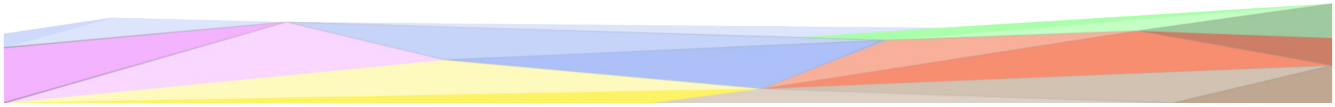
Dipartimento di Ingegneria e Scienza dell'Informazione

– 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
2.0	18.10.2020	Fivos Kapidis	Draft of the Scenario and of the Storytelling Definition
3.0	19.10.2020	Fivos Kapidis	Draft 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
3.1	20.10.2020	Fivos Kapidis, Antonio Stefani	Final revision of the CQs
6.0	20.10.2020	Alberto Carbognin	Draft of the metadata documentation
2.1	21.10.2020	Antonio Stefani	Final revision of the Scenario Description and of the Storytelling Definition
3.2	21.10.2020	Antonio Stefani	Final revision of the CQs
4.1	21.10.2020	Antonio Stefani	Final revision of the Inception Schema
6.1	21.10.2020	Alberto Carbognin	Final revision of the metadata documentation
5.1	21.10.2020	Omid Jadidi	Final revision of the dataset description
7.0	03.11.2020	Antonio Stefani	Draft Informal Modeling Schema
8.0	03.11.2020	Fivos Kapidis	Draft variance respect the defined CQs
9.0	04.11.2020	Antonio Stefani	Draft ETypes and attributes
10.0	04.11.2020	Omid Jadidi	Draft metadata documentation
11.0	04.11.2020	Alberto Carbognin	Draft datasets management process
7.1	05.11.2020	Antonio Stefani, Fivos Kapidis	Final revision of ETypes and attributes, Informal Modeling schema
10.1	05.11.2020	Omid Jadidi, Alberto Carbognin	Final revision datasets management process
11.1	05.11.2020	Omid Jadidi, Alberto Carbognin	Final revision metadata documentation
8.1	05.11.2020	Antonio Stefani, Fivos Kapidis	Final revision of the variance respect the defined CQs
12.0	18.11.2020	Antonio Stefani	Draft Ontology documentation
13.0	25.11.2020	Omid Jadid	Draft datasets management
14.0	25.11.2020	Alberto Carbognin	Draft variance respect previous datasets
12.1	26.11.2020	Antonio Stefani, Fivos Kapidis	Final revision Ontology documentation
13.1	26.11.2020	Omid Jadidi, Alberto Carbognin	Final revision datasets management
14.1	26.11.2020	Omid Jadidi, Alberto Carbognin	Final revision variance respect previous datasets
15.0	15.11.2020	Antonio Stefani	Revision of the whole documents
16.0	16.12.2020	Fivos Kapidis	Draft evaluation
17.0	16.12.2020	Antonio Stefani	Draft codebook
18.0	16.12.2020	Omid Jadidi	Draft data processing
19.0	16.12.2020	Alberto Carbognin	Draft karmalinker and process documentation
16.1	17.12.2020	Antonio Stefani	Final revision evaluation
17.1	17.12.2020	Antonio Stefani	Final Revision codebook, draft presentations
19.1	17.12.2020	Omid Jadidi, Alberto Carbognin	Final revision karmalinker and process documentation
18.1	17.12.2020	Omid Jadidi, Alberto Carbognin	Final revision data processing
20.0	18.12.2020	All	Final revision presentation and documents

1 Knowledge Graph Codebook

In the first chapter of this report we are going to resume all the activities done during the project. Each step will be then illustrated into details in the second chapter.

1.1 Knowledge Graph general description

Improving the quality of traveling is essential to live in an easier way: you always lose time moving from one place to another and you lose even much more time looking for a route or mode of transport to get somewhere as fast as possible and often spending as little as possible. Our project wants to solve this problem: it works under the transportation domain and it goals is to answer to those questions regarding the road system in Trentino, this include of course the railway system, the mountain paths and the cycling routes as well. So, more in general, we want to provide a system able to suggest to the users the fastest and/or (or whichever information they want to know) cheapest way to get to another place by respecting their will: using a specific public mode of transport or a private one.

1.2 Data level

In the final version of the datalevel datasets have been broken in many pieces to be as much as possible the same as schema. For instance one dataset might have many fields and attributes in it, so we devided these attributes in classes that have been described in the schema **(Address, Location, ...)** and defined a key for each connection between classes. In this sense hierarchy of the classes or etypes have been saved through key indexes. Later we will explane how we import these data in karma linker and define relation between nodes. After importing data in karmalinker and defining relations we extract the final data as rdf format and all have been saved in github directory.

1.2.1 Datasets general details

- **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
Geographic coverage	Trento
URI of GeoNames	https://www.geonames.org/3165241
Holder	Autonomous Province of Trento
Author	Name: Public Transport Service, IPA VAT: 0OK0PZ
Url	https://www.trentinotrasporti.it/opendata/google_transit_urbano_tte.zip

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
Geographic coverage	Trento
URI of GeoNames	https://www.geonames.org/3165241
Languages of the dataset	Italian
Holder	Autonomous Province of Trento
Author	Name: Public Transport Service, IPA / VAT: 0OK0PZ
Url	https://www.trentinotrasporti.it/opendata/google_transit_extraurbano_tte.zip

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
Geographic coverage	Trento
URI of GeoNames	https://www.geonames.org/3165241
Holder	Autonomous Province of Trento
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

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
Geographic coverage	Trento
URI of GeoNames	https://www.geonames.org/3165241
Holder	Autonomous Province of Trento
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

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

Info	Description
Identifiers	p_TN:3ff5db13-8a3d-4bd8-8d6f-9b2fdf1aeb41_resource
Url	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

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
Identifiers	p_TN:0211d261-70d8-485e-9265-b1c27b1a84e1_resource
Url	https://siat.provincia.tn.it/IDT/vector/public/p_tn_0211d261-70d8-485e-9265-b1c27b1a84e1.zip
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

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
Publisher	Open.it
Author	OpenStreetMap http://www.datiopen.it/it/catalogo-opendata/openstreetmap-org
Url	http://www.datiopen.it/it/opendata/Mappa_dei_parcheggi_in_Italia
Contact	mailto: info@datiopen.it

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
Publisher	Trentino Trasporti
Author	Trentino Trasporti https://www.trentinotrasporti.it/
Url	https://www.comune.trento.it/Aree-tematiche/Cartografia/Download/Stazioni-treno
Contact	mailto: Servizio.innovazionedigitale@comune.trento.it

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
Author	Name: Municipality of Trento IPA / VAT: c_l378
Published	Trentino Trasporti
Url	https://dati.trentino.it/dataset/car-sharing-open-data
Contact	mailto: Servizio.innovazionedigitale@comune.trento.it

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
Author	OpenStreetMap
Publisher	DatiOpen.it
Url	http://www.datiopen.it/it/opendata/Mappa_dei_distributori_di_carburante_in_italia
Contact	mailto: Servizio.innovazionedigitale@comune.trento.it

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
Author	SAT (Society of Tridentine Alpinists)
Published	DatiOpen.it
Url	http://www.datiopen.it/it/opendata/Provincia_di_Trento_Sentieri
Contact	mailto: sentieri@sat.tn.it

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
Author	Trentino Alto Adige Region
Published	DatiOpen.it
Url	http://www.datiopen.it/it/opendata/Provincia_Autonoma_di_Trento_Elenco_strutture_extra_alberghiere?metadati=showall
Contact	mailto: info@open.it

1.2.2 Datasets metadata documentation

All the metadeta at attribute level can be found in the json format in our Github repository.

1.3 Ontology level

In this section we are going to describe the final ontology defining our Knowledge Graph illustrating in particular all parties involved.

1.3.1 Ontology general details

As previously said in the introduction, our ontology aims to model the transportation domain. We have built an ontology as modular as possible taking into consideration not only the road system but also the railway system, the cycling routes and the mountain paths going to cover almost all the ways and modes of transport. This way of proceeding gave us the chance to not only providing an ontology working with different datasets from those of departure but also one which can be easily integrated into other ontologies describing different problems: for instance a Geo-Spatial ontology can integrate our one just by adding links to the class "Location" (i.e. the coordinates of a point).
Our work was done starting from a blank sheet of paper in order to develop a schema as modular and complete as possible. This leaded us to not consider any ontology already existing in our domain but just hiring some constructs of them. In particular one thing we focused on was the GTFS Format structure, indeed we used connection among the classes guided by the format itself (Calendar ⇒ CalendarDates).

1.3.2 Ontology metadata documentation

In this section all the metadata describing each element of the ontology are reported, in particular we illustrate firstly the ETypes used and the concept which they represent, then we show all the attributes of a circumstantial class added to define the type of certain elements (i.e. the class Enumeration comprising the several lists of the modes of transport, the facilities and the fuel types used by vehicles)

EType	GID	Concept	Data Properties	Object Properties
Address	36400	A sign in front of a house or business carrying the conventional form by which its location is described	has City (45988) has Number (34489) has Province (46567) has Street Address (45807) has Zip Code (34110)	has Location (132)
Agency	45084	An administrative unit of government	has Agency ID (120057) has Agency Language (120060) has Agency Name (120058) has Agency Timezone (120059)	has Contact (39136)

EType	GID	Concept	Data Properties	Object Properties
Calendar	44719	A tabular array of the days	has Start Date (120091) has End Date (120090) has Monday (80758) has Tuesday (80759) has Wednesday (80760) has Thursday (80761) has Friday (80762) has Saturday (80763) has Sunday (80757) has Service ID (120056)	has Exception (31741)
Calendar Dates	120045	Tabular array of dates which are associated to a specific events	has Date (103420) has Service ID (120056) has Exception (31741)	
Contact	120109	All useful information to get in touch with someone	has Phone (34485) has Website (34126) has Email (105296)	
Duration	80581	The property of enduring or continuing in time		has Time Stamp (120046)
Enumeration	34789	A numbered list		
Facility	3012	Something designed and created to serve a particular function and to afford a particular convenience or service		Is Specified By (120047) has Calendar (44719) has Facility Address (36400) has Facility Calendar (44719) has Facility Contact (39136) has Facility Price (28431)
Location	132	A point or extent in space	has Altitude (28272) has Latitude (46263) has Longitude (46270)	
Mode Of Transport	120044	Way in which transportation happens		Is Specified By (120048)
Path	46379	An established line of travel or access		has Road (22592)
Price	28431	Value measured by what must be given or done or undergone to obtain something	has Cost (70407) has Currency Type (120061)	

EType	GID	Concept	Data Properties	Object Properties
Private Transport	120043	Personal or individual use of transportation vehicle	has Cost (70407) has Currency Type (120061)	has Individual Price (28431) has Fuel Type (120049)
Public Transport	22138	Conveyance for passengers or mail or freight		has Agency (45084) has Stop Time (120042) has Ticket (111874)
Road	22592	An open way (generally public) for travel or transportation	has Length (28259) has Road ID (120051) has Speed Limit (35726)	has Address (36400) has Duration (80581) has Facility (3012) has Mode Of Transport (120044)
Stop	5446	A brief stay in the course of a journey	has Stop Code (120053) has Stop ID (120051) has Stop Name (120054) has Wheel Chair Boarding (120055)	has Location (132)
Stop Time	120042	Temporal parameters for a specific stop	has Cost (70407) has Currency Type (120061)	has Calendar (44719) has Stop Time (5446) has Time Stamp (120046)
Ticket	111874	Provide with a ticket for passage or admission	has Fare ID (70599) has Payment Method (120061) has Time Table Duration (120088)	has Price (28431)
Time Stamp	120046	Hour, minutes and seconds of a duration	has Hour (81114) has Minutes (81154) has Seconds (72173)	

Table 13: Ontology Elements Metadata

The class Enumeration is a class needed to collect all those lists of elements which can be selected to specify a particular attribute of a class. In particular we have designed three different lists:

Parent EType	Element	GID	Concept
ModeOftransport_GID-120044			
	Bicycle	15188	A wheeled vehicle that has two wheels and is moved by foot pedals
	Bus	15732	A vehicle carrying many passengers

Parent EType	Element	GID	Concept
	CableCar	15797	A conveyance for passengers or freight on a cable railway
	Car	15945	A wheeled vehicle adapted to the rails of railroad
	Foot	1429	The act of traveling by foot
	Train	18679	Wheelwork consisting of a connected set of rotating gears by which force is transmitted or motion or torque is changed
Facility_GID-3012			
	BikeSharing	843	The act of maneuvering a vehicle into a location where it can be left temporarily
	BusStation	15745	A terminal that serves bus passengers
	CampsiteParking	45940	A site where people on holiday can pitch a tent
	FuelStation	18641	A service station that sells gasoline
	ParkingArea	46375	A lot where cars are parked
	RailwayStation	22321	Terminal where trains load or unload passengers or goods
PrivateTransport_GID-120043			
	Diesel	17309	An internal combustion engine that burns heavy oil
	Electric	61771	A physical phenomenon associated with stationary or moving electrons and protons
	Gas	79121	A fluid in the gaseous state having neither independent shape nor volume and being able to expand indefinitely
	Methane	79566	A colorless odorless gas used as a fuel
	Petrol	78042	A volatile flammable mixture of hydrocarbons (hexane and heptane and octane etc.) derived from petroleum

Table 14: Enumeration Metadata

1.4 Knowledge Graph Evaluation

In this final section we are going to show the results obtained by computing the evaluation metrics. In order to understand how much is good our ontology we have compared it with several others found in the web, here we are reporting just two of them, please look at the section xxx to have more information.
In particular we computed 4 different metrics: Coverage, Flexibility, Extensiveness and Sparsity.The first reference schema is taken by :

Information	Data
Name	km4city
URL	http://wlode.disit.org/WLODE/extract?url=http://www.disit.org/km4city/schema#Support{_}activities{_}for{_}transportation
Ontology IRI	http://www.disit.org/km4city/schema
Authors	DISIT lab
Publisher	DISIT Lab, University of Florence, Italy, http://www.km4city.org
Coverage	0.01
Flexibility	0.02
Extensiveness	0.01
Sparsity	0.97

Comparing our ontology with this one, the first thing to say is that there is a very huge difference in the number of ETypes considered: the ontology provided by the DISIT Lab is indeed composed by 667 ETypes while our one just 18. In this case the most important metric is the Sparsity: it indicates that there is an important difference between the ETypes defined by us and those defined in the km4city ontology. The second thing to highlight is the similarity among the other metrics, indeed it indicates that our ontology is not well represented by the other schema.

Information	Data
Name	Tickets Ontology
URL	http://www.heppnetz.de/ontologies/tio/ns#TransportationService
Ontology IRI	
Authors	Martin Hepp
Publisher	
Coverage	0.26
Flexibility	0.42
Extensiveness	0.11
Sparsity	0.63

In this case it is possible to see as Sparsity is lower than in the case before, this is due to the amount of classes: in this case indeed the schema took as reference is composed by 42 ETypes and this means that there is no a huge difference between the schemes. In addition an higher Coverage value indicates that the ontology above is more similar to our one, this means that if going ahead in exploring the domain we could obtain a very interesting ontology. A good indicator is instead the flexibility, being almost at 50 means that our schema could potentially become a very good graph if integrated in order to explore more into details the domain.