KGE 2023

Trentino Territory & Transportation

Rubens Rissi Onzi   rubens.rissionzi@unitn.it
Eugenio Ferrari   eugenio.ferrari-1@studenti.unitn.it

Project Purpose

- The main idea of the project was to try to predict possible delays of the bus bus lines in the city of trento

Project Purpose

- Provide data to applications and services that need information suitable for predicting bus transportation delays in Trento.
- Incorporate data that could affect transportation delays in urban areas:
    - facilities;
    - population density;
    - length and number of right and left turns in the bus path;
- We've integrated the data into knowledge graphs (KGs), utilizing bus transportation and available territorial data as the foundation.

Domain of Interest

- Bus transportation data in Trento urban areas, covering the period from 1 September 2022, to 1 October 2023.
- Territorial and facilities data from Trentino OSM Places, collected up to 28 February 2023.
- Population data for the year 2018, sourced from the most recent survey conducted by the Comune di Trento.

Purpose Formalization

- 4 Scenarios
- 4 Personas
- 6 Competency Questions. Examples:
    - Isabella, after lunch, wants to reach the city center, where she can find lot of shops to buy groceries to prepare sweets to her daughter. She wants to know how much time it is gonna take to reach the center, and arrive home for dinner.
    - After enjoying a dinner with his friends, Giovanni decide to head to one of his friend's houses in Martignano. They are fortunate to be right on schedule for the last bus. The buses to Martignano are usually punctual, as the area had fewer residents compared to the city center, resulting in less traffic. They want to confirm if the bus will run on time so that they can arrive at their destination on time.

ER Model
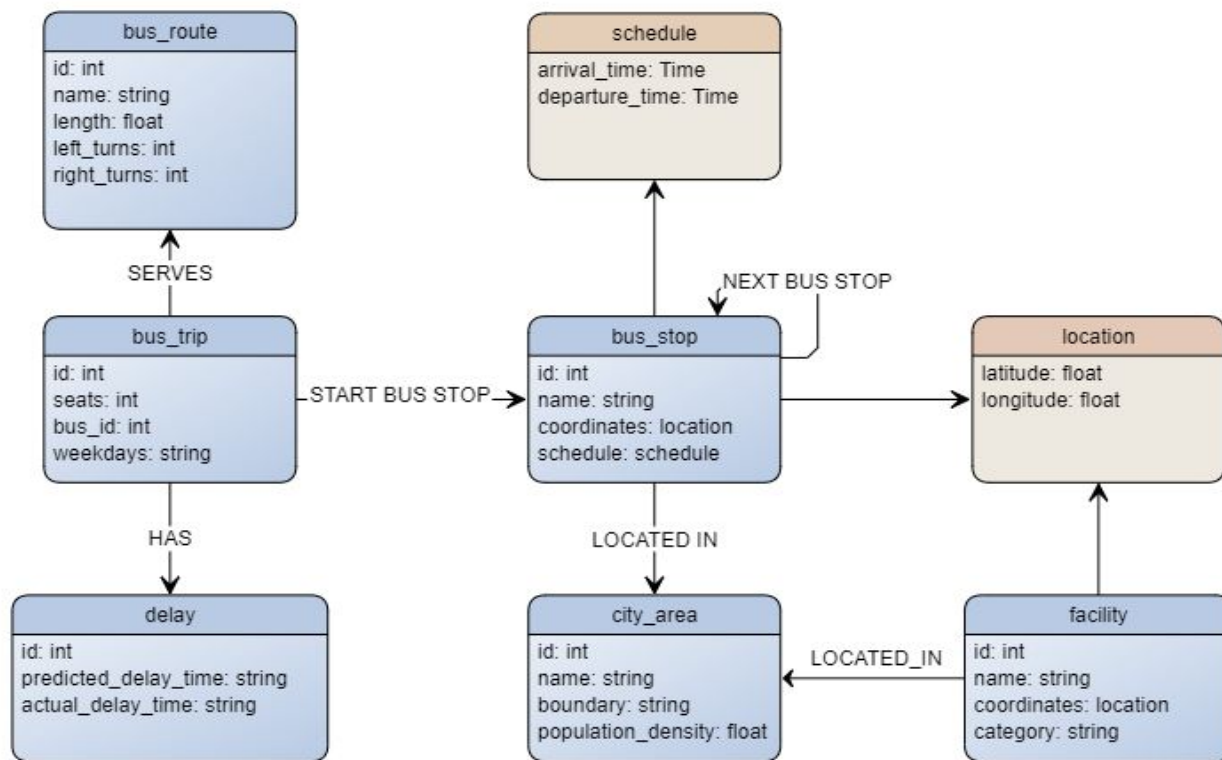
Information Gathering

- Bus Transportation: since the KGE 2022 - Urban Transportation was missing some data for achieving our purpose, we decided to use [Trentino Trasporti Open data](#).

- City areas: from [City of Trento Districts](#), it contains the official subdivision of Trento with the name and the boundaries of each area.

- Population: from [City of Trento statistical data](#), it contains PDFs about the most recent survey on the population in Trento divided by the areas.

- Facilities: from [City of Trento Economic activities](#) we identified where the facilities are located.

Parsing the Trentino Trasporti opendata

For the data parsing we used the areas of each circoscrizione and checked which ones were inside and which were not.

For this we used pandas and opencv.

Creating the route distance

For each route distance we used the points for each shape from the trentino trasporti open data. Using each point added in the shape file we determined the it's length.

Language Definition

This step aims at formalize the concepts of the Etypes and properties to GIDs (General Identifiers) from the UKC (Universal Knowledge Core).

As the main objective of this project is to predict delays, here we have an example with the delay etype formalization:

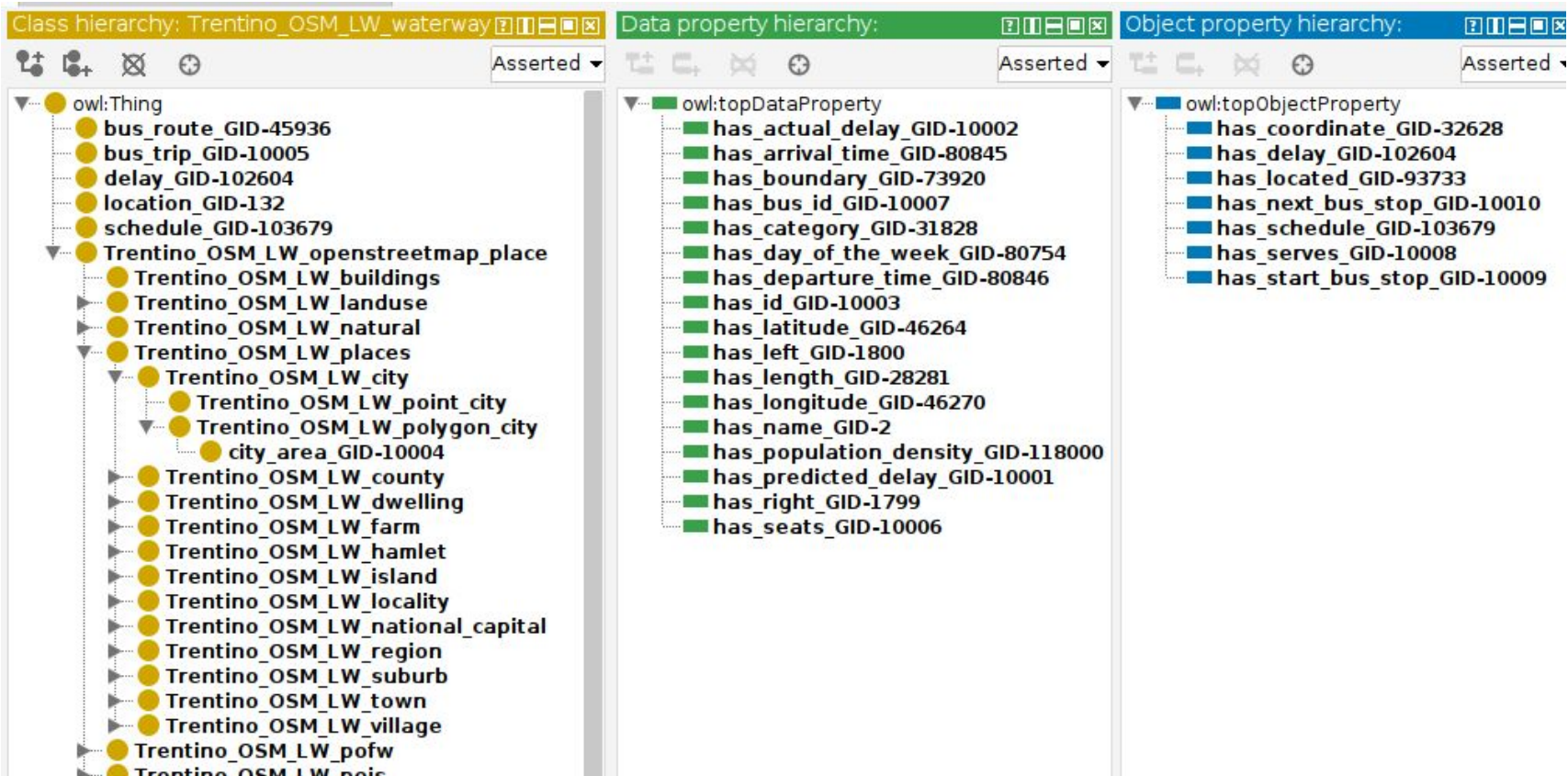| Concept Labels	Description
delay_GID-102604	cause to be slowed down or delayed
id_GID-10003	unique identifier, being it any entity, for it's collection
predicted delay_GID-10001	delay predicted by any mean of any event until it's end
actual delay_GID-10002	actual delay of any event until it's end

Knowledge Definition

- Top-Down: reuse of a Lightweight Ontology, [Trentino OSM LW Ontology](#)
- Bottom-Up: modelling of a Teleology
- Middle-Out: aligning of a Teleology grounded into the Lightweight Ontology to generate a Teleontology
- Knowledge annotation

# Teleology in Protégé

# Teleontology in Protégé

Data Definition

-   Entity Matching. Not needed for us, because we didn't had entities coming from multiple datasets.
-   Entity Identification. Not needed for us, because each entity already had an unique id assigned during the previous phases of the project.
-   Entity Mapping. Link the data to the teleontology.

Facility mapping with Karma
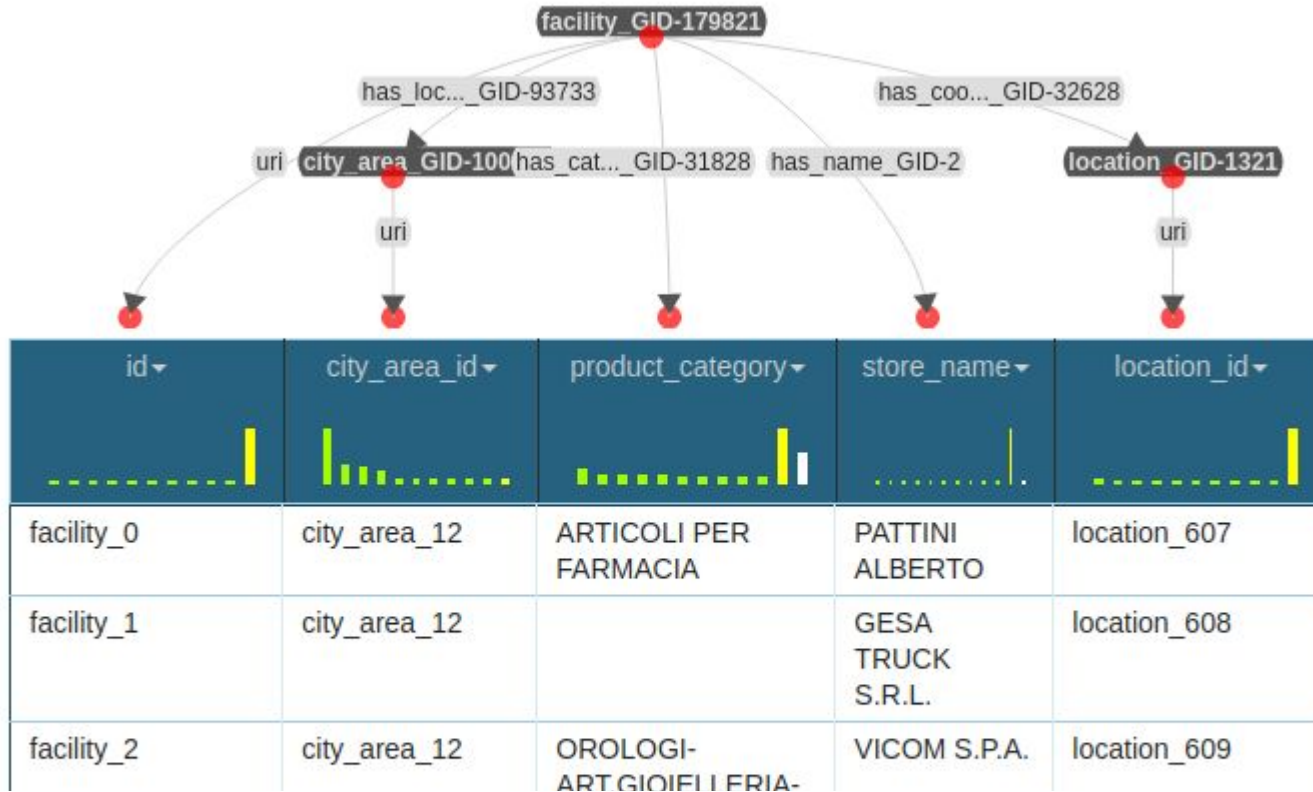
Evaluation

For the item ER evaluation we compare our created teleology and the ER created in the beginning of the project. In the ER we have 6 entities and in the competency questions we have 8 entities as well (as facility is a super class of some). The quantity size of the competency questions was modified from 6 to 8 to include schedule and location.
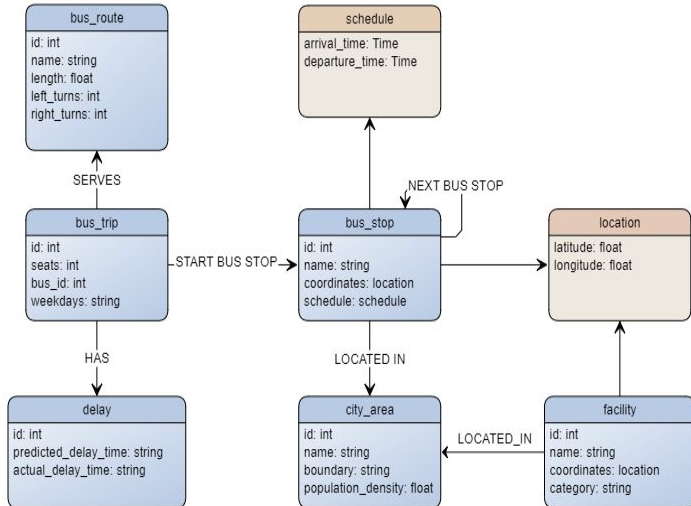
Scenarios	Personas	Competency Questions	Entities	Properties	Focus	Popularity
1 - 2 - 3	2	1	bus_stop	(id: int, name: string, coordinates: location, arrival_times: schedule)	Contextual	Common
1 - 2 - 3	2	1	delay	(id: int, predicted_delay_time: string, actual_delay_time: string)	Contextual	Contextual
1 - 2 - 3	2	1	supermarket	(id: int, name: string, coordinates: string,)	Common	Common
1 - 3 - 4	1	2	education_facilities	(id: int, name: string, coordinates: string,)	Common	Common
1 - 3 - 4	2	3	bus_route	(id: int, name: string, length: float, left_turns: int, right_turns: int,)	Contextual	Contextual
1 - 3 - 4	2	3	city_area	(id: int, name: string, boundary: string, population_density: float)	Core	Contextual
1 - 2 - 3 - 4	3	3	tourism_destinations	(id: int, name: string, coordinates: string,)	Common	Common
1 - 2 - 4	1	4	catering	(id: int, name: string, coordinates: string,)	Common	Common
1 - 3	4	5	bus_trip	(id: int, seats: int, bus_id: int, weekdays: string,)	Contextual	Core

$$Cov_E R = \frac{|Cq_E \cap T_E|}{Cq_E} = \frac{|8|}{8} = 1$$

Evaluation

In the evaluation of the proprieties of the Teleology is made the same way as in last slide. This results reflects on the lack of data in the Delay propriety and not having the seats number propriety for the buses in their route.

Being the ones missing predicted_delay, actual_delay, left_turn, right_turn and seats.

$$Cov_E R = \frac{|C_{qp} \cap T_p|}{C_{qp}} = \frac{|16|}{21} = .76$$

Evaluation

Now, with the evaluation of the teleontology and the ontologies we use the Trentino Trasporti Open data reference ontology, that contain 9 entities. This Reference Ontology has 4 Entities in common with our Teleontology.

Being common ones route, trip, schedule and bus stop.

$$Cov_E = \frac{|RO_E \cap T_E|}{RO_E} = \frac{|4|}{9} = .44$$

Evaluation

For the evaluation of the teleontology and the ontologies, now on the propriety side, we use the Trentino Trasporti Open data reference ontology, that contain 43 proprieties and object properties. This Reference Ontology has 8 Properties in common with our Teleontology

Being common ones longitude, latitude, name(bus stop), arrival time, departure time, weekdays, route short name and direction.

$$\frac{|RO_P \cap T_P|}{RO_P} = \frac{|8|}{46} = .17$$

Open Issues

- No data available for Delay, for this reason we used mocks for the queries
- No good data on services on the city of Trento
- Missing Right and left turns

Demo

We chose Competency Questions 1 and 5 because they effectively demonstrate the integration across the various datasets:

- Bus transportation and bus route structure
- Facilities
- Population density

Demo - CQ 1

Isabella, after lunch, wants to reach the city center, where she can find lot of shops to buy groceries to prepare sweets to her daughter. She wants to know how much time it is gonna take to reach the center, and arrive home for dinner.

Assumptions:

- after lunch, 13:30 - 14: 30
- Isabella home from Personas is Povo, Spré Pinara bus stop
- city center, Venezia "Port'Aquila" bus stop

```sparql
SELECT DISTINCT ?trip ?firstStopOfTrip ?busRoute ?routeName ?routeLength ?actualDelay ?predictedDelay ?startStop ?destinationStop
?destinationStopSchedule ?destinationStopArrivalTime ?cityArea ?cityAreaName (COUNT(?facility) as ?facilityCount)
WHERE{
    ?startStop rdf:type etype:bus_stop_GID-45937;
        etype:has_name_GID-2 "Spré Pinara" .

    ?destinationStop rdf:type etype:bus_stop_GID-45937;
        etype:has_name_GID-2 "Venezia \"Port'aquila\"" .

    ?destinationStop etype:has_schedule_GID-103679 ?destinationStopSchedule .
    ?destinationStopSchedule etype:has_arrival_time_GID-80845 ?destinationStopArrivalTime .

    FILTER (STRDT(STR(?destinationStopArrivalTime), xsd:time) > "13:30:00"^^xsd:time &&
        STRDT(STR(?destinationStopArrivalTime), xsd:time) < "14:30:00"^^xsd:time)
```

```
    ?firstStopOfTrip rdf:type etype:bus_stop_GID-45937 .
    ?trip rdf:type etype:bus_trip_GID-10005;
          etype:has_serves_GID-10008 ?busRoute;
          etype:has_delay_GID-102604 ?delay;
          etype:has_start_bus_stop_GID-10009 ?firstStopOfTrip.

    ?firstStopOfTrip etype:has_next_bus_stop_GID-10010* ?startStop .
    ?startStop etype:has_next_bus_stop_GID-10010* ?destinationStop .

    ?busRoute etype:has_name_GID-2 ?routeName;
              etype:has_length_GID-28281 ?routeLength .

    ?delay etype:has_actual_delay_GID-10002 ?actualDelay;
           etype:has_predicted_delay_GID-10001 ?predictedDelay .

    ?destinationStop etype:has_located_GID-93733 ?cityArea .
    ?cityArea etype:has_name_GID-2 ?cityAreaName .

    ?facility rdf:type etype:facility_GID-17982;
              etype:has_located_GID-93733 ?cityArea .
}
GROUP BY ?trip ?firstStopOfTrip ?busRoute ?routeName ?routeLength ?actualDelay ?predictedDelay ?startStop ?destinationStop
?destinationStopSchedule ?destinationStopArrivalTime ?cityArea ?cityAreaName
LIMIT 5
```

	trip	firstStopOfTrip	busRoute	routeName	routeLength	actualDelay	predictedDelay	startStop	destinationStop	destinationStopSchedule	destinationStopArrivalTime	cityArea	cityAreaName	facilityCount
1	http://localhost:8080/source/bus_trip_0003988502023091120240611	http://localhost:8080/source/bus_stop_161	http://localhost:8080/source/bus_route_400	"5"	"8.594295448800244"	"38.35307587388043"	"55.96302612152342"	http://localhost:8080/source/bus_stop_180	http://localhost:8080/source/bus_stop_2927	http://localhost:8080/source/schedule_238	"13:36:00"	http://localhost:8080/source/city_area_11	"S.GIUSEPPE-S.CHIARA"	"278"^^xsd:integer
2	http://localhost:8080/source/bus_trip_0003988502023091120240611	http://localhost:8080/source/bus_stop_161	http://localhost:8080/source/bus_route_400	"5"	"8.594295448800244"	"38.35307587388043"	"55.96302612152342"	http://localhost:8080/source/bus_stop_180	http://localhost:8080/source/bus_stop_2927	http://localhost:8080/source/schedule_247	"13:46:00"	http://localhost:8080/source/city_area_11	"S.GIUSEPPE-S.CHIARA"	"278"^^xsd:integer
3	http://localhost:8080/source/bus_trip_0003988502023091120240611	http://localhost:8080/source/bus_stop_161	http://localhost:8080/source/bus_route_400	"5"	"8.594295448800244"	"38.35307587388043"	"55.96302612152342"	http://localhost:8080/source/bus_stop_180	http://localhost:8080/source/bus_stop_2927	http://localhost:8080/source/schedule_248	"13:47:00"	http://localhost:8080/source/city_area_11	"S.GIUSEPPE-S.CHIARA"	"278"^^xsd:integer
4	http://localhost:8080/source/bus_trip_0003988502023091120240611	http://localhost:8080/source/bus_stop_161	http://localhost:8080/source/bus_route_400	"5"	"8.594295448800244"	"38.35307587388043"	"55.96302612152342"	http://localhost:8080/source/bus_stop_180	http://localhost:8080/source/bus_stop_2927	http://localhost:8080/source/schedule_252	"13:53:00"	http://localhost:8080/source/city_area_11	"S.GIUSEPPE-S.CHIARA"	"278"^^xsd:integer

Demo - CQ 5

After enjoying a dinner with his friends, Giovanni decide to head to one of his friend's houses in Martignano. They are fortunate to be right on schedule for the last bus. The buses to Martignano are usually punctual, as the area had fewer residents compared to the city center, resulting in less traffic. They want to confirm if the bus will run on time so that they can arrive at their destination on time.

Assumptions:

- after dinner, 20:00
- From "Port'Aquila" bus stop
- To "Martignano P.zza Meneghin" bus stop

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

SELECT DISTINCT ?trip ?busRoute ?routeName ?routeLength ?actualDelay ?predictedDelay ?startStop ?destinationStop ?startStopDepartutreTime
?cityAreaName ?populationDensity
WHERE{
    ?startStop rdf:type etype:bus_stop_GID-45937;
        etype:has_name_GID-2 "Cervara \"Port'aquila\"" .

    ?destinationStop rdf:type etype:bus_stop_GID-45937;
        etype:has_name_GID-2 "Martignano P.Zza Menghin" .

    # get the departure time from the start stop
    ?startStop etype:has_schedule_GID-103679 ?schedule .
    ?schedule etype:has_departure_time_GID-80846 ?startStopDepartutreTime .

    # departure after dinner
    FILTER (STRDT(STR(?startStopDepartutreTime), xsd:time) > "20:00:00"^^xsd:time)

    ?firstStopOfTrip rdf:type etype:bus_stop_GID-45937 .
    ?trip rdf:type etype:bus_trip_GID-10005;
        etype:has_serves_GID-10008 ?busRoute;
        etype:has_delay_GID-102604 ?delay;
        etype:has_start_bus_stop_GID-10009 ?firstStopOfTrip.

    ?firstStopOfTrip etype:has_next_bus_stop_GID-10010* ?startStop .
    ?startStop etype:has_next_bus_stop_GID-10010* ?destinationStop .
```

```
        ?busRoute etype:has_name_GID-2 ?routeName;
                  etype:has_length_GID-28281 ?routeLength .
        FILTER (?routeName = "10") # to martignano

        ?delay etype:has_actual_delay_GID-10002 ?actualDelay;
               etype:has_predicted_delay_GID-10001 ?predictedDelay .

        ?destinationStop etype:has_located_GID-93733 ?cityArea .
        ?cityArea etype:has_name_GID-2 ?cityAreaName;
                  etype:has_population_density_GID-118000 ?populationDensity.
}
LIMIT 5
```

	trip	busRoute	routeName	routeLength	actualDelay	predictedDelay	startStop	destinationStop	startStopDepartureTime	cityAreaName	populationDensity
1	http://localhost:8080/source/bus_trip_00039903220230911120240611	http://localhost:8080/source/bus_route_408	"10"	"4.831162794882017"	"14.591257612850441"	"43.88784615349611"	http://localhost:8080/source/bus_stop_177	http://localhost:8080/source/bus_stop_107	"20:40:00"	"ARGENTARIO"	"9.56"
2	http://localhost:8080/source/bus_trip_00039903220230911120240611	http://localhost:8080/source/bus_route_408	"10"	"4.831162794882017"	"14.591257612850441"	"43.88784615349611"	http://localhost:8080/source/bus_stop_177	http://localhost:8080/source/bus_stop_107	"20:20:00"	"ARGENTARIO"	"9.56"
3	http://localhost:8080/source/bus_trip_00039903220230911120240611	http://localhost:8080/source/bus_route_408	"10"	"4.831162794882017"	"14.591257612850441"	"43.88784615349611"	http://localhost:8080/source/bus_stop_177	http://localhost:8080/source/bus_stop_107	"20:24:00"	"ARGENTARIO"	"9.56"
4	http://localhost:8080/source/	http://localhost:8080/source/	"10"	"4.831162794882017"	"14.591257612850441"	"43.88784615349611"	http://localhost:8080/source/	http://localhost:8080/source/	"20:10:00"	"ARGENTARIO"	"9.56"