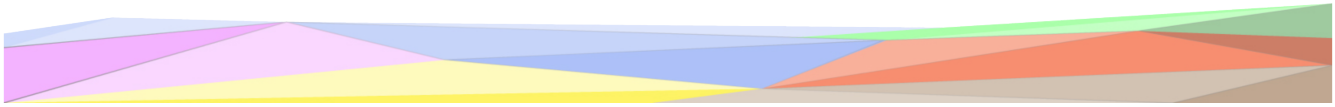
Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

Transportation KG [KDI 2020-21]

Document Data:	Reference Persons:
- date -	- authors -



Contents

Revision History:

Revision	Date	Author	Description of Changes
1.0	17.10.2020	Fivos Kapidis, Antonio Stefani	Draft of the Scope Definition
1.1	18.10.2020	Fivos Kapidis, Antonio Stefani	First revision of the Scope Definition
2.0	18.10.2020	Fivos Kapidis	Draft of the Scenario and of the Storytelling Definition
2.1	18.10.2020	Antonio Stefani	First revision of the Scenario and of the Storytelling Definition
3.0	19.10.2020	Fivos Kapidis	Draft of the CQs
3.1	19.10.2020	Fivos Kapidis, Antonio Stefani	First revision of the CQs
4.0	19.10.2020	Fivos Kapidis, Antonio Stefani	Draft of the Inception Schema
5.0	19.10.2020	Omid Jadidi	Draft of the dataset description
3.2	20.10.2020	Fivos Kapidis, Antonio Stefani	Second revision of the CQs
4.1	20.10.2020	Fivos Kapidis	First revision of the Inception Schema
5.1	20.10.2020	Omid Jadidi	First revision of the dataset description
6.0	20.10.2020	Alberto Carbognin	Draft of the metadata documentation
2.3	21.10.2020	Antonio Stefani	Second revision of the Scenario Description and of the Storytelling Definition
3.3	21.10.2020	Antonio Stefani	Third revision of the CQs
4.2	21.10.2020	Antonio Stefani	Second revision of the Inception Schema
6.1	21.10.2020	Alberto Carbognin	First revision of the metadata documentation

1 Knowledge Graph Codebook

The first of the two sections, in the current document, contains the codebook of the whole KG (Knowledge Graph), including the description of all the data and information that it contains.

1.1 Knowledge Graph general description

This sub section aims to give a general description of the KG, reporting:

- the context/domain in which the KG lives and works;
- *The Problem* the KG aims to solve;
- How the KG can solve *The Problem*

1.2 Data level

The data level section aims to describe in details the (final version of) datasets collected and managed by the KG, with a description of each variable involved.

1.2.1 Datasets general details

In this section are reported the metadata at datasets level, so the metadata regarding the sources, the authors, the collection methods, and so on.

1.2.2 Datasets metadata documentation

In this section are reported the metadata at dataset attribute level, through a description of each variable involved in the datasets collected, specifying the variable types, meanings, value-set (possible values), and every other meaningful variable information.

1.3 Ontology level

The ontology level section aims to describe the underlying KG ontology, through the description of its elements at each level, reporting so the language, conceptual and schema resources used within it.

1.3.1 Ontology general details

This first sub section of the ontology level description, report the general details such as authors, sources and the description of external ontology eventually adopted to generate the final one.

1.3.2 Ontology metadata documentation

In this section instead, are reported the more specific metadata describing the single elements of the ontology (terms, concepts, ETypes and relations).

1.4 Knowledge Graph Evaluation

In the final section of this first chapter, the KG Evaluation is reported. It aims to describe, through specific metrics, the quality of the overall KG on different aspect, like domain coverage, usability, domain representation, and other meaningful aspects.

2 Knowledge Graph Development Process

The second chapter of this document aims to describe, in a detailed way, the KG development process. The sections below describe each phase of the KG building project, reporting for each phase, the description of the datasets and their evolution respect the previous phases, the schema construction which will generate the KG ontology in the end, as well as the description of the procedures adopted to manage the data and finally achieve those results. Moreover for each phase is reported an evaluation section, which aims to evaluate the quality of the results achieved at the end of each phase.

2.1 Scope Definition

Even if working, studying, visiting a city or a relative, practicing sports, tacking a trip seem totally different activities they have one common thing: they take us out of our houses. To get to our workplace, school or gym we have to move and this makes us spend one of the greatest parts of our day on transportation. To save even a little of our time we often spend much of it looking for a faster and a cheaper way to get to our destination. Transportation involves several parameters that each traveler evaluate carefully based on his needs. The main parameters taken into account are:

- routes
- modes of transport
- time spent
- cost

Our system wants to provide a solution able to solve two different problems in the Trentino area: giving all the useful information on the way of transport and providing all those useful facilities related to the road system. To do this our solution is based on an integration of data regarding the road system like routes, time spent on each path, public transports schedules, and its relative facilities like parking areas, petrol pumps, campsites.
To better link all the available datasets and create a system able to give important information instantaneously, the best way to proceed is using a Knowledge Graph. To construct it is very important to have a standardized methodology like the iTelos one. Thanks to the iTelos Methodology indeed, it is possible to divide the problem in several sub-problems being able to solve each one in an easier way. Those parts are, in particular:

- Inception, in this step the goal is to define the queries that could be posed and how to shape up the system itself.
- Informal Modeling, in this step the goal is to design an EER Model such to highlight all possible relations among data and entities.
- Formal Modeling, in this step the goal is to build up the real KG provided of a SKG and a LKG.

- Data Integration, in this step the goal is to finalize all the system providing a final DKG with the reference to the LKG and the SKG.

But before proceeding with the system design and its implementation there is a step 0 called "Scope Definition". The goal of this step is to provide a description of the scenario. To achieve this goal the idea is to describe several situations in which it is shown how the system can be extremely relevant to solve certain problem.

2.1.1 Scenario Description

In order to show in a better way the application scenario and how the system interact with people, let define some personas which could be used to generate examples of competency queries.
Claudio: he is currently studying at the High School and since he is still too young to drive a vehicle he always moves using public transport. In particular he makes use of it to go to his school which is located in the centre of the city while he lives in its suburbs. Claudio has always practiced basketball and three times a week he goes training and once a week he have a match with other locals basketball teams. In the weekend he enjoys spending time with his friends at some locals parks or hanging out to a nearby basketball court.
Andrea: he is currently working as salesman of a big company so he travels a lot all around the Trentino area. To move he usually makes use of the company car but sometimes, if his destination is really close to a railway station or a bus station he takes the public transport. In his free time he enjoys going hiking in the mountains and exploring new paths.
Elisa: she is a university off-premise student. To reach her faculty she often makes use of public transportation, she makes use of it also to go back home. She lives in the center of the city so, to move easier, she uses her bicycle or during the summer she likes walking. Elisa is very attached with her family so once every two weeks she goes back home to visiting them. In her free time she likes visiting the historical city centers of the nearby cities, while when it is warmer she likes exploring cycle paths.
Maria: she is currently working in Bolzano but she lives in Trento with her husband and her children. She is a very ordinary mother: she is always in rush. During the holiday she likes to plan her family trips and usually they travel by camper. Once a month they usually go to visit her mother in Molveno.

2.1.2 Storytelling Definition

Claudio studies at the Trento Scientific Highschool "Leonardo Da Vinci" and he lives in Gardolo, due to this every morning he must take the bus to go to the school. Claudio enjoys basketball very much and since he was a child he has always played with a team at the Sanbàpolis gym on the south of Trento.
His team is currently competing in a regional basketball tournament so once a week he plays against other teams. Sometimes the match location us pretty far so he has to take public transports to get there and the evaluating of the several routes and modes of transport could be very annoying and time-consuming. In order to save some of his time, Claudio often asks for a lift to his parents but they are not always available.
In his limited free time Claudio likes hanging out with his friends spending some hours along the Adige shore or in some other parks near the city center, therefore he often checks the bus schedule to get there.

Andrea is a salesman at "Dolomiti Energia", an energy provider company based on Trento. Since he has started working there, he spends much of his working day travelling around to the several agencies collocated in: Mezzolombardo, Pergine Valsugana, Rovereto and many others. To do these tours a company car has been provided to him.

Another main part of Andrea's job consists in writing emails and check stuff on the web, for this reason when he has a lot of work to do and he needs to use his laptop he often decides to take trains in order to feel sufficiently comfortable to work. Since the tours change from one day to another he spends a lot of time checking the train schedule and this is often the most annoying part of his job.
Andrea is a mountain lover and in his free time he likes hiking in order to explore new mountains paths of nearby mountains. Since he does not have its own car he often makes use of public transports as buses to get there.

Elisa is studying at the University of Trento, since she is from Tuscany she is currently living in Rovereto as off-site student even if she studies Sociology in Trento. All her activities are carried out in the city of Trento and often she involves also her boyfriend, they likes very much walking on the Adige riverside and often they take the cable car to go to the Bondone.
Elisa is an eco-friendly person, she really likes going around using just her bike and indeed she uses it to go to the supermarket, to the library, to the coffee and all those other places which she enjoys. Even when she has not the possibility to take her bike she often use the "e-motion" service to move around. But while for her ordinary movement she takes the bike, to go to the university or to go back home she has always to use the public transports and since her schedules change everyday she spends a lot of time to plan very well every trip in order to save both time and money.
In her free time she likes very much camping with her friends and since she has acquired all the skills needed to plan a cheap tour she is always picked as planner of their trips. Elisa is very good at it in fact she always evaluate the cheapest and shortest way to get to the selected campsite, moreover she always think to what they could need during the trip as hospitals, supermarkets, fuel station.

Maria is a 58 years old woman who for the last three years has worked in Bolzano. She has always lived in Trento with her husband and her children and even if the movement between Trento and Bolzano has always been annoying she has never wanted to move there. To respect her daily routine she has always preferred to move autonomously using the family car but when her husband needs it she use the car sharing service. Due to all this rush she is always careful on the best route to take and she usually spend a lot of time by studying the several possible paths to get her destination.
During the family holidays she often rent a camper in order to visit all the Trentino and sometimes to move also to other countries. Even if she knows that this kind of holiday can be more expensive than staying in an hotel she really likes planning the several routes to be taken and the evaluating of the several rest areas.
One of her children really likes the cable cars, so during the weekend she often goes with her family to Sardagna. Moreover, once a month, they all go to visit her mother who lives on the shore of the Molveno Lake and there they often take a walk all around the lake.

2.1.3 People Description

In the following table we want to resume the main interests of the people described in the storytelling section highlighting in particular the usage of the system we want to provide.

Name	Age	Interest	Usage
Claudio	17	Go to school, to his basketball training or matches all around the Trentino. Hang out with his friends	Check the public transportation schedules and the position of the facility he needs
Andrea	35	Go to the work by car or sometimes by public transports. Go hiking in the Trentino mountains	Check the fastest ways to get his several destination. Planning his hiking
Elisa	23	Go to the university, go cycling and camping with friends	Check cycling routes and facilities, check public transportation schedules and cheaper campsites
Maria	58	Go to the work by using her car or the car-sharing service, save as much time as possible. Visit the Trentino during the holidays	Check fastest and cheapest routes, check the schedule of the public transports, check the position of all the facilities she needs

2.2 Inception

2.2.1 CQs definition

Starting from the people described in the scenario we have imagined which could be their needs and tried to identify those possible questions which could be done to our system. The following table contains person doing the questions and the possible answer that the system should provide.

Person	CQ #	Question	Answer
Claudio	1.1	List all the possible paths from Gardolo to Trento	Return all the routes from Gardolo Square to Trento Square highlighting modes of transport, time spent on each route, prices of each path, useful facilities available encountered in the path
Claudio	1.2	List each path from Gardolo to Trento duration	Return the "duration" associated to each path from Gardolo Square to Trento Square
Claudio	1.3	List all the train itineraries from Gardolo to Trento	Return paths with associated route, time spent, price, Railway Stations and mode of transportation that has field type: "train" from the closest Railway Station to Gardolo Square to the closest Railway Station to Trento Square
Claudio	1.4	List all the parks location in Trento	Return all the locations and addresses associated to each park in Trento
Claudio	1.5	List all the possible ways to get a specific Park	Return all the possible paths with associated routes, modes of transportation, time spent and prices from Trento Square to the Park

Person	CQ #	Question	Answer
Claudio	1.6	Give the cost of the cheapest path from Gardolo to Trento	Find the path from Gardolo Square to Trento Square with the associated smallest price and return the value of "price"
Claudio	1.7	List all the bus itineraries from Gardolo to the High School "Leonardo Da Vinci"	Return the paths with associated route, time spent, price, buses stops which have the mode of transport field as: "bus" from Gardolo Square to the High School "Leonardo Da Vinci"
Claudio	1.8	Give the time spent of the fastest path from Gardolo to Trento	Find the path from Gardolo Square to Trento Square with the associated smallest time spent and return the value of "time spent"
Claudio	1.9	List all the costs from Gardolo to Trento	Return the associated "price" value of each path from Gardolo Square to Trento Square
Claudio	1.10	List all the Bus Stops from Gardolo to Trento	Find all the paths from Gardolo Square to Trento Square which have the mode of transport field as: "bus" and return for each path the location, address and schedule of the stops
Claudio	1.11	Give the fastest path from Gardolo to the High School "Leonardo Da Vinci"	Return route, mode of transport, price, time spent and the list of useful facilities with the lower associated field: "duration" from Gardolo Square to the High School "Leonardo Da Vinci"
Claudio	1.12	Give the schedule of a specific bus stop in the weekend	From the stop with the specific "StopID" return the schedule of "Saturday" and "Sunday"
Claudio	1.13	Give the closest bus stop to a specific location	Search in a range of 5km all the bus stops and return location, address, schedule and route associated to the bus stop in which the value of the route field "length" is the lowest
Claudio	1.14	Give the closest train stop to a specific location	Search in a range of 5km all the Railway Stations and return location, address, schedule and route associated to the Railway Station in which the value of the route field "length" is the lowest

Person	CQ #	Question	Answer
Claudio	1.15	Give the schedule of the closest bus stop	In a range of $5km$ from all the Routes with arrival point the location of a Bus stop and starting point the current location, return the associated schedule of the stop considering the route with the minimum value of "length"
Claudio	1.16	Give the availability of the pubic transportation for a specific route	Return the several values of the attribute "Type" of the ModeOfTransport for a specific route
Andrea	2.1	Give the fastest path from Trento to Mezzolombardo	Search for the path with the lowest value of the field "duration" from Trento Square to Mezzolombardo Square and return route, mode of transport, time spent, cost and all the associated facilities
Andrea	2.2	List all the possible paths from Trento to Cles	Return route, mode of transport, time spent, price and associated facilities of all the possible paths from Trento to Cles
Andrea	2.3	Give all the fastest train itinerary from Trento to Mezzolombrardo	Return the path from Trento Square to Mezzolombardo Square with associated route, price and Railway Stations which has the field type of the mode of transport as "train" and the lowest value of the "time spent"
Andrea	2.4	List all possible paths from Trento to Pergine Valsugana	Return route,mode,price,time spent and all associated facilities of all possible paths from Trento to Pergine Valsugana
Andrea	2.5	List all the hiking paths in Trentino	Return route, time spent, starting point, final point and associated facilities of all hiking paths in Trentino
Andrea	2.6	Give longest hiking path in Trentino	From all the hiking paths in Trentino, return route, time spent, initial position, final position and facilities associated of the path with the maximum value of the route field "length"
Andrea	2.7	List all the available parking areas close to a specific point	Return contact, locations, addresses and price of the parking areas which has the value of the filed "length" of the associated route $<$ 5km from the specific point

Person	CQ #	Question	Answer
Andrea	2.8	List all the available fuel station between Trento and Tione	From the list of all the paths starting in Trento Sqaure and arriving in Tione Square return all the available facilities of types "Fuel Station" with contact, locations, address, schedule and price
Andrea	2.9	List all the buses itinerary from Trento to Mezzolombardo	Return route, cost, time spent and facilities associated of all the paths starting in Trento Square and arriving in Mezzolombardo which have the field type of the mode of transport as "Bus"
Andrea	2.10	Give the closest parking area to a specific point	Return contact, location, address, route, time spent, price of the path starting in the specific point and arriving in the parking area with the lowest value of the filed "length" of the route
Andrea	2.11	Give the closest fuel station to a specific point	Find the facilities of which the value of the field type is "Fuel Station" and return contact, location, address and price of the one with the lowest value of the field "length" of the route
Andrea	2.12	Give the length of a specific hiking path	Return the value of the field "length" of the specific path
Andrea	2.13	Give the location of the cheapest fuel station close to Trento	From all the fuel station with the value of the field route $<$ 5km from the Trento Square return the location and the address of the one with the lowest value of the field price
Andrea	2.14	Give the cost of an autonomous transportation from Trento to Mezzolombardo	Return the attribute "Price" of the ModeOfTransport
Andrea	2.15	Give the closest bus Stop from a specific location with the least waiting time	From all the routes with starting point the specific location and arrival point the location of the bus stop return the one with associated route with the minimum value of "Length" and least waiting time
Elisa	3.1	List all the possible paths from Rovereto to Trento	Return route, time spent, prices and mode of transportation of all the possible paths starting from Rovereto Square and arriving to Trento Square

Person	CQ #	Question	Answer
Elisa	3.2	List all the bus itineraries from Rovereto to Trento	Return route, time spent, costs, schedule and bus stops of all the paths with the value of the field type of the associated mode of transportation equal to "bus" from the Rovereto Square to the Trento Square
Elisa	3.3	List all the e-bikes facilities location in Trento	Return location and address of all the e-bikes facilities in Trento
Elisa	3.4	Give the availability of all the e-bikes facilities in Trento	Return location and address of all the available e-bikes facilities
Elisa	3.5	List all the cycling itineraries in Trento	Return route, time spent, starting position, arrival position of all the cycling routes with the distance from the Trento Square < 5km
Elisa	3.6	Give the cost of the cheapest path from Rovereto to Trento	Return the value "price" of the path with the lowest value of the associated price with starting position in Rovereto Sqaure and arrival position in Trento Square
Elisa	3.7	Give the cheapest train itinerary from Trento to Rovereto	Return route, time spent, price and all the facilities associated of the path starting in Roverto Square and arriving in Trento Square with lowest value of the field "price" and the value of the mode of transportation field type equal to "train"
Elisa	3.8	Give the longest cycling route of Trento	Return route, time spent, starting position, arrival position of the cycling routes with the distance from the Trento Square < 5km and the maximum value of the field "length" of the route
Elisa	3.9	Give the closest e-bike location to a specific point	Return location and address of the e-bikes facility with the lowest value of the field "length" of the route between the specific point and the facility location
Elisa	3.10	List all the campsites close to Trento	Return contacts, location, address, and price of all the campsites with the value of the field route < 5km starting from Trento Square
Elisa	3.11	Give the closest campsite to a specific point	Return contact, locations, addresses and price of the campsite which the lowest value of the filed "length" of the associated route starting from the specific point
Elisa	3.12	List the cost of all the campsites close to Trento	Return the value of the field "price" of all the campsites with the value of the field route < 5km starting from Trento Square

Person	CQ #	Question	Answer
Elisa	3.13	Give the cost of the cheapest campsite close to Trento	Return the value of the field "price" of the campsites with the value of the field route < 5km starting from Trento Square and the lowest value of the field "price"
Elisa	3.14	Give the most economic path from Rovereto to the Sociology Department in Trento	From all the paths starting from the Rovereto Square and arriving to the Sociology Department return route, time spent, price, mode of transport and associated facilities of the one with the lowest value of the field "price"
Elisa	3.15	Give the available campsites on the weekend in Trentino	Return facilities of type: "Campisite" which are available during the weekend
Elisa	3.16	Give the dates of unavailability of a specific bus stop	Return the list of the dates in which the schedule indicates that the specific bus stop is unavailable
Elisa	3.17	Give the next arrival time of a specific bus stop	Return the schedule of a specific bus stop
Maria	4.1	List all the possible paths from Bolzano to Trento	Return route, time spent, price, mode of transport and all the associated facilities of all the paths from Bolzano Square to Trento Square
Maria	4.2	List all the fuel station between Bolzano and Trento	From all the path starting in Bolzano Square and arriving in Trento Square return contact, location, address, price of all the facilities with the value of the field type equal to "Fuel Station"
Maria	4.3	Give the cheapest fuel stations in Bolzano	From all the facilities with the route value < 5km from the Bolzano Square search the one with value of the field type equal to "Fuel Station" and lowest value of the filed "price", return contact, location, address and price of it
Maria	4.4	List all the train itineraries from Trento to Bolzano	From all the paths starting in Trento Square and arriving in Bolzano Square return route, time spent, price and facilities of those with the field type of the mode of transport equal to "train"
Maria	4.5	List all the paths costs from Bolzano to Trento	For each path starting in Bolzano Square and arriving in Trento Square return the value of the field "price"

Person	CQ #	Question	Answer
Maria	4.6	Give the closest fuel station to a specific point	Return contact, location, address, price of the facility with type "Fuel Station" which has the lowest value "Length" of the associated route
Maria	4.7	Give the closest available "car sharing" position to a specific point	Return location and address of the car-sharing facility with availability "1" and the lowest value of the field "length" of the route between the specific point and the facility location
Maria	4.8	Give the fastest path from Bolzano to Trento by using public transports	From the list of paths starting in Bolzano Square and arriving in Trento Square with the field Public of the mode of transport equal to "1" return the one with the lowest value of "time spent"
Maria	4.9	Give the cost of the fastest public transportation path from Bolzano to Trento	From all the paths starting in Bolzano Square and arriving in Trento Square with value of the field Public of the mode of transport equal to "1" search the one with lowest value of the field "time spent" and return its associated "price"
Maria	4.10	List all the possible paths from Trento to Molveno by using the car	Return route, time spent, price and facilities mode of all the possible path from Trento to Molveno with the field Public of mode of transport equal to "0"
Maria	4.11	Give the cable car schedule from Trento to Sardagna	From all the path from Trento Square to Sardagna Square find those with the field type of the mode of transport equal to "cable car" and return the schedule
Maria	4.12	Give the fastest path from Trento to Molveno	From the list of paths starting in Trento Square and arriving in Molveno Square return the one with the lowest value of "time spent"
Maria	4.13	Give the cheapest path from Trento to Molveno	From the list of paths starting in Trento Square and arriving in Molveno Square return the one with the lowest value of "price"
Maria	4.14	Give the schedule of a specific bus stop	Return the schedule of a specific bus stop
Maria	4.15	Give the schedule of the fuel stations of a specific path	From the facilities associated to the specific path return the schedule of all the ones with the field "Type" equal to FuelStation

2.2.1.1 Inception Schema

In this paragraph we are going to schematize the several queries. The next tables has 3 properties: the number of the query, the the associated type which is going to be returned to answer to the question and last the properties defining the type.

CQ #	Type	Properties
1:1-4-11 **2**:1-2-4-5-6-9 **3**:3-4-7-9-11-14 **4**:1-4-7-10-11	Generic (Facility)	Type, Availability
1:1 **2**:8-11-13 **3** **4**:2-3-6-15	Fuel Station (Facility)	Price
1:1 **2**:7-10 **3** **4**:	Parking Area (Facility)	Price
1:1-3-14 **2**:3 **3** **4**:	Railway Station (Facility)	
1:1-7-14 **2**:15 **3**:7 **4**:	Buses Station (Facility)	
3:10-11-15	Campsites (Facility)	
1:1-3-5-7-8-11-13 **2**:1-2-3-4-5-6-7-8-9-12-15 **3**:1-2-5-7-8-10-11-12-14 **4**:1-4-7-8-10-12-13	Route	Length, Speed limits
1:1-5-7-11-16 **2**:1-2-3-4-14 **3**:1-8-14 **4**:1-5-13	Mode of transport	Type
1:1-3-5-6-7-9-11 **2**:1-2-3-4-7-8-9-10-11-13-14 **3**:1-2-5-7-6-9-10-12-13-14 **4**:1-3-4-5-6-8-9-10-13	Price	Price
1:1-4-5-10-12-13-14-15 **2**:5-6-7-8-10-11-12-13 **3**:3-4-5-7-8-9-10-11**4**:2-3-6-7-10-12-14	Location	Latitude, Longitude, Altitude
1:1-2-3-5-7-8-11 **2**:1-2-3-4-5-6-7-8-9-10-11 **3**:1-2-5-7-8-10-11-14 **4**:1-4-7-8-9-10-12-13	Time spent	Duration
1:7-10-12-13-14-15 **2**:15 **3**:16-17 **4**:11-14	Stop	Schedule
2:7-8-10-11-13 **3**:3-4-5-7-10-11 **4**:2-3-6-7-10	Contact	Website, Phone, Email address
1:1-4-5-10-12-13-14 **2**:7-8-10-11-13 **3**:3-4-5-7-9-10-11 **4**:2-3-6-7-10-12	Address	Province, City, Village, Street, Number, CAP

2.2.2 Initial Datasets description

In this section are reported the metadata at datasets level involved in the Inception phase, so those metadata regarding the sources, the authors, the collection methods, and other meaningful information. In this step we have identified twelve datasets which are useful to answer to the several quires.

- **Trentino public transport Urban TTE**: this dataset contains the *core data* about the city of Trento;

- **Trentino public transport Extra-Urban TTE**: this dataset contains the *core data* of routes in the Trentino province;
- **Trentino public transport rates for Urban TTE**: contains the *core data* of prices of routes in Trento;
- **Trentino public transport rates for Extra-Urban TTE**: contains the *core data* of prices of routes in Trentino Province;
- **Cycle paths**: contains the *common data* of cycle paths routes;
- **Cycling points of interest**: contains the *common data* of cycle point of interest;
- **Italian Parking Areas**: contains the *contextual data* of the parking places;
- **Railway Stations**: contains the *common data* of the railway's stations;
- **Car sharing (Open Data)**: contains the *common data* of the Car sharing information.;
- **Map of petrol stations in Italy**: contains the *contextual data* of the gas stations around Italy;
- **Mountain Paths**: contains the *common data* of the mountain paths;
- **Campsite and other accommodation facilities**: contains the *contextual data* of the campsites and their relative prices.

1. Trentino public transport Urban TTE.
   This dataset is beased on the GTFS standard and have eleven different attributes (agency, calendar, calendar_dates, feed_info, routes, shapes, stop_times, stops, stoplevel, transfers, trips). Each attribute is in separate txt file. General description of this dataset is in table below:

Info	**Description**
Dataset Identifier	p_TN: d3c9f167-3271-4a43-b5c1-e0879aa5ad3f
Dataset Publisher	Name: Public Transport Service, IPAVAT Code: 0OK0PZ
Date of modification	2017-10-24
Geographic coverage	Trento
URI of GeoNames	https://www.geonames.org/3165241
Languages of the dataset	Italian
Holder	Autonomous Province of Trento
Refresh Rate	Half yearly
Author	Name: Public Transport Service, IPA VAT: 0OK0PZ
Url	https://www.trentinotrasporti.it/opendata/google_transit_urbano_tte.zip
License	Creative Commons Attribution 4.0 International (CC BY 4.0)
License_Type	https://w3id.org/italia/controlled-vocabulary/licences/A21_CCBY40
Format	txt

2. Trentino public transport Extra-Urban TTE.

This dataset is beased on the GTFS standard and have eleven different attributes (agency, calendar, calendar_dates, feed_info, routes, shapes, stop_times, stops, stoplevel, transfers, trips). Each attribute is in separate txt file. General description of this dataset is in table below:

Info	**Description**
Dataset Identifier	p_TN: d3c9f167-3271-4a43-b5c1-e0879aa5ad3f
Dataset Publisher	Name: Public Transport Service, IPA / VAT Code: 0OK0PZ
Date of modification	2017-10-24
Geographic coverage	Trento
URI of GeoNames	https://www.geonames.org/3165241
Languages of the dataset	Italian
Holder	Autonomous Province of Trento
Refresh Rate	Half yearly
Author	Name: Public Transport Service, IPA / VAT: 0OK0PZ
Url	https://www.trentinotrasporti.it/opendata/google_transit_extraurbano_tte.zip
License	Creative Commons Attribution 4.0 International (CC BY 4.0)
License_Type	https://w3id.org/italia/controlled-vocabulary/licences/A21_CCBY40
Format	txt

3. Trentino public transport rates for Urban TTE.
This dataset is beased on the GTFS standard and have seven different attributes (fare_attributes_urbano.txt, fare_attributes_urbano_cartascalare, fare_attributes_urbano_mobile, fare_rules_urbano, fare_rules_urbano_cartascalare, fare_rules_urbano_mobile, zones_urbano). Each attribute is in separate txt file. General description of this dataset is in table below:

Info	**Description**
Dataset identifier	p_TN: d3c9f167-3271-4a43-b5c1-e0879aa5ad3f
Dataset Publisher	Name: Public Transport Service, IPA / VAT Code: 0OK0PZ
Date of modification	2019-05-09
Geographic coverage	Trento
URI of GeoNames	https://www.geonames.org/3165241
Languages of the dataset	Italian
Holder	Autonomous Province of Trento
Refresh rate	Half yearly
Author	Name: Public Transport Service, IPA / VAT: 0OK0PZ
Url	https://dati.trentino.it/dataset/6d5c2000-972e-4c21-aef6-fdbba94418a8/resource/44efc0bd-223a-49c7-b3b0-16128e32813c/download/tariffegtfsurbano.zip
License	Creative Commons Attribution 4.0 International (CC BY 4.0)

Info	Description
License_Type	https://w3id.org/italia/controlled-vocabulary/licences/A21_CCBY40
Format	txt

4. Trentino public transport rates for Extra-Urban TTE.
   This dataset is beased on the GTFS standard and have seven different attributes (fare_attributes_urbano.txt, fare_attributes_urbano_cartascalare, fare_attributes_urbano_mobile, fare_rules_urbano, fare_rules_urbano_cartascalare, fare_rules_urbano_mobile, zones_urbano). Each attribute is in separate txt file. General description of this dataset is in table below:

Info	Description
Identifier	p_TN: d3c9f167-3271-4a43-b5c1-e0879aa5ad3f
Publisher	Name: Public Transport Service, IPA / VAT Code: 0OK0PZ
Date of modification	2019-05-09
Geographic coverage	Trento
URI of GeoNames	https://www.geonames.org/3165241
Languages of the dataset	Italian
Holder	Autonomous Province of Trento
Refresh rate	Half yearly
Author	Name: Public Transport Service, IPA / VAT: 0OK0PZ
Url	https://dati.trentino.it/dataset/6d5c2000-972e-4c21-aef6-fdbba94418a8/resource/10e93dd1-3463-4664-8c24-300a7403780a/download/tariffegtfsextraurbano.zip
License	Creative Commons Attribution 4.0 International (CC BY 4.0)
License_Type	https://w3id.org/italia/controlled-vocabulary/licences/A21_CCBY40
Format	txt

5. Piste ciclabili (Cycle paths).
   General description of this dataset is in table below:

Info	Description
Landing page	https://siat.provincia.tn.it/IDT/vector/public/p_tn_3ff5db13-8a3d-4bd8-8d6f-9b2fdf1aeb41.zip
Created	26.09.2008
Coordinates	[[[10.41, 46.6], [11.97, 46.6], [11.97, 45.6], [10.41, 45.6], [10.41, 46.6]]] Type: Polygon
Contact	mailto: serv.naturambiente@provincia.tn.it
Identifiers	p_TN:3ff5db13-8a3d-4bd8-8d6f-9b2fdf1aeb41_resource
Format	cpg,dbf,prj,shp,shx

6. Punti di interesse ciclabili (Cycling points of interest).
   Representation of the punctual elements present on the Trentino cycle paths: bicigrill (refreshment point, assistance and information), counters (instrumentation for detecting pedestrian and cycle paths), cippi km and fountains. General description of this dataset is in table below:

Info	**Description**
Landing page	https://siat.provincia.tn.it/IDT/vector/public/p_tn_0211d261-70d8-485e-9265-b1c27b1a84e1.zip
Created	30.04.2019
Coordinates	[[[10.41, 45.6], [10.41, 46.6], [11.97, 46.6], [11.97, 45.6], [10.41, 45.6]]] Tipo: Polygon
Contact	mailto: serv.naturambiente@provincia.tn.it
Identifiers	p_TN:0211d261-70d8-485e-9265-b1c27b1a84e1_resource
Format	cpg,dbf,prj,shp,shx

7. Parking map in Italy.
   Archive, which can be represented on the map, which contains the non-exhaustive list of over 21,000 car parks in Italy. The source of the data is OpenStreetMap.org which has been assigned with automated procedures the classification by municipality, province and region. Data updated on: 23 February 2016 The data was created by DatiOpen.it on 23 February 2016

Info	**Description**
Landing page	http://www.datiopen.it/it/opendata/Mappa_dei_parcheggi_in_Italia
Created	23.02.2016
Contact	mailto: info@datiopen.it
AUTHOR	OpenStreetMap http://www.datiopen.it/it/catalogo-opendata/openstreetmap-org
PUBLISHED BY	Open.it
Format	xlsx

8. Train stations (Open data).
   Station of the railway stations in the municipal area of Trento. It includes the Brenner railway, the Trento_Malè_Marilleva railway and the Valsugana railway. Data provided by Trentino Trasporti.

Info	**Description**
Landing page	https://www.comune.trento.it/Aree-tematiche/Cartografia/Download/Stazioni-treno
Created	01.01.2017
Contact	mailto: Servizio.innovazionedigitale@comune.trento.it
AUTHOR	Trentino Trasporti https://www.trentinotrasporti.it/
PUBLISHED BY	Trentino Trasporti
Format	gml,shp,kml,dxf

9. Car sharing (Open Data).
   Location of Car sharing stalls Parking spaces dedicated to the collection and delivery of Car sharing vehicles. Data taken directly from the site https://www.carsharing.tn.it Car sharing allows you to have a car suitable for family or business needs without owning one and without incurring fixed costs (road tax, insurance, maintenance, garage or parking), but paying only in proportion to use.

Info	Description
Landing page	https://dati.trentino.it/dataset/car-sharing-open-data
Created	04-05-2018
Contact	mailto: Servizio.innovazionedigitale@comune.trento.it
AUTHOR	Name: Municipality of Trento IPA / VAT: c_l378
PUBLISHED BY	Trentino Trasporti
License	http://creativecommons.org/publicdomain/zero/1.0/deed.it
Format	gml,shp,kml,dxf

10. Map of petrol stations in Italy.
    Archive, which can be represented on the map, which contains the non-exhaustive list of over 13,000 petrol stations in Italy. The source of the data is OpenStreetMap.org which has been assigned with automated procedures the classification by municipality, province and region.

Info	Description
Landing page	http://www.datiopen.it/it/opendata/Mappa_dei_distributori_di_carburante_in_italia
Created	23-02-2016
Contact	mailto: Servizio.innovazionedigitale@comune.trento.it
AUTHOR	OpenStreetMap
PUBLISHED BY	DatiOpen.it
License	http://opendatacommons.org/licenses/odbl/
Format	xlsx

11. Province of Trento Paths.
    Paths of the entire network of the Società degli Alpinisti Tridentini (SAT) that insist on the territory of the Autonomous Province of Trento: each path consists of the spatial coordinates that allow it to be correctly positioned on the territory and presents a series of additional textual information (attributes) that describe.

Info	Description
Landing page	http://www.datiopen.it/it/opendata/Provincia_di_Trento_Sentieri
Created	15-12-2015
Contact	mailto: sentieri@sat.tn.it
AUTHOR	SAT (Society of Tridentine Alpinists)

Info	Description
PUBLISHED BY	DatiOpen.it
License	http://opendatacommons.org/licenses/odbl/
Format	shp,json,xml,csv

12. Autonomous Province of Trento List of non-hotel structures.
    The archive contains information on non-hotel accommodation facilities in the territory of the Autonomous Province of Trento: rural businesses, bed-and-breakfasts, campsites, hostels, holiday homes, etc. Where available, the data contains the address, telephone, e-mail address, website and other information.

Info	Description
Landing page	http://www.datiopen.it/it/opendata/Provincia_Autonoma_di_Trento_Elenco_strutture_extra_alberghiere?metadati=showall
Created	25-10-2017
Contact	mailto: info@open.it
AUTHOR	Trentino Alto Adige Region
PUBLISHED BY	DatiOpen.it
License	http://opendatacommons.org/licenses/odbl/
Format	xls,csv,json

2.2.3 Datasets metadata documentation

Datasets following the GTFS format have a standard metadata definition[1].
Regarding the datasets containing the information for urban and extra urban:

- **agency.txt**: it contains information about the agency - Trentino Trasporti;
- **calendar.txt**: it contains information about the operational days of the week;
- **calendar_dates.txt**: exceptions to the pattern described into *calendar.txt*;
- **feed_info.txt**: information about the publisher;
- **routes.txt**: we will use the information contained into these files to obtain urban and extra-urban paths;
- **shapes.txt**: it contains the path that the bus travels along a route;
- **stop_times.txt**: we can extract all the information regarding travel time and time scheduling;
- **stops.txt**: into this file we can extract all the stops information like name and coordinates;
- **transfers.txt**: information about transfers between lines;
- **trips.txt**: information about service available a certain route.

[1] Google metadata description: https://developers.google.com/transit/gtfs/reference

Moreover, the datasets containing the prices will be integrated with the above information to provide a complete overview over a certain route; all the metadata descriptions about the Trentino Trasporti's datasets are found here. Further meta description and origin of the attributes are available at this github repository.

2.2.4 Datasets collection process

In this section we are going to explain the method used to collect the several datasets. In particular we have divided the process in 4 different steps as described in the iTelos Methodology. Entering into details:

2.2.4.1 Iteration Zero In the first step we verified that datasources were compliant with the information we needed for the competency queries in *common data* typology.
Indeed, we realized that we were missing *the campsite and accomodation dataset*; we then proceed to add this dataset to the datasources sheet.

2.2.4.2 Iteration One In the second step we downloaded the following datasets in *common data* typology: Cycle paths, Cycling points of interest, Railway Stations and Mountain Paths. The kind of information we cared the most were the coordinates of the cycling points.
We then added a metadata description of both the dataset and the attributes to have a more descriptive information.
A lot of **CQs** require locations points that are found in these datasets, so that the reason why we decide to categorized them as *common data*.

2.2.4.3 Iteration Two In the third iteration we extracted all the data for the *core data* typology. The datasets collected by the datasources were using the GFTS standard, so that metadata description were already available together with the Trentino Trasporti dataset.
In this iteration the following datasets were extracted and categorized as *core data* according to the **CQs**: Trentino public transport Urban TTE, Trentino public transport Extra-Urban TTE, Trentino public transport rates for Urban TTE, Trentino public transport rates for Extra-Urban TTE.

2.2.4.4 Iteration Three In the last iteration we collected the *contextual data* typology by extracting the following datasets: Italian Parking Areas, Car Sharing, Italian Gas Stations, Campsite and other accommodation facilities.
As long as some prices were missing in some *campsite and accomodations* dataset, we elaborated the data price attribute by calculating **mean** and **variance** over the available information on a specific category and assumed that they the fit a **normal distribution**. We then randomly fill the missing data accordingly to obtain cleaned dataset.

2.2.5 Inception level evaluation

The last section of the Inception phase report the evaluation of the outcomes obtained in this phase, through specif evaluation metrics.

2.3 Informal Modeling

The Informal Modeling step is divided in two main tasks: one at Schema Level and one at Data Level. Concerning the Schema Level the definition of the several ETypes that the KG is going to involve and the providing of the EER

Schema are aimed, while on the data level the goal is to highlight the evolution of the datasets (those filtered and those new) and provide all their metadata.

2.3.1 Schema level

The Schema level is divided in two paragraphs in order to highlight in a better way the EER Schema and the evolution from the CQs defined in the Inception step.

2.3.1.1 ETypes and EER Model definition

The first step to design the EER Schema is to define the ETypes and identify which are Core ETypes, Common ETypes or Contextual ETypes and then defined their attributes and the relationship among them.
The Core ETypes are the sine qua non entities defining our project. In the next lists we are going to highlight those ETypes considered as the most important for the Transportation Domain:

1. Path: it is the all-encompassing entity and it represents the set of all the routes, mode of transport, prices and related facilities of the taken trip.
    - Route: it is an array of the entity Route.
2. Route: it is the entity linked to all the other entities and involving all those attributes to which the user can be interested.
    - RouteID: it is the identifier of the route.
    - SpeedLimit: it is the value of the speed limit of the route.
    - StartingPoint: it is an attribute defined by the entity Address and it is the starting point of the route.
    - ArrivalPoint: it is an attribute defined by the entity Address and it is the arrival point of the route.
    - ModeOfTransport: it is an attribute defined by the entity ModeOfTransport and indicates the mode of transport used to go through the route.
    - TimeSpent: it is an attribute defined by the entity TimeSpent and indicates the amount of time spent to go through the route.
    - Facility: it is an array of the entity Facility and it represents all the facilities encountered going through the route.
    - Length: it is the length of the route (i.e. the distance from the arrival point and the starting point).
3. ModeOfTransport: it is the entity describing the set of the mode of all the mode of transport involved.
    - Type: it is an attribute defined by the entity TransportEnum and indicates the mode of transport used to go through the associated route.
4. PublicTransport: it is a direct child of the ModeOfTransport entity.
    - Agency: it is an attribute defined by the entity Agency and represents all the specifics of the Agency providing the service to go through the associated route.
    - Ticket: it is an attribute defined by the entity Ticket and indicated the price of the fare to use the transportation service.

- StopTime: it is an array of the entity StopTimes and it indicates the schedule of the public mode of tranport.

5. AutonomousTransport: it is a direct child of the ModeOfTransport entity.
   - FuelType: it is an attribute defined by the entity FuelTypeEnum and it provides the type of fuel used by the autonomous vehicle to go through the associated route.
   - Price: it is an attribute defined by the entity Price and it provides the cost of the mode of transport used.
6. Ticket (GTFS Format): it is the entity which describes all those attributes concerning the cost of the tickets.
   - FareID: it is the identifier of the fare.
   - Price: it is an attribute defined by the entity Price and it provides the cost of the mode of transport used.
   - PaymentMethod: it indicates the payment method as defined by the GTFS Format.
   - TransferDuration: it indicates the duration of the validity of the ticket.
7. StopTimes (GTFS Format): it is the entity specifying the time schedule of the stops made by the mode of transport selected.
   - ArrivalTime: it is an attribute defined by the entity TimeStamp and it represents the arrival time at a certain stop.
   - DepartureTime: it is an attribute defined by the entity TimeStamp and it represents the departure time from a certain stop.
   - StopID: it is the identifier of a stop.
   - StopSequence: it indicates the number of the stop during the trip.
   - Calendar: it is an attribute defined by the entity Calendar and it highlights the days in which the service is available for a certain stop.
   - Stop: it is an attribute defined by the entity Stop and it makes reference to all those information about the Stop (shown below).
8. Stop (GTFS Format): it is the entity involving all those attribute needed to characterize a stop of the public mode of transport.
   - StopID: it is the identifier of a stop.
   - StopCode: it is the code identifying a stop.
   - StopName: it is the name of the stop.
   - Location: it is an attribute defined by the entity Location and it represents the location of the stop in the world.
   - WheelchairBoarding: it indicates the presence of the wheelchair boarding along the stop.
9. TimeSpent: it is the entity indicating the time spent to go through a route.

- TimeSpent: it is an attribute defined by the entity TimeStamp and it indicates the time spent to go through a route.

The Common Etypes are not necessary to design the KG but they are needed to improve the quality of the description of the whole environment in which the system is going to be established. In case of the Transportation Domain they characterize for example the prices, the facilities or the scheduling of the bus stop:

1. Facility: it is the EType defining all the facilities associated to a route, so those which can be encountered going through the route.
   - Type: it is an attribute defined by the entity FacilityEnum and it characterizes the nature of the facility.
   - Price: it is an array of the entity Price and it contains the prices of the service guarantee by the facility.
   - Contact: it is an attribute defined by the entity Contact and it indicates the several ways to get in touch with the facility.
   - Ranking: it indicates the service evaluation of the facility.
   - Address: it is an attribute defined by the entity Address and it indicates the address and the information about the location of the facility.
   - Calendar: it is an attribute defined by the entity Calendar and it indicates the time scheduling and the availability of the facility.
2. Agency (GTFS Format): it is the entity describing those information about the agency providing the service of public transportation.
   - AgencyID: it is the identifier of the agency.
   - AgencyName: it is the name of the agency.
   - AgencyTimezone: it is the timezone of the agency.
   - AgencyLang: it is the language used by the agency.
   - Contact: it is an attribute defined by the entity Contact and it indicates the several ways to get in touch with the agency.
3. Price: it is the entity describing the costs of a ticket more than the costs of all the services provided by a facility or even the cost of an autonomous trip.
   - Cost: it is the cost.
   - CurrencyType: it is the currency of the payment.
4. Address: it is the entity describing all those information about the location of a facility more than an arrival or starting point of a route .
   - Province: it represents the Province of interest.
   - City: it represents the City of interest.
   - Street: it represents the Street of interest.
   - Number: it represents the Number of interest.

- CAP: it is the CAP of the City of interest.
- Location: it is an attribute defined by the entity Location and it represents the location in the world.

5. Calendar (GTFS Format): it represents the time scheduling of a public transport stop or of a facility.
    - ServiceID: it is the service identifier.
    - Monday: it indicates if the service is available on Monday.
    - Tuesday: it indicates if the service is available on Tuesday.
    - Wednesday: it indicates if the service is available on Wednesday.
    - Thursday: it indicates if the service is available on Thursday.
    - Friday: it indicates if the service is available on Friday.
    - Saturday: it indicates if the service is available on Saturday.
    - Sunday: it indicates if the service is available on Sunday.
    - StartDate: it indicates the service starting date.
    - EndDate: it indicates the service ending date.
    - Exceptions: it is an array of the entity CalendarDates and it represents those dates in which the service is interrupted.
6. CalendarDates (GTFS Format):
    - ServiceID: it is the service identifier.
    - Date: it indicates the date in which the service is interrupted.
    - ExceptionType: it indicates the reason of the interruption accordingly to the GTFS Standard.

The Contextual ETypes are the less important ones, they are not used to characterized the system designed but they are needed to characterize in a better way the Common and the Core ETypes, usually they are structures highlighting the several options which can be selected by an enumeration or again to put together several easier types in order to characterized a more complex attribute.

1. FuelTypeEnum: it is an entity indicating the several fuel types which can be used in an autonomous trip.
    - NoFuel: it indicates when the trip can be taken without using any fuel (in our case when it is possible to go through a route on foot or by bike).
    - Diesel: it indicates when the trip is taken by using a vehicle with diesel engine.
    - Petrol: it indicates when the trip is taken by using a vehicle with petrol engine.
    - Gas: it indicates when the trip is taken by using a vehicle with gas engine.
    - Methane: it indicates when the trip is taken by using a vehicle with methane engine.
    - Electric: it indicates when the trip is taken by using a vehicle with electric engine.
2. TransportEnum: it is the entity indicating the several modes of transport which can be used to move through a route.
    - Train: it indicates the train as mode of transport.

- Bus: it indicates the bus as mode of transport.
- Car: it indicates the car as mode of transport.
- CableCar: it indicates the cable car as mode of transport.
- Bike: it indicates the bike as mode of transport.
- Foot: it indicates no mode of transport.

3. FacilityEnum: it is the entity indicating the several facilities involved by our system, in particular those considered are all the facilities related to the road system.
   - FuelStation: it indicates a facility which provides fuel pumps.
   - ParkingAreas: it indicates a parking area.
   - RailwayStation: it indicates a railway station.
   - BusStation: it indicates a bus station.
   - CarSharingPark: it indicates a parking area which is provided by the car sharing service.
   - BikeSharingPark: it indicates a parking area for bikes which is provided by the bike sharing service.
   - Campsite: it indicates an area which provides the possibility of camping or parking campers or caravans.
4. Contact: it is the entity describing all the ways to get in touch with a facility more than a transportation agency.
   - Phone: it indicates the phone number.
   - Email: it indicates the email.
   - Website: it indicates the url of the website.
5. Location: it is the entity describing the world location of an address, more than the location of a stop.
   - Latitude: it indicates the latitude coordinate.
   - Longitude: it indicates the longitude coordinate.
   - Altitude: it indicates the altitude coordinate.
6. TimeStamp: it is the entity describing the time information format for the time spent going through a route more than the arrival time or departure time from a stop.
   - Hour: it indicates the hours.
   - Minutes: it indicates the minutes.
   - Seconds: it indicates the seconds.

Starting from the definition of all the ETypes we have designed the EER Schema. To highlight the difference among the Core, the Common and the Contextual ETypes they have been represented by using different colours:

- Core ETypes: blue.
- Common ETypes: yellow.
- Contextual ETypes: red.

Moreover several arrows have been used to highlight the links among the ETypes, in particular the head of the arrows have been changed according to the nature of the relationship between two ETypes:

- black dot cursor: it represents the structural attribute.
- full triangle cursor: it represents an inheritance relationship.
- simple arrow cursor: it represents other relationship types, furthermore a label on each arrow has been added so that it is possible to identify the nature of the relation itself.

In order to show in a better way the schema, it has been plotted hierarchically, the starting point is the EType Path that includes the most involved EType Route which in turn is connected to almost all the Core ETypes.
In the schema it is possible to recognize three different branches: the one related to the modes of transport domain, considering their costs and their "high level" specifications as the agency, the fuel type and the mode of transport type; the one related to the space domain, which involves the several specifications about the facilities like the address, the type of the facility, the contact, the location; the one related to the time domain, which includes the stop times of the public transportation, the time spent on each route, the calendar and the calendar dates ETypes.
The Informal Modeling Schema is shown in the figure 1.

2.3.1.2 Variance respect CQs definition

Between the Inception Schema and the Informal Modeling Schema several changes have been made, this in order to satisfy two main milestone of the project:

- Exploiting the already existing standards, our main one is the GTFS Format;
- Being able to create a modular system in order to adapt its functionality to any situation.

In order to describe in a clearer way the several changes we are going to: list all the entities added specifying when they belong to a particular format; list all the changes related to each already existing entity highlighting when necessary the attributes introduced.
Starting from the new entities:

- Ticket (GTFS Format): the ModeOfTransport entity is no longer directly attached to the Price entity but it goes through the Ticket entity, here are described all the attributes related to the fares.
- StopTimes (GTFS Format): thanks to this entity it is possible to totally describe the daily scheduling of each stop, including the arrival time and the departure time and the calendar.
- Agency (GTFS Format): this entity has been introduced in order to describe in a better way the service provided.
- directly attached to the Calendar entity the CalendarDates entity is used to identify those particular days in which a service is unavailable.
- FuelTypeEnum: this contextual entity is used to list all the possible fuel types, it has been introduce to give a better qualification of the ModeOfTransport entity.
- FacilityEnum: this entity is involved in an important change. While in the Inception Schema the several types of facilities where treated with an inheritance relationship from the parent entity "Facility", now thanks to this entity it is possible to enclose all of them in just one.

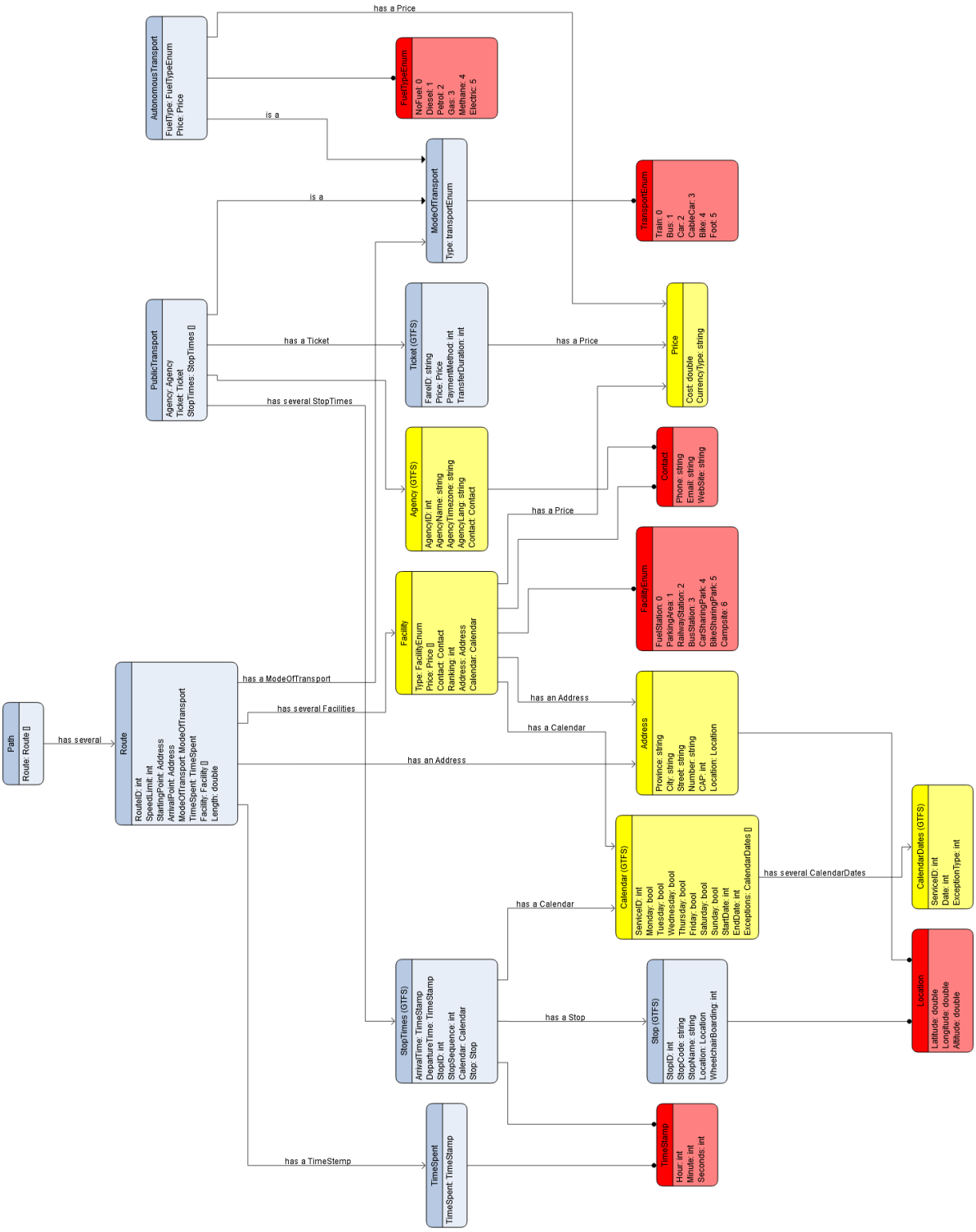


Figure 1: Informal Modeling Schema

While the changed entities are:

- Path: is no longer the most important entity which involved all the entities hierarchically below, but even if it maintains its role of top level entity now it is directly attached just to the Route entity.
- Route: this entity has now become the center of the whole schema, in particular each Route element involves a ModeOfTransport, a TimeSpent, a list of Facility and two Address.
- ModeOfTransport: this entity is the other one playing one of the most important role. While in the Inception Schema it was treated as a single entity now it has been split into two children entity: "PublicTransport" and "AutonomousTransport", in this way it is possible to highlight the difference between the ways of transport adoptable.
- Stop (GTFS Format): this entity already existing has been modified in order to follow the GTFS standard, so several attributes have been introduced qualifying in a better way the Stop entity.
- Facility: as it has already been shown above, the Facility entity has been subject to change so several attributes have been includes, the most important is the "Ranking" one in which it is shown the quality of the service provided.
- Price: to the entity Price it has been introduced the attribute CurrencyType in order to respect the GTFS format and to give an improvement of the description.

2.3.2 Data level

From the *Data Level* point of view we decide to export the **filtered** data into **json** format. The filtering scripts were developed based on the data that we needed from the dataset to model our knowledge. We use python along with different libraries such as pandas to read the csv datasets and filter and rename the different attributes. Moreover, we provided together with the exported json datasets their metadata description in **DCAT** specification. Encoded in both **RDF/Turtle** and **JSON-LD**. The metadata description of the entity's attribute of a certain dataset is found in the previous phase section.

2.3.2.1 Datasets management process

The first thing that we have done was to check if more data were needed to fulfill the *CQs* requirements. We found and added a new dataset with information about **e-motion** bikesharing service in the Trentino province.
Regarding the data filtering, at the beginning we evaluated the usage of existing libraries to generate the **DCAT** specification and the filtering altogether. We abandoned this path because we found out it was time consuming and checking the **Socrata** library we saw that the online catalog that we should be exploited to *publish* our elaborated dataset was deprecated since July 2020. We then decided a manual approach with python scripts and mostly the *pandas* library to read the csv dataset, filter the needed columns, checking the informal Modeling Schema and exporting the datasets in the json format.
We decided to maintain the datasets in the *GTFS* format as they were because we were concerned that converting them in a single json could slow us down in the next phase.

2.3.2.2 Datasets metadata documentation

As described in the previous section, we generated the **dcat metadata** description of our datasets, both in *RDF/TURTLE* and *JSON-LD* file. We also renamed some columns to make them more explicit, even if they were already described in the metadata in the Inception phase.
Moreover, we have tried to install a Socrata catalog without success, but we were able to manage to install a **CKAN Catalog** on a server, where we are planning to share the datasets. The server is hosted in digitalocean and it is available at this url `http://167.99.139.12/` for the moment. All the metadata files are available in the respective metadata folder of the Informal Schema folder of the github repository.

2.3.2.3 Variance respect Inception datasets

In the **accomodation** dataset we extracted only: *address information*, *typology* and *price information*, renaming the field "Altitudine località turistica" to simply "Altitude"; this dataset contains also synthetic information about prices estimated from available data in the Inception phase.
We keep all the information in the added **e-motion bikesharing** dataset, but we merged all the files into one.
In the **car sharing** dataset we extracted: *nomepos*, *via* and *geometry*.
In the **cycling points** we filtered: *OBJECTID*, *DES_TIPO_E*, *DESCRIPTION* and *GEOMETRY*.
In the **cycling routes** script we extracted from the dataset: *OBJECTID* and *geometry*.
In the **fuel pumps maps** dataset we extracted: **coordinates information**, *address information* and *name*.
In the **mountain paths** dataset we extracted: *Path information*, *difficulty*, *start place* and *end place*.
In the **parking areas** dataset we extracted *coordinates information*, *address information* and *name*.
Finally in the **stazioni** dataset we extracted: *name* and *geometry*.
In the dataset pre-processing we also taking into account the evolution of the Informal Modeling Schema during the different iterations.

2.3.3 Informal Modeling Evaluation

The last section of the Informal Modeling phase report the evaluation of the outcomes obtained in this phase, through specif evaluation metrics.

2.4 Formal Modeling

This section is dedicated to the Formal Modeling phase description. The Section is divided in Schema and Data level in order to report the details regarding both the ontology generated and the datasets version in the current phase.

2.4.1 Schema level

The schema level section in the current phase, reports the detailed description of the ontology generation.

2.4.1.1 Ontology definition

This section reports in details how the ontology is generated stating from the informal schema of the previous phase, which tools are used to do that, as well as usage of external ontology resources adopted to obtain the final KG ontology. Moreover a list of metadata is reported in this section, in order to describe all the elements of the ontology defined.

2.4.1.2 Variance respect to the EER Model

Once the ontology has been built, this section report the differences, and so the variance, respect the EER model defined in the previous phase. This a way to define the quality of the outcomes for the current phase as well as the alignment of the overall project development process.

2.4.2 Data level

As in the previous phase the data level section here, reports the description of the new version of the datasets, after formatting operations.

2.4.2.1 Formal Modeling datasets management

In this section are reported the operations and the tools adopted to format the dataset collected, in order to align them to the ontology definitions generated at schema level.

2.4.2.2 Datasets metadata documentation

In this section eventually new metadata information are added in order to describe the evolution of the datasets.

2.4.2.3 Variance respect Informal Modeling datasets

This section aims to define the variance between the data elements (datasets and attributes within them) produced in this phase, and the initial datasets collected in the previous phase. This a way to define the quality of the outcomes for the current phase as well as the alignment of the overall project development process.

2.4.3 Formal Modeling Evaluation

The last section of the Formal Modeling phase report the evaluation of the outcomes obtained in this phase, through specif evaluation metrics.

2.5 Data integration

This section is dedicated to the Data Integration phase description.

2.5.1 Data integration operations and tool

This section is dedicated to the description of the usage of the data integration tool that allows to map the datasets generated and well formatted in the previous phases, with the final ontology generated. The last datasets adaptation performed using the tool, as well as the mapping operation are here detailed.

2.5.2 Variance respect Formal Modeling datasets

The last section of the data integration phase aims to describe the variance, analyzing the differences, between the datasets integrated with the ontology, in the data integration platform which contain the KG, and the datasets collected in the previous phase. This analysis can highlight the results of the operations performed during the final phase of the data integration process.