |

Figure 17: Urban bus segment and final walking access to Ospedale Santa Chiara.
Query code: CQ-5(P3–S6)-ab.sparql, Requires P3-S6_Triples.brf to run

#### 7.3.6 CQ-6 (P4-S6)

*Chiara finishes her shift at Ospedale S. Camillo at 3:00 p.m. and needs to travel to Centro Medico di Rovereto for a training session at 4:30 p.m., but there is ongoing construction causing delays on the usual bus route. What alternative bus routes can she use to reach the training center on time despite the disruption?*

CQ-6 (P4-S6) cannot be answered based on the current KG.
The Event Etype and its related properties, initially envisioned in the CQs and the ER diagram, were not formalized in the final Teleontology schema due to lack of direct representation within the available datasets. As a consequence, event-driven phenomena such as strikes and disruptions are not explicitly modeled in the KG (see Section 5.2).

#### 7.3.7 CQ-7 (P4–S4)

*Chiara is finishing her night shift at the hospital at 7:00 a.m. on a rainy weekday.*
*a) What public transportation options are available for her to get home after her shift?*
*b) If the bus arrives but is overcrowded due to the heavy rain, how long will she have to wait for the next available bus?*

##### 7.3.7.1 CQ-7 (P4-S4) – a)

*What public transportation options are available for her to get home after her shift?*

The query assumes a weekday scenario (**Thursday, 10 October 2024**) and computes feasible trips from her workplace area (**Via Benedetto Giovanelli**, near the hospital) to her home (**Viale degli Olmi**). The query integrates spatial, temporal, and scheduling constraints. First, it selects candidate trips connecting a boarding stop near the start location (**A-stop**) to an alighting stop near the destination (**B-stop**). It then filters trips based on weekday service calendars, taking into account both regular weekly schedules and possible schedule exceptions. Only trips operating on the specified weekday and within the valid service date range are considered. Temporal feasibility is ensured by comparing Chiara's departure time (**07:00**) with the scheduled departure time at the boarding stop. Finally, the query orders candidate solutions by walking access distance and departure time (Fig. 18).

Filter query results Compact view Hide row numbers Showing results from 0 to 1 of 1. Query took 2s, moments ago.

| | datetime | weekday_... | start_point | end_point | trip | route | bus_name | time_until... | start_poi... | A_stop | A_stop_n... | B_stop | B_stop_n... | A_stop_d... | B_stop_ar... | stop_count | Bstop_to_... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | "2024-10-10T07:00:00" ^^xsd:dateTime | "Thursday" | src: ViaBenedettoGiovanelli-19-38122-Trento | src: VialedegliOlmi-17-38123-Trento | src:trip-urb-0004202542024090920250612 | src:route-urb-541 | "6" | "0:11" | "(0.17 km, 2 min)" | src:stop-urb-277 | "Barbacovi "Itg Pozzo" | src:stop-urb-271 | "Asiago "Cernidor" | "07:11:00" | "07:16:00" | "4"^^xsd:integer | "(0.39 km, 5 min)" |

Figure 18: Optimal public transport journey from Via Benedetto Giovanelli to Viale degli Olmi on Thursday, 10 October 2024.
Query code: CQ-7(P4–S4)-a.sparql, Requires P4Trips.brf to run

#### 7.3.7.2 CQ-7 (P4-S4) – b)

*If the bus arrives but is overcrowded due to the heavy rain, how long will she have to wait for the next available bus?*

This query reuses the same spatial and scheduling constraints applied in CQ-7 a) but shifts the reference time to **07:11**, simulating the scenario in which Chiara is unable to board the first arriving bus due to overcrowding caused by heavy rain. The query searches for the next available departure from the **same boarding stop**. The waiting time is computed as the difference between the new reference time and the scheduled departure time at the boarding stop.

Specifically, starting from **07:11** on **Thursday, 10 October 2024**, the next feasible bus serving the boarding stop **Barbacovi "ITG Pozzo"** departs at **07:30** on **bus line 6**. Therefore, Chiara would have to wait **19 minutes** before being able to board the next bus (Fig. 19).

Filter query results Compact view Hide row numbers Showing results from 0 to 1 of 1. Query took 1.6s, moments ago.

| | datetime | A_stop | A_stop_name | trip | route | bus_name | time_until_departure | A_stop_departure_time |
|---|---|---|---|---|---|---|---|---|
| 1 | "2024-10-10T07:11:00"^^xsd:dateTime | src:stop-urb-277 | "Barbacovi "Itg Pozzo" | src:trip-urb-0004194742024090920250612 | src:route-urb-541 | "6" | "0:19" | "07:30:00" |

Figure 19: Waiting-time computation for the next available bus under overcrowding conditions.
Query code: CQ-7(P4–S4)-b.sparql, Requires P4Trips.brf to run

### 7.3.8 CQ-8 (P5–S2)

*Edoardo finishes his late shift at the restaurant at 1:00 a.m. and needs to stop by a 24-hour pharmacy to pick up sleep aids. Given the limited late-night bus service, what is the quickest way for him to reach the nearest open* *pharmacy and then head home to Caldonazzo?*

A preliminary query is executed to identify pharmacies that are open at **1:00 a.m.**. The query filters health facilities classified as *pharmacies* and checks their opening intervals against the query time. The result shows that **Farmacia Grandi**, located at **Via Alessandro Manzoni 7A, Trento**, is the only pharmacy in the KG that can be confirmed as open at that time, operating continuously from **00:00** to **23:59** on weekdays (Fig. 20).

Filter query results Compact view Hide row numbers Showing results from 0 to 5 of 5. Query took 0.1s, moments ago.

| | health_facility | position | legal_name | day | interval_start | interval_end |
|---|---|---|---|---|---|---|
| 1 | src:pharm-89 | src:ViaAlessandroManzoni-7A-38122-Trento | "Farmacia Grandi" | "Monday" | "00:00" | "23:59" |
| 2 | src:pharm-89 | src:ViaAlessandroManzoni-7A-38122-Trento | "Farmacia Grandi" | "Tuesday" | "00:00" | "23:59" |
| 3 | src:pharm-89 | src:ViaAlessandroManzoni-7A-38122-Trento | "Farmacia Grandi" | "Wednesday" | "00:00" | "23:59" |
| 4 | src:pharm-89 | src:ViaAlessandroManzoni-7A-38122-Trento | "Farmacia Grandi" | "Thursday" | "00:00" | "23:59" |
| 5 | src:pharm-89 | src:ViaAlessandroManzoni-7A-38122-Trento | "Farmacia Grandi" | "Friday" | "00:00" | "23:59" |

Figure 20: Identification of 24-hour pharmacies open at 1:00 a.m.
Query code: CQ-8(P5–S2)-01.sparql

A second query then uses **Farmacia Grandi** as the destination. It evaluates all public transport options departing after 1:00 a.m., filters them according to weekday validity and schedule exceptions, and selects the solution with the **minimum combined walking time** (from Edoardo's starting position to the boarding stop and from the arrival stop to the pharmacy). Additional constraints ensure that any selected service departs only after Edoardo has sufficient time to reach the stop on foot. The query results show that no public bus services are operating in the immediate late-night window around 1:00 a.m. The earliest feasible public transport option is **bus line 11**, departing from the stop **Gazzoletti – Piazza Dante** at **05:29** and arriving at **Largo Sauro** at **05:31** (Fig. 21).

Filter query results Compact view Hide row numbers Showing results from 0 to 1 of 1. Query took 56s, minutes ago.

| | DateTime | WeekDay | A | C | AC_Trip | AC_Route | BusNumber | A_stop | A_stop_na... | A_to_Asto... | C_stop | C_stop_na... | C_to_Csto... | Departure... | ArrivalTime | StopCount |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | "2024-11-07T01:00:00"^^xsd:dateTime | "Thursday" | src:PiazzettaBrunoLunelli-5-38122-Trento | "Farmacia Grandi" | src:trip-urb-0004203762024090920250612 | src:route-urb-425 | "11" | src:stop-urb-2680 | "Gazzoletti Piazza Dante" | "(0.35 km, 4 min)" | src:stop-urb-76 | "Largo Sauro" | "(0.10 km, 1 min)" | "05:29:00" | "05:31:00" | "1"^^xsd:integer |

Figure 21: Earliest feasible late-night public transport option after 1:00 a.m.
Query code: CQ-8(P5–S2)-02.sparql

Finally, a third query evaluates the feasibility of Edoardo's return journey from the selected 24-hour pharmacy back to his home in **Caldonazzo**. The earliest valid connections are available from **05:05**, both via train services on the **Valsugana line** and through regional bus routes (Fig. 22). This confirms that, due to limited night-time service availability, Edoardo must rely on walking during the night and wait until early morning to use public transport for his return trip.

| | DateTime | WeekDay | C | B | CB_Trip | CB_Route | BusNumber | C_stop | C_stop_na… | C_to_Csto… | B_stop | B_stop_na… | B_to_Bsto… | Departure… | ArrivalTime | StopCount |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | "2024-11-07T01:00:00"^^xsd:dateTime | "Thursday" | "Farmacia Grandi" | src:ViaGuglielmoMarconi-20-38052-Caldonazzo | src:trip-train-16105 | src:route-train-valsugana | "Valsugana" | src:stop-train-2264 | "trento" | "(0.51 km, 6 min)" | src:stop-train-1798 | "caldonazzo" | "(0.51 km, 6 min)" | "05:05:00" | "05:46:00" | "8"^^xsd:integer |
| 2 | "2024-11-07T01:00:00"^^xsd:dateTime | "Thursday" | "Farmacia Grandi" | src:ViaGuglielmoMarconi-20-38052-Caldonazzo | src:trip-urb-0003263362024093020241130 | src:route-ext-695 | "R25" | src:stop-ext-74 | "Trento. Staz. Fs" | "(0.42 km, 5 min)" | src:stop-ext-1102 | "Caldonazzo. Staz. Fs" | "(0.48 km, 6 min)" | "05:05:00" | "05:46:00" | "8"^^xsd:integer |
| 3 | "2024-11-07T01:00:00"^^xsd:dateTime | "Thursday" | "Farmacia Grandi" | src:ViaGuglielmoMarconi-20-38052-Caldonazzo | src:trip-train-16107 | src:route-train-valsugana | "Valsugana" | src:stop-train-2264 | "trento" | "(0.51 km, 6 min)" | src:stop-train-1798 | "caldonazzo" | "(0.51 km, 6 min)" | "05:35:00" | "06:16:00" | "8"^^xsd:integer |
| 4 | "2024-11-07T01:00:00"^^xsd:dateTime | "Thursday" | "Farmacia Grandi" | src:ViaGuglielmoMarconi-20-38052-Caldonazzo | src:trip-urb-0003257282024090920250627 | src:route-ext-520 | "B401" | src:stop-ext-74 | "Trento. Staz. Fs" | "(0.42 km, 5 min)" | src:stop-ext-1102 | "Caldonazzo. Staz. Fs" | "(0.48 km, 6 min)" | "05:35:00" | "06:16:00" | "8"^^xsd:integer |

Figure 22: Feasible return connections from Farmacia Grandi to Caldonazzo in the early morning.
Query code: CQ-8(P5–S2)-03.sparql

# 8 Metadata Definition

In this section the report collects the definitions of all the metadata defined for the different resources produced along the whole process. The metadata defined in this phase describes both the final outcome of the project, and the intermediate outcome of each phase (language, schema, and data source standardised values).

The definition of the metadata, is crucial to enable the distribution (sharing) of the resource produced, through the data catalogs. For this reason it is important to describe also where such metadata will be published to distribute the resources it describes (for example the DataScientia catalogs).

In particular the structure of this section is organized as follows, with the objective to describe the metadata relative to all the type of resources produced by the project.

- Project metadata description
- Language resources metadata description
- Knowledge resources metadata description
- Data resources metadata description

## 8.1 Project metadata description

| | |
|---|---|
| ProjTitle | Trentino Transportation & Health Facilities 2024 |
| ProjURL | `https://amassacci.github.io/HealthRoute_Trentino.html` |
| ProjKeyword | Transportation, Trento, Trentino, Health Facilities, Knowledge Graph |
| ProjType | Knowledge Graph Generation |
| ProjDescription | This project aims to develop a Knowledge Graph (KG) that provides access to up-to-date information about healthcare facilities across the Trentino Province, along with their integration into the public transportation network. |
| ProjStartDate | 2024-10-16 |
| ProjEndDate | 2026-01-30 |
| ProjFundingAgency | None |

Table 11: Metadata description of the Trentino Transportation & Health Facilities 2024 project

## 8.2 Language resources metadata description

| | |
|---|---|
| DatLicense | Apache-2.0 license |
| DatURL | `https://github.com/amassacci/HealthRoute-Trentino/blob/main/Phase%203%20-%20Language%20Definition/Language_Resources.csv` |
| DatKeyword | Concepts, UKC |
| DatPublisher | Massacci Alice |
| DatCreator | Massacci Alice |
| DatOwner | Massacci Alice |
| DatLanguage | English |
| DatSize | 7.65 KB |
| DatName | Language_Resources.csv |
| DatPublication | 2024-12-10 |
| DatDescription | Description of all the language used in the HealthRoute-Trentino Knowledge Graph |
| DatDomain | Transport–Health |
| DatFileFormat | CSV |

Table 12: Metadata description of the Language Resources

For more details about the language resources metadata, refer to the following tables: E-types (Table 2), Data properties (Table 3), and Object properties (Table 4).

### 8.3 Knowledge resources metadata description

| | |
|---|---|
| DatLicense | CC0 1.0 DEED |
| DatURL | `https://github.com/amassacci/HealthRoute-Trentino/blob/main/Phase%204%20-%20Knowledge%20Definition/teleontology/teleontology.rdf` |
| DatKeyword | Teleontology, HealthRoute-Trentino |
| DatPublisher | Massacci Alice |
| DatCreator | Massacci Alice |
| DatOwner | Massacci Alice |
| DatLanguage | English |
| DatSize | 47.5 KB |
| DatName | teleontology.rdf |
| DatPublication | 2026-02-03 |
| DatDescription | The Teleontology file contains entity types, data properties, and object properties within the HealthRoute-Trentino. |
| DatDomain | Transport–Health |
| DatFileFormat | RDF |

Table 13: Metadata description of the Knowledge Resources

### 8.4 Data resources metadata description

For details about the data resources metadata, refer to the Dataset Metadata Sheet on the GitHub repository.

# 9 Open Issues

This project presented the design and implementation of the HealthRoute Trentino Knowledge Graph that combines public transport data, healthcare facility information, and user-specific spatial and temporal constraints to support journey planning for healthcare access.

The final results are able to satisfy the initial purpose to a large extent. The KG successfully supports multi-modal journeys, route transfers, service availability, temporal constraints, and spatial proximity reasoning (e.g., walking distances to and from stops and healthcare facilities) for accessing healthcare services.

Despite the successful construction of the KG and the systematic application of the iTelos methodology, several open issues remain open at the end of the project. A primary open issue concerns query performance and scalability. The complexity of SPARQL queries required to support multi-leg journeys with multiple temporal, spatial, and semantic constraints led to long execution times in some cases. To address this, a two-phase query strategy and intermediate materialization were adopted as partial solutions. While this approach improved manageability and clarity, it does not fully solve performance limitations.

The Event Etype and its related properties, initially envisioned in the CQs and the ER diagram, were not formalized in the final Teleontology schema due to lack of direct representation within the available datasets. As a consequence, event-driven phenomena such as strikes and disruptions are not explicitly modeled in the KG. This results in partial coverage of the originally defined competency questions.

Finally, the availability and completeness of certain operational data, particularly the opening hours of pharmacies, represent a further limitation. In the integrated datasets, opening hours information is sparse and often missing. This reduces the reliability of queries that depend on temporal availability of healthcare facilities and limits the accuracy of recommendations based on time constraints. Improving data completeness through the integration of higher-quality sources would enhance the KG's ability to support time-sensitive healthcare access scenarios.