KDI : Knowledge and Data Integration

Transportation

KDI Demo Presentation

Contributors

- Data Scientist: Alberto Carbognin
- Data Scientist: Omid Jadidi
- Domain Expert: Fivos Kapidis
- Knowledge Engineer: Fivos Kapidis
- Knowledge Engineer: Antonio Luigi Stefani
- Project Manager: Antonio Luigi Stefani
- Tutor Data: Alessio Zamboni
- Tutor Knowledge: Mayukh Bagchi

Table of Contents

Table of Contents

Project description

Our ontology wants to explore the transportation domain in order to solve the daily problems of finding cheapest and shortest routes to reach a certain place.

Our goal is to provide all the more useful information regarding the road system, exploring also the railway system, the cycling routes and the mountain trails. In particular to reach our goal we have circumscribed the problem into two main issues:

- information regarding the road system itself: roads, modes of transport, prices, scheduling of the public transport.
- information regarding the facilities related to each road and public transport service.

Project description

To design our system first of all we put ourselves in the users shoes:

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

Figure: Users Description

Table of Contents

SKG description

The SKG has been developed starting from the Informal Modeling Phase. Accordingly to the iTelos Methodology we have divided our work in several steps: first of all we identified all the ETypes needed to design our EER Model focusing in highlighting the role of each EType: Core, Common, Contextual. In particular we have designed our schema in order to adapt it as much as possible to the GTFS standard.

The second step was to formalize all the model designed previously using Protégé. The goal of the project was to provide an Ontology accordingly to the concepts defined in the KOS software. To do this we have integrated the knowledge already available in the KOS adding several classes and properties such to obtain a unique global ID to each one.

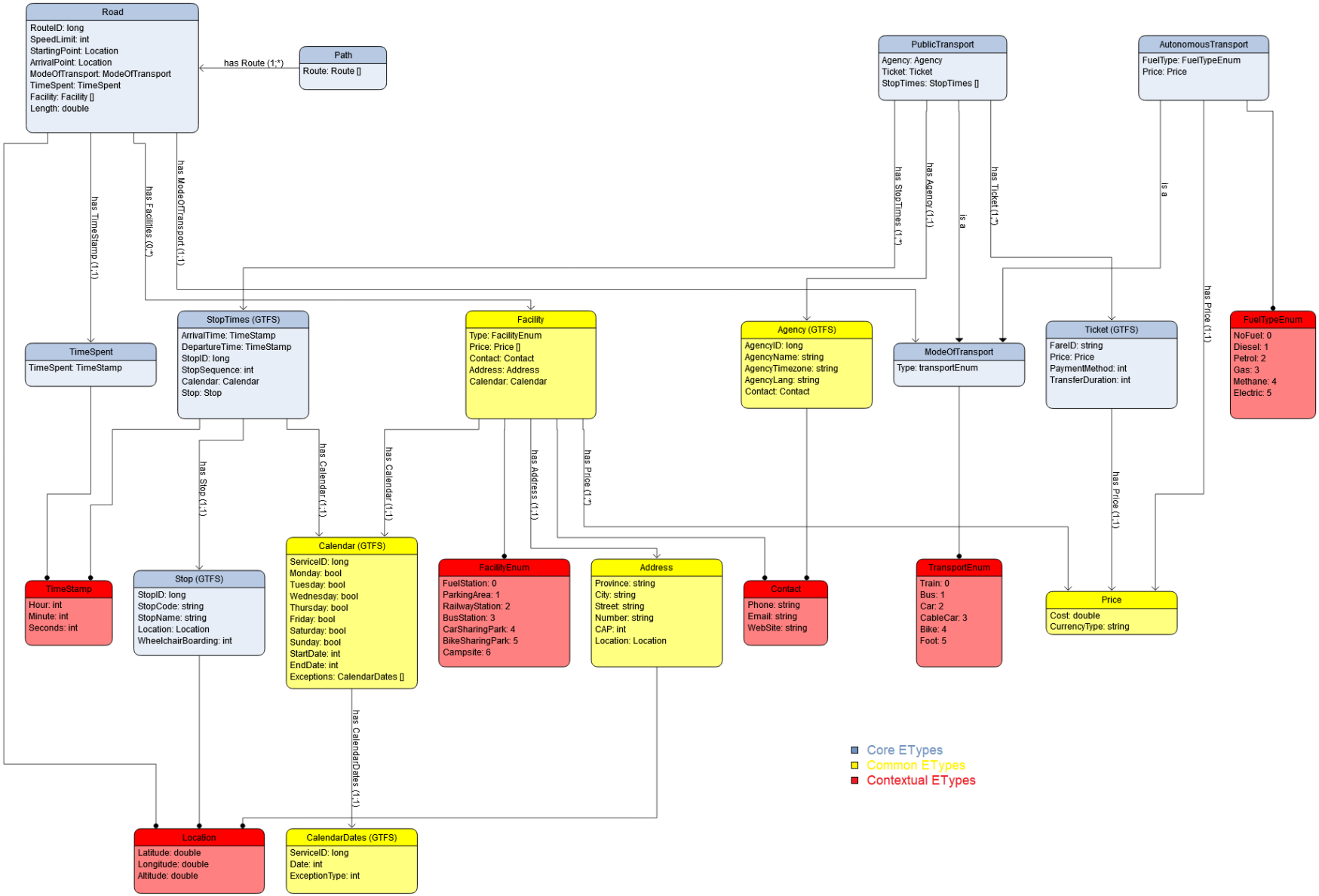
SKG description: Informal Modeling Phase

During the Informal Modeling Phase we designed the EER model accordingly the iTelos Methodology. Firstly we have identified all the ETypes taking part in the projecy:

- Common: Facility, Agency, Calendar, CalendarDates, Address, Price.
- Core: Path, Road, ModeOfTransport, PublicTransport, PrivateTransport, Stop, StopTimes, Ticket.
- Contextual: TimeStamp, Location, Contact.
- Enumeration: FacilityEnum, TransportEnum, FuelEnum.

Then we passed to build a model on a graph visualizer: yEd.

SKG description: EER Model



SKG description: Formal Modeling Phase

After having designed the EER model we passed to the software Protégé in which we have imported all the classes and their properties defining for each class a unique Global ID (GID).

Once obtained the RDF code of our ontology we proceed in its visualization by using the online tool: WebVOwl.

SKG description: Ontology

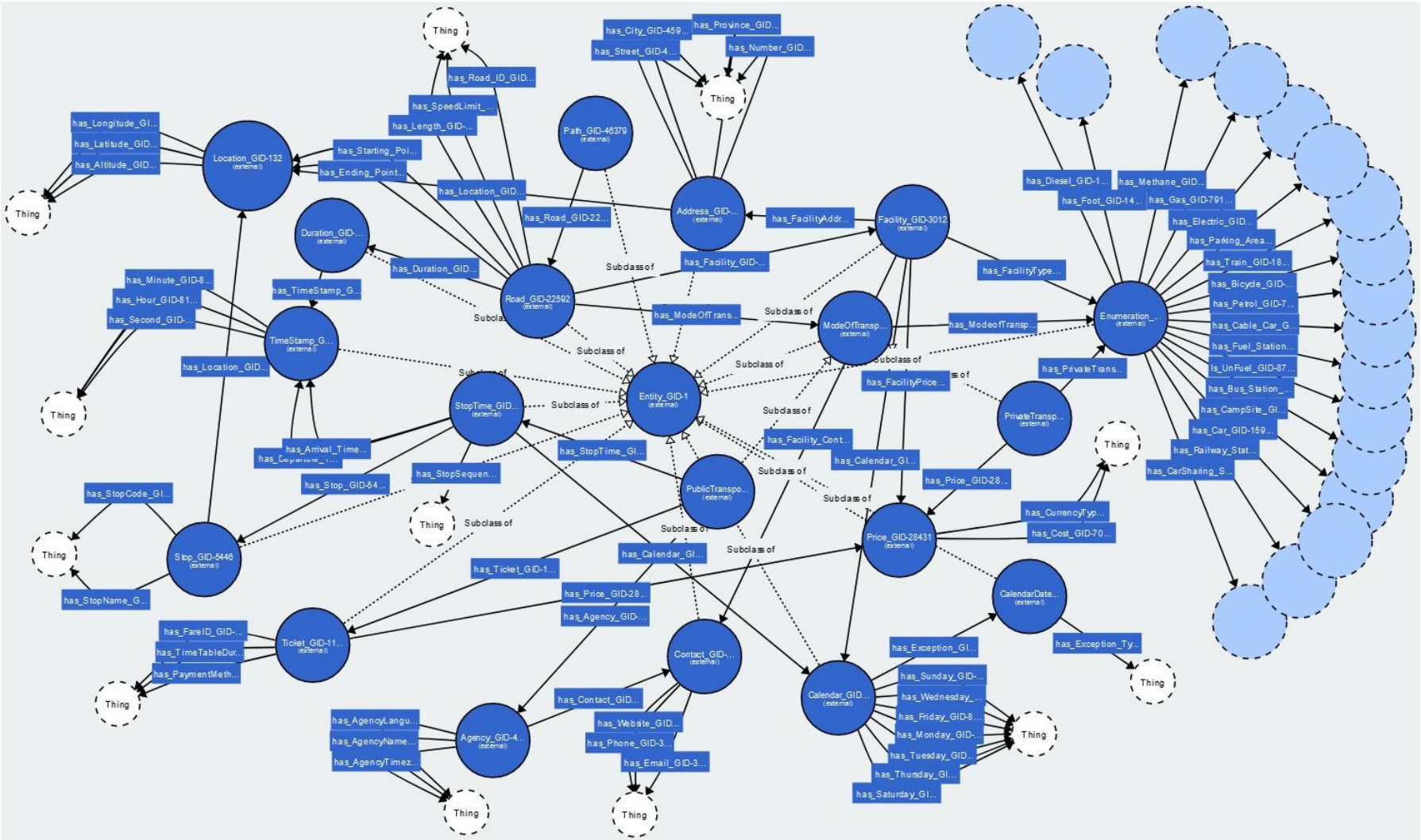


Figure: EER Model

Table of Contents

Data description

To achieve our goal we identified several data sources, those which provided the most important datasets are:

- Trentino Trasporti S.r.L.
- OpenStreetMap
- DataTrentino
- Provincia.tn.it
- comune.trento

In particular our needs were focused on: public transport (Urban and ExtraUrban) involving routes and scheduling, Cycling routes, Cycling facilities and point of interest, Italian Parking Areas, Railway Stations, Car sharing, Map of Petrol Stations in Italy, Mountain Trails, Campsite and other accommodation facilities.

Data description

During all steps of the project (According to the iTelos Methodology) we have transformed, filtered and aligned the datasets in order to have the same format and to use them in the next phases. In particular the format adopted has been the ttl. The datasets are possible to be found in our Github repository.

To acquire knowledge on the datasets we have also collect a series of metadata describing them. They are all in json format and are possible to be found in our Github repository

Table of Contents

DKG description

To obtain the DKG the goal was to align the data collected during the project to the ontology designed in the Formal Modeling Phase. To achieve this result we went through several steps:

- filter and transform datasets in order to obtain them in the ttl format;
- map the data to the ontology by using the KarmaLinker software;
- import ttl files into the GraphDB in order to be able to query and visualize entities.

We used **KarmaLinker** to map data with ontology. All data after filtering, seperation and alignment imported in karma, defined unique **URI**, connected through outgoing link; which they were generated using pyTransformer functionality and added to a new coloumn. At the end we exported the rdf to be queried and visualized in graphDB.

DKG description

Queries: query 1:1.11 List all the cycling starting and ending points closer than 10 Km to the San Bartolomeo train station.

SPARQL Query & Update

Unnamed | Unnamed | Unnamed

```
prefix onto:<http://www.ontotext.com/>
prefix ontology:<http://knowdive.disi.unitn.it/etype#>
PREFIX omgeo:<http://www.ontotext.com/owlim/geo#>

select ?road ?location ?latitude ?longitude
where
{

    {
        #strting facelet for find sanbartalomeo


        select ?mylat ?mylong where {
            ?address ontology:has_street_address_GID-45807_Type-36400 ?street.
            FILTER regex(?street, "Malpensada").
            ?address ontology:has_location_GID-132_Type-36400 ?location.
            ?location ontology:has_latitude_GID-46263_Type-132 ?mylat.
            ?location ontology:has_longitude_GID-46270_Type-132 ?mylong.
        }
    }

    ?road ontology:has_location_GID-132_Type-138 ?location.
    ?location ontology:has_latitude_GID-46263_Type-132 ?latitude.
    ?location ontology:has_longitude_GID-46270_Type-132 ?longitude.
    FILTER( omgeo:distance(?mylat, ?mylong, ?latitude,?longitude) < 5).
}
```

Figure: Query 1 Result

DKG description

Filter query results

⚠ Showing results from 1 to 68 of 68. Query took 0.2s, yesterday at 21:09.

	road	location	latitude	longitude	address	mylat	mylong
1	http://localhost:8	http://localhost:8	"46.035290875 26111"	"11.12459717697 7254"		"46.047565"	"11.1352074"
2	http://localhost:8	http://localhost:8	"46.035419235 46627"	"11.1245893147 3671"		"46.047565"	"11.1352074"
3	http://localhost:8	http://localhost:8	"46.035127025 44549"	"11.1223057542 51887"		"46.047565"	"11.1352074"
4	http://localhost:8	http://localhost:8	"46.035608035 58216"	"11.1223462202 88074"		"46.047565"	"11.1352074"
5	http://localhost:8	http://localhost:8	"46.035608035 58216"	"11.1223462202 88074"		"46.047565"	"11.1352074"
6	http://localhost:8	http://localhost:8	"46.0351606198	"11.1222701405		"46.047565"	"11.1352074"

Figure: Query 1 code

DKG description

query 2:4.7 Give the closest available ”car sharing” position to a specific point.

SPARQL Query & Update

Unnamed | Unnamed | Unnamed

```
prefix onto:<http://www.ontotext.com/>
prefix ontology:<http://knowdive.disi.unitn.it/etype#>
PREFIX omgeo:<http://www.ontotext.com/owlim/geo#>

select ?street ?latitude ?longitude ?enumvalue
where
{
    {
        #strting facelet for find sanbartalomeo
        select ?mylat ?mylong where {
            ?address ontology:has_street_address_GID-45807_Type-36400 ?street.
            FILTER regex(?street, "Malpensada").
            ?address ontology:has_location_GID-132_Type-36400 ?location.
            ?location ontology:has_latitude_GID-46263_Type-132 ?mylat.
            ?location ontology:has_longitude_GID-46270_Type-132 ?mylong.
        }
    }
    ?facility ontology:has_address_GID-36400_Type-22592 ?address.
    ?address ontology:has_street_address_GID-45807_Type-36400 ?street.
    ?facility ontology:has_type_GID-103418_Type-3012 ?enum.
    ?enum ontology:has_class_GID-43482_Type-1 ?enumvalue.
    ?address ontology:has_location_GID-132_Type-36400 ?location.
    ?location ontology:has_latitude_GID-46263_Type-132 ?latitude.
    ?location ontology:has_longitude_GID-46270_Type-132 ?longitude.
    FILTER( omgeo:distance(?mylat, ?mylong, ?latitude,?longitude) < 10).
    FILTER regex(?enumvalue ,"CarSharingPark").

}
```

Figure: Query 1 code

DKG description

Filter query results

⚠ Showing results from 1 to 7 of 7. Query took 0.1s, today at 01:45.

	street	latitude	longitude	enumvalue
1	"Via Giovanni Segantini"	"46.0746447"	"11.1217133"	"CarSharingPark"
2	"Via Santa Croce"	"46.0647501"	"11.1232546"	"CarSharingPark"
3	"Piazza Venezia"	"46.0695192"	"11.1276056"	"CarSharingPark"
4	"Piazza Dante"	"46.0712884"	"11.1213118"	"CarSharingPark"
5	"Via dei Solteri"	"46.08609955"	"11.1214243670073 15"	"CarSharingPark"
6	"Via Don Tommaso Dallafior"	"46.0654572"	"11.1543055"	"CarSharingPark"
7	"Largo Donatori volontari del	"46.0563978"	"11.131146627108679"	"CarSharingPark"

Figure: Query 2 code

DKG description

query 3:2.6 Give the longest path in Trentino.

```
# Returns the longest path in Trentino

prefix onto:<http://www.ontotext.com/>
prefix ontology:<http://knowdive.disi.unitn.it/etype#>
PREFIX omgeo:<http://www.ontotext.com/owlim/geo#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

select ?road ?lengthF ?length
where
{
    ?road ontology:has_length_GID-28259_Type-22592 ?length .
    BIND(REPLACE(STR(?length), ",", "") AS ?lengthF) FILTER (xsd:float(?lengthF) > 1).
}
ORDER BY DESC(xsd:float(?lengthF)) LIMIT 1
```

Run

Table | Raw Response | Pivot Table | Google Chart

Download as

Filter query results

Showing results from 1 to 1 of 1. Query took 0.1s, yesterday at 22:03.

	road	lengthF	length
1	http://localhost:8080/api/route_150691	"67300"	"67,300"

Figure: Query 3 result

DKG description

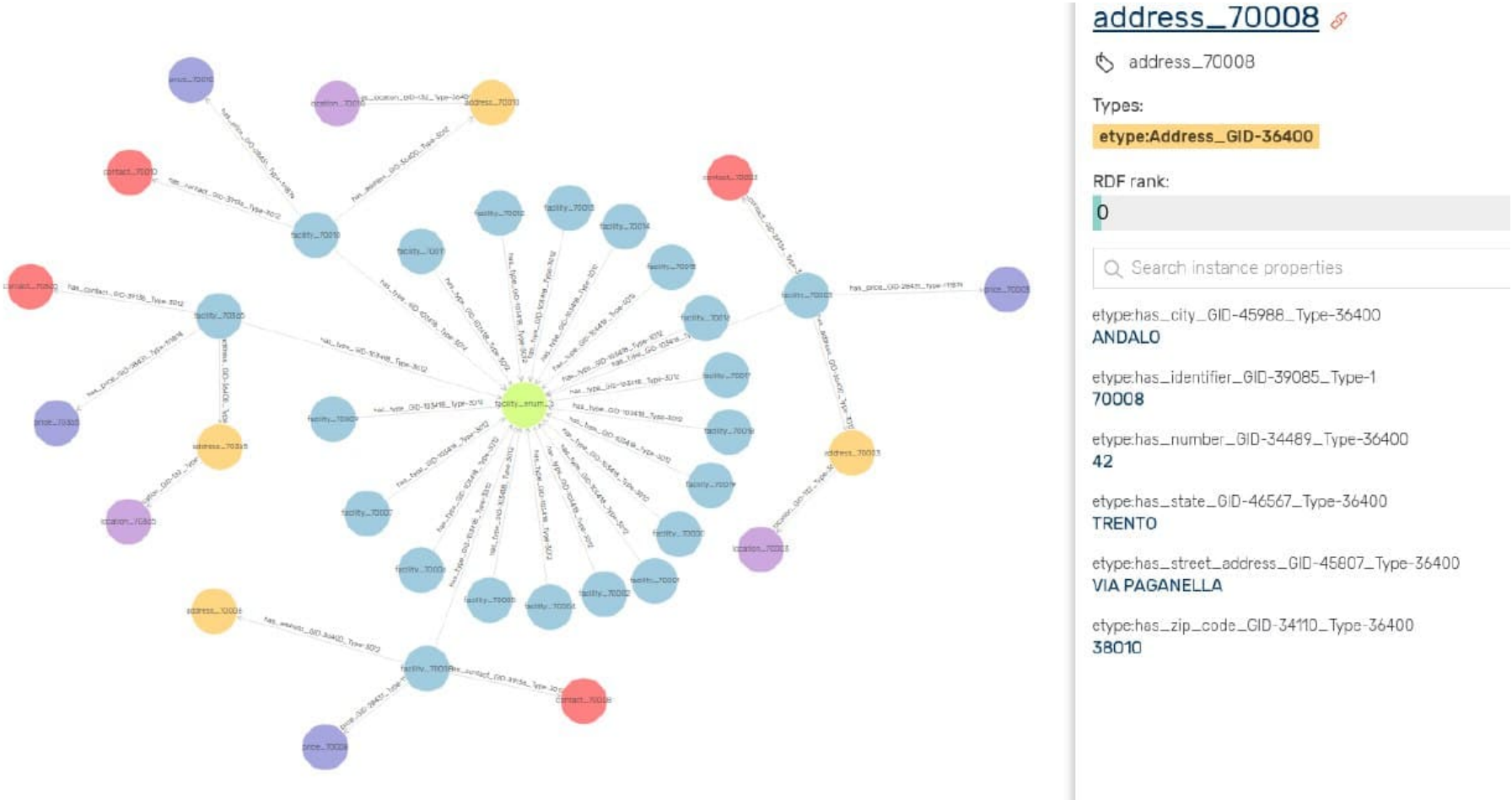


Figure: Graph Visualization of some instances of Facility

Table of Contents

Conclusions

The final DKG results working and well structured. Despite of this we are not totally satisfy because of a coverage of the 93% on the queries defined at the beginning. This is due to a lack of information in the dataset, in particular we are missing some information on the prices of the several facilities considered. This because of the owner of datasets requiring a payment to obtain the needed data.

Thanks for the attention,
Omid Jadidi, Alberto Carbognin, Fivos Kapidis, Antonio Luigi Stefani

KDI : Knowledge and Data Integration

Transportation

KDI Demo Presentation