



KDI • Knowledge and Data Integration

‘Geospatial domain’

KDI Demo Presentation

Contributors

- Data Scientist: Michele Tessari
- Data Scientist: Daniele Mattedi
- Domain Expert: Matteo Bortolon
- Knowledge Engineer: Matteo Bortolon
- Knowledge Engineer: Sergio Povoli
- Project Manager: Michele Tessari
- Tutor Data: Alessio Zamboni
- Tutor Knowledge: Mayukh Bagchi

Table of Contents

1 Project description

2 SKG description

3 Data description

4 DKG description

5 Conclusions

Table of Contents

1 Project description

2 SKG description

3 Data description

4 DKG description

5 Conclusions

Project description



Table of Contents

1 Project description

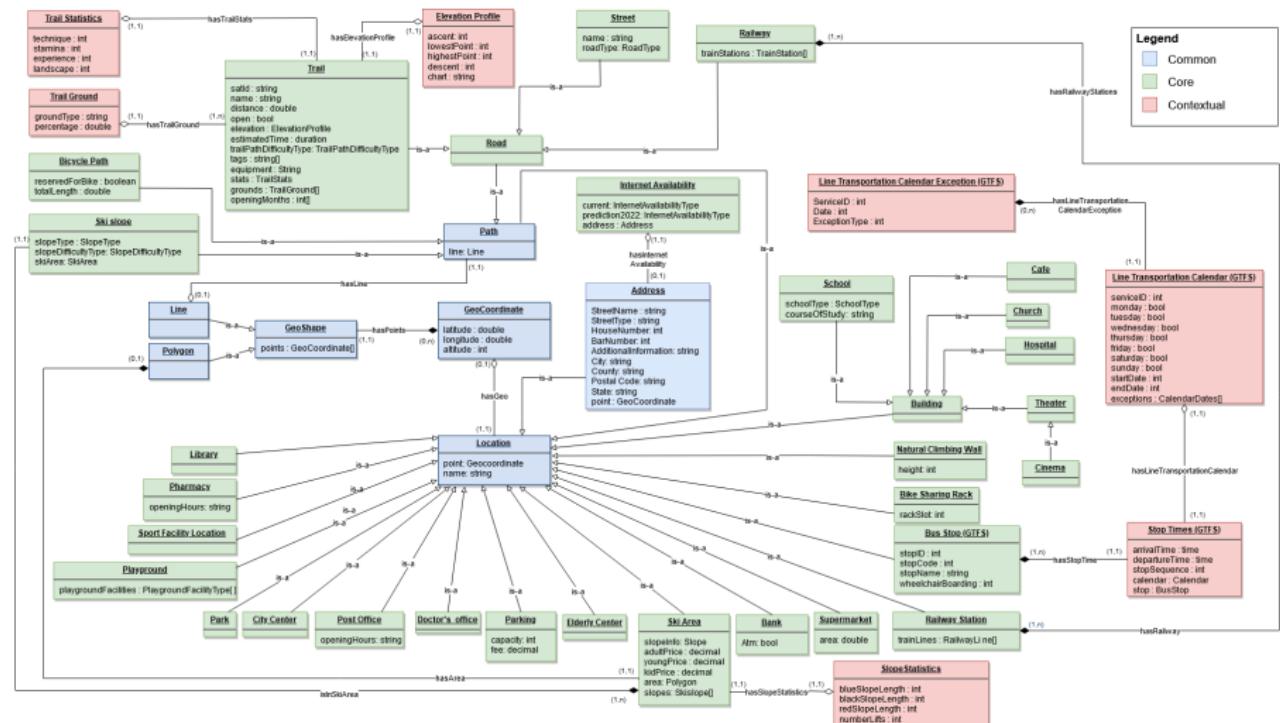
2 SKG description

3 Data description

4 DKG description

5 Conclusions

EER



SKG

The image shows two separate ontology editors side-by-side, both titled "OntologyID(Anonymous-6)".

Left Editor (Pharmacy Ontology):

- Annotations:** A tooltip for "Place" defines it as "Entity describing a chemist or drug store".
- Description: Pharmacy**
 - Equivalent To: None
 - SubClass Of:
 - spinningheels:has_facilizing
 - Place
 - General class axioms: None
 - Instances: None
 - Target for key: None
 - Disjoint With: None
 - Disjoint Union Of: None

Right Editor (Drugstore Ontology):

- Annotations:** A tooltip for "Drugstore" defines it as "Drugstore_GD-17637".
- Description: Drugstore_GD-17637**
 - Equivalent To: None
 - SubClass Of:
 - Locator_GD-12
 - General class axioms: None
 - Instances: None
 - Target for key: None
 - Disjoint With: None
 - Disjoint Union Of: None

Complete ontology

KOS ontology

'Geospatial domain'

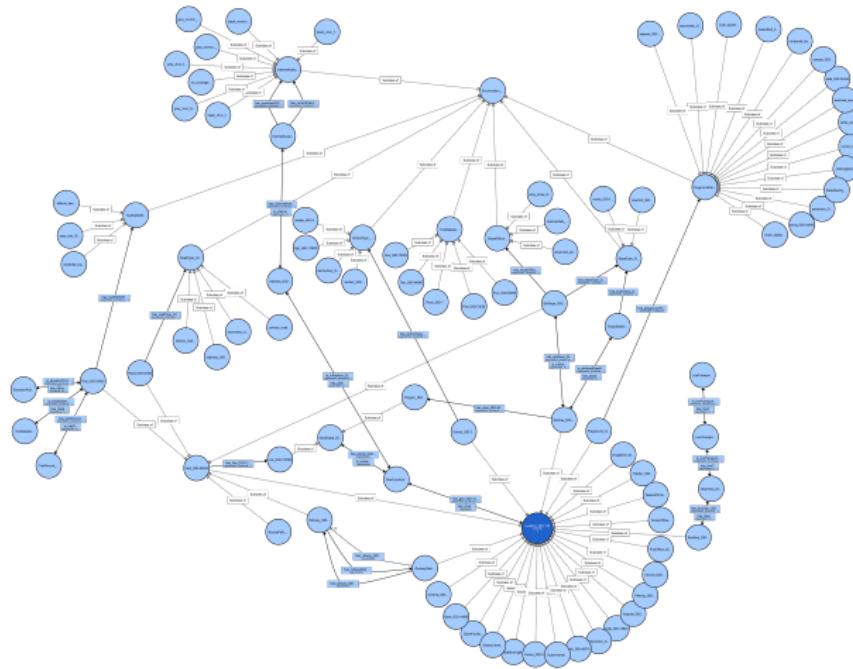


Table of Contents

1 Project description

2 SKG description

3 Data description

4 DKG description

5 Conclusions

Data description

32 datasets representing all the etypes, coming from:

- **Openstreetmap**, extracted thanks to *overpass-turbo.eu*
- **DataTrentino**
- **webscraping**

All the data are in `json` format and stored in this [Data folder](#)

Metadata documentations is stored in [this folder](#) (DCAT standard).

Meaning of attributes are examined in depth in the files stored in [this folder](#).

Scraping scripts are in this [folder](#).

Data description

An other important aspect was the collaboration with other 2 groups, Tourism facilities and Transportation. They provided us 4 datasets:

- **Trails**
- **Bus stops and stop times**
- **Ski Resort**, which has been merged later with **Ski Area**

Also to these Dataset the operations which later we are going to mention have been applied.

Data description

Main operations applied on datasets are:

- **filtering** (applied in order to remove sparse attributes and aligning to the ontology) with script that are stored in this [folder](#)
- **separation of datasets** (e.g *Point of Interest*, from which information about single etypes have been extracted)
- **separation of attributes** (e.g. address)
- small manual operations have been applied.

Other operations have been carried out in order to provide effective relationship between Etypes and can be found [here](#). These script were applied to *Ski Area*, *Ski Slopes*, *Railway* and *Railway Stations*.

Table of Contents

1 Project description

2 SKG description

3 Data description

4 DKG description

5 Conclusions

DKG description

The tool used to align data to reference ontology is **KarmaLinker**. All the datasets has been imported in KarmaLinker, where the manual linking of new data sources has been performed.

We obtain for each dataset 3 types of data: the model, the `RDF` and the `EML`. These data are all stored in this [folder](#).

To visualize the graph we used the `GraphViz` tool and installing the python library `rdflib`.

The results obtained are stored [here](#).

The queries have been performed thanks to the `GraphDatabase` `GraphDB`, importing our `RDFs` and using the query language `SPARQL` to have the following [queries](#).

DKG description

CQ: Luciano and Corrado, 4.8

- Which and where are the hospitals in Trentino?

The screenshot shows a SPARQL query interface with the following details:

- Title:** CQ_Luciana_4_10
- Prefixes:**

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

```
select ?hospital_name ?longitude ?latitude
where {
  ?hospital etype:has_name_GID-2_Type-132 ?hospital_name.
  ?hospital rdf:type etype:Hospital_GID-19345.
  ?hospital etype:has_geocoordinate_GID-120951_Type-132 ?geocoordinate.
  ?geocoordinate etype:has_longitude_GID-46270_Type-120951 ?longitude.
  ?geocoordinate etype:has_latitude_GID-46261_Type-120951 ?latitude
}
```
- Buttons:** geospatial, Run
- Table Headers:** hospital_name, longitude, latitude
- Data:**

	hospital_name	longitude	latitude
1	"Casa di Cura "Villa Bianca"	"11.12668658333333"	"46.06253391666666"
2	"Pronto Soccorso Ospedale Santa Maria del Carmine"	"11.0401616"	"45.8811734"
3	"Ospedale San Camillo"	"11.132001558333332"	"46.06490974166667"
4	"Pronto Soccorso Ospedale di Arco"	"10.875114"	"45.919077"
5	"Pronto Soccorso Ospedale Tione di Trento"	"10.7245566"	"46.0422014"
6	"Ospedale Riabilitativo Villa Rosa"	"11.24912533888888"	"46.06447616666669"
7	"Ospedale Valli del Noce"	"11.032800792307693"	"46.361310757692316"
- Bottom Buttons:** Table, Raw Response, Pivot Table, Google Chart, Download as

DKG description

CQ: Giuseppe, 1.1

■ Where is the closest library?

SPARQL Query & Update

geospatial

Editor only Editor and results Results only

Run

Unnamed CQ_Luciana_4_10 Unnamed 

```
11 ?library etype:has_geocoordinate_GID-120051_Type-132 ?geocoordinate.
12 ?geocoordinate etype:has_longitude_GID-46270_Type-120051 ?longitude.
13 ?geocoordinate etype:has_latitude_GID-46263_Type-120051 ?latitude.
14 BIND (xsd:float(?latitude) AS ?lat2).
15 BIND (xsd:float(?longitude) AS ?lon2).
16 BIND (46.06253916666666 AS ?lat1).
17 BIND (11.12668658333333 AS ?lon1).
18 BIND ((6371e3 * 2 * ofn:atan2(ofn:sqrt(ofn:sin((?lat2-?lat1) * ofn:pi()/180)/2) * ofn:sin((?lat2-?lat1) * ofn:pi()/180)/2) + ofn:cos(?lat1 * ofn:pi()/180) * ofn:cos(?lat2 * ofn:pi()/180) * ofn:sin((?lon2-?lon1) * ofn:pi()/180)/2) * ofn:sin((?lon2-?lon1) * ofn:pi()/180)/2), ofn:sqrt(1 - (ofn:sin((?lat2-?lat1) * ofn:pi()/180)/2) * ofn:sin((?lat2-?lat1) * ofn:pi()/180)/2) + ofn:cos(?lat1 * ofn:pi()/180) * ofn:cos(?lat2 * ofn:pi()/180) * ofn:sin((?lon2-?lon1) * ofn:pi()/180)/2) * ofn:sin((?lon2-?lon1) * ofn:pi()/180)/2))) AS ?distance)
19 }
20 }
```

Table Raw Response Pivot Table Google Chart Download as

Filter query results Showing results from 1 to 1 of 1. Query took 0.1s, moments ago.

	library_name	latitude	longitude	distance
1	"DAE Biblioteca centrale di via Roma"	"46.069688"	"11.120852"	"913.877214470929"

DKG description

CQ: Giovanna, 3.6

- Which are the supermarkets within a radius of 2 km? How big they are?

SPARQL Query & Update ①

Editor only Editor and results Results only □

Unnamed × Unnamed × CO_Giovanna_3.6 × @

```
1 PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
2 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 PREFIX ofn: <http://www.ontotext.com/sparql/functions/>
5
6 select ?supermarket_name ?area ?latitude ?longitude ?distance
7
8 where {
9   ?supermarket etype:has_name_GID-2_Type-132 ?supermarket_name.
10  ?supermarket rdf:type etype:Supermarket_GID-24168.
11  ?supermarket etype:has_geocoordinate_GID-120051_Type-132 ?geocoordinate.
12  ?geocoordinate etype:has_longitude_GID-46270_Type-120051 ?longitude.
13  ?geocoordinate etype:has_latitude_GID-46263_Type-120051 ?latitude.
14 }
```

Run

Table Raw Response Pivot Table Google Chart

Download as ...

Filter query results

Showing results from 1 to 4 of 4. Query took 0.4s, minutes ago.

	supermarket_name	area	latitude	longitude	distance
1	"EuroSpar"	"1583.1989318240705"	"46.0542897375"	"11.127855075"	"700.2009659917982"**xsd:double
2	"Tovazzi Europa"	"468.3791687083169"	"46.07006891428573"	"11.118663742857143"	"1041.61933217474"**xsd:double
3	"Margherita Conad"	"536.7363991967187"	"46.053261355555555"	"11.123914755555555"	"1053.124409558152"**xsd:double
4	"Poli Regina"	"1180.1041217235936"	"46.04710611249999"	"11.1261839"	"1715.8036545689306"**xsd:double

Table of Contents

1 Project description

2 SKG description

3 Data description

4 DKG description

5 Conclusions

Conclusions

Thanks for attention.

If you have any question, please ask.



KDI • Knowledge and Data Integration



'Geospatial domain'

KDI Demo Presentation