

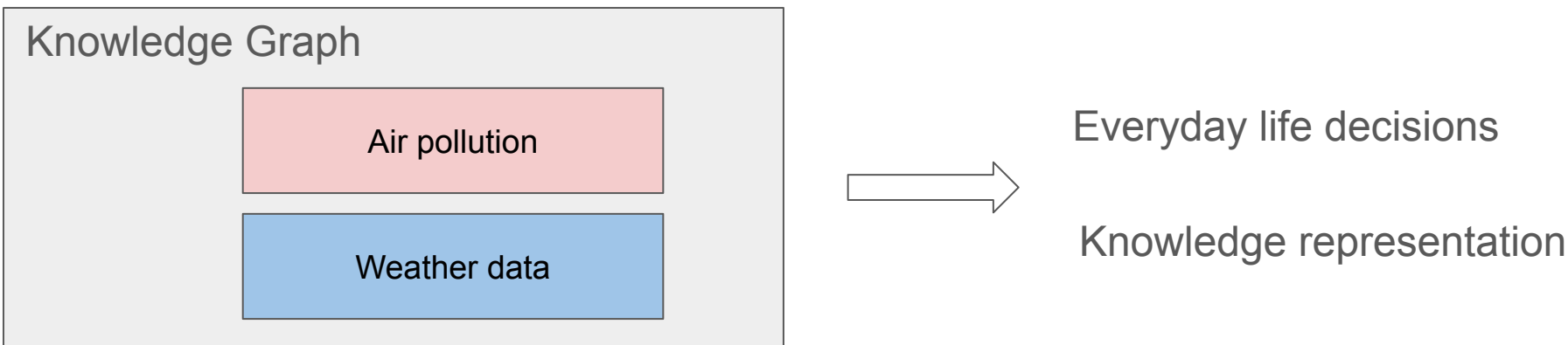
# Weather and Climate Change in Trentino

Veronika Deketova, Hasan Aldhahi

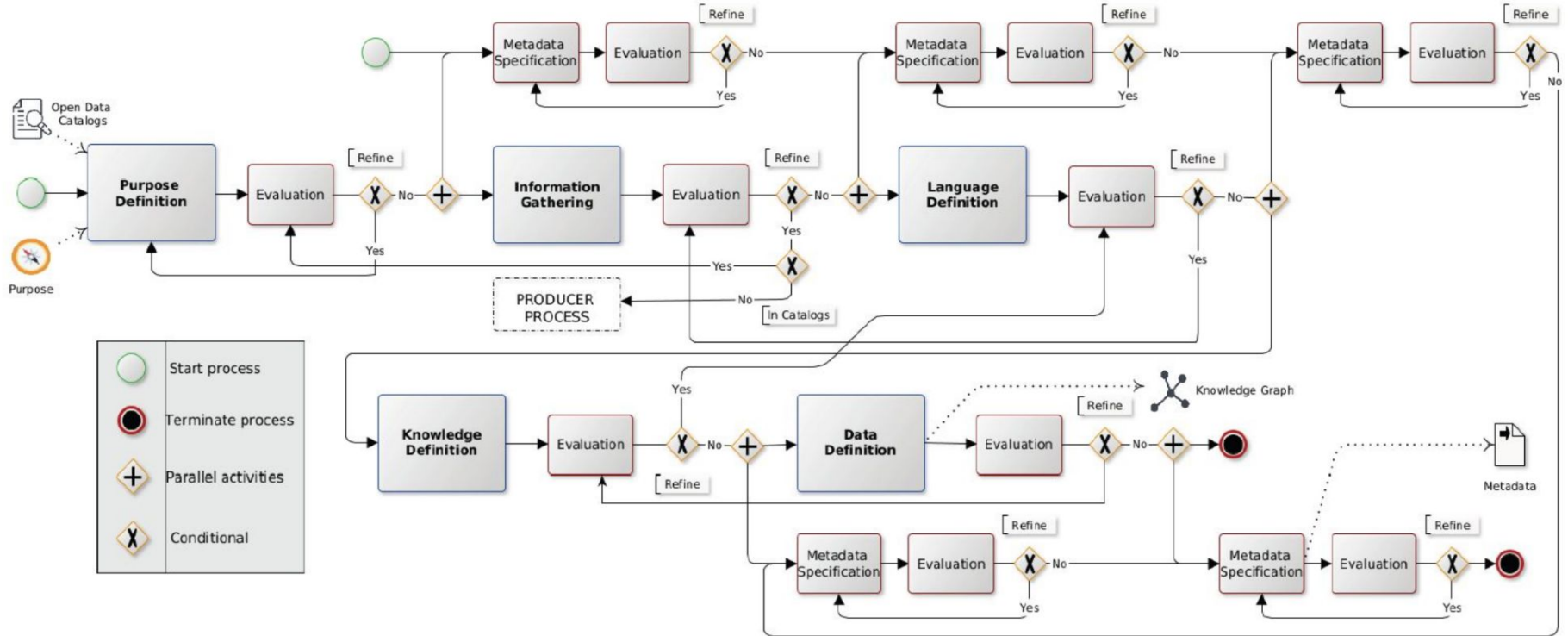
University of Trento  
Knowledge Graph Engineering class 2023/2024

# Introduction

- Air pollutant concentration can have a significant effect on human health
- Connection of climate change and air pollution
- EU statistics: Premature deaths due to exposure PM2.5
- Our KGE allow:
  - Monitor situation regarding the air quality data and weather data in the Trentino region
  - Give a perception of the situation in given area
  - Different time scenarios: history, current and prediction



# ITelos Approach



# Scenarios

1

Non-negligible number of individuals dealing with lowered immunity or lowered breathing conditions (fe. asthma).

Growing awareness about air pollution effects, increase in number of individuals concerned about their health.

2

3

Trip planning for spending time outdoors – what will the weather situation be like.

Deciding location for sports event organization

4

5

Effect of air quality and weather stability on agriculture.

# Personas and CQ

**Persona 1:** *Matteo, 35-year-old man living in Borgo Valsugana, asthmatic*

**CQ 1:** Provide Matteo the most current air pollution data for a given day closest to Borgo Valsugana, so he can make a decision about the time he wants to spend outside given his health condition. Return a list of all pollutants together with their concentration.

**Persona 2:** *Lenka, a 24-year-old professional triathlete living close to Bolzano*

**CQ 2:** Give Lenka air pollution prediction for next day close to Bolzano, so she can adjust her training plan accordingly and potentially move some of her training sessions indoors. The data are suppose to be from a station which is closest to Lenka's home.

**Persona 3:** *Sarah is a 29-year-old digital nomad from Norway planning her next destination in Trentino*

**CQ 3:** Sarah is interested in the long-term history of temperature, rainy days, and wind speed in cities in the Trentino area.

# Personas and CQ

**CQ4:** Provide Sarah with a list of the average values of O3 and NO2 for stations available through the history of the dataset, ordered by the smallest average value.

**Persona 4:** *Paolo is a 19-year-old student who just moved to Trento, he is planning trip with his family*

**CQ 5:** Paolo is interested in the precipitation and temperatures status for current day for Trento.

**Persona 5:** *Mario and Suse, family farmers from Rome family farmers aiming to set up a new ecological farm in the Trentino area*

**CQ6:** Provide Mario and Suse with data about areas in Trentino and history of temperatures, number of rainy days and levels of air pollution and which area has the best values. This means provide Mario and Suse a list of all historical reports for all weather and pollution stations in Trentino area.

# Data Gathering

- Main issue: find quality data from open sources
- Different possible platforms



**European  
Environment  
Agency**

## *Current pollution data + history*

- Data in **csv** format
- Archived **parquet** data
- Missing description, mappings



IMPLEMENTED BY



## *Pollution prediction*

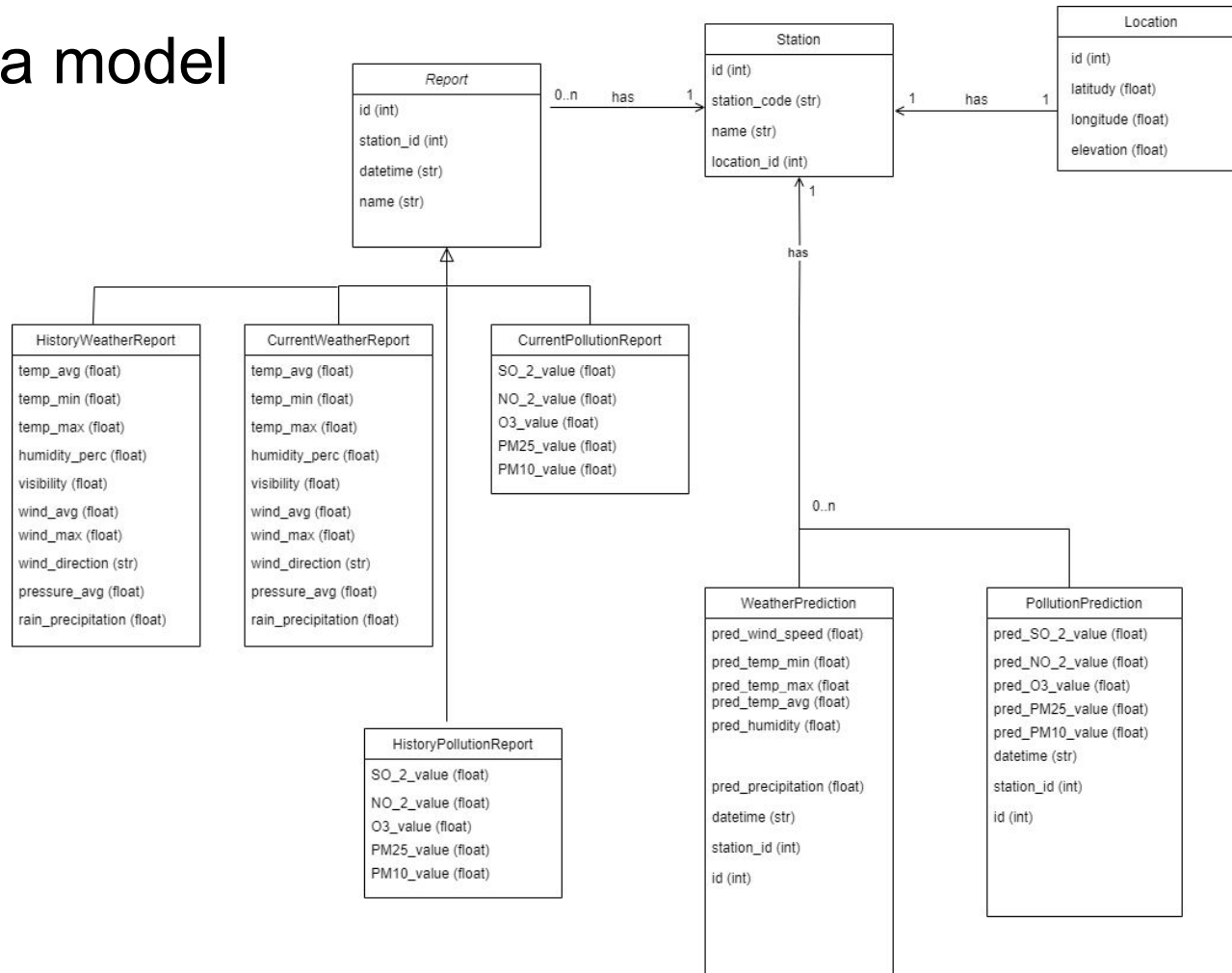
- Data in **grib** format
- Documentation available on Confluence
- ecCodes package



## *Weather data for Trentino*

- Data aligned with previous KGE project resources

# ER data model

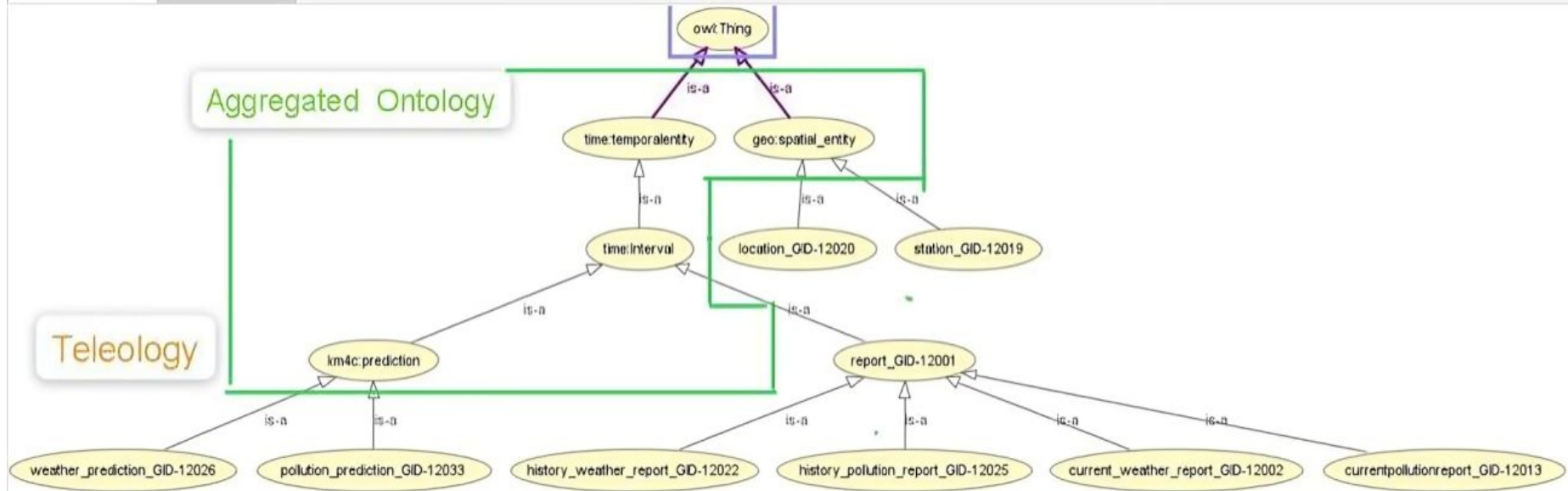




# Language Modeling

Concept Labels	Description
forecast_GID-36179	a prediction about how something (as the weather) will develop
Report_GID-12001	Its measurment that could be in the present or past or future
CurrentWeatherReport_GID-12002	A real-time report providing up-to-the-minute data on current weather conditions, including average temperature, temperature range, humidity, visibility, wind speed and direction, atmospheric pressure, and precipitation.
temp_avg_GID-12003	The average temperature, typically in degrees Celsius, at the time of the report.
temp_min_GID-12004	The minimum temperature, typically in degrees Celsius, recorded at the time of the report.
temp_max_GID-12006	The maximum temperature, typically in degrees Celsius, recorded at the time of the report.
humidity_perc_GID-12007	The percentage of relative humidity in the air at the time of the report.
visibility_GID-26112	capability of providing a clear unobstructed view
wind_avg_GID-12008	The average wind speed, often in meters per second, at the time of the report.
wind_max_GID-12009	The maximum wind speed, often in meters per second, observed at the time of the report.
wind_direction_GID-12010	The cardinal or compass direction from which the wind is blowing.
pressure_avg_GID-12011	The average atmospheric pressure, often in millibars or hectopascals, at the time of the report.
rain_precipitation_GID-12012	The amount of precipitation, typically in millimeters, that has fallen or is expected to fall at the time of the report.
CurrentPollutionReport_GID-12013	A real-time report that provides current data on pollution levels, including measurements of sulfur dioxide (SO2), nitrogen dioxide (NO2), ozone (O3), fine particulate matter (PM2.5), and coarse particulate matter (PM10) in the environment.
SO2_value_GID-12014	The current level of sulfur dioxide (SO2) in the environment, typically measured in parts per million (PPM) or micrograms per cubic meter ( $\mu\text{g}/\text{m}^3$ ).
NO2_value_GID-12015	The current level of nitrogen dioxide (NO2) in the environment, typically measured in parts per million (PPM) or micrograms per cubic meter ( $\mu\text{g}/\text{m}^3$ ).
O3_value_GID-12016	The current level of ozone (O3) in the environment, typically measured in parts per million (PPM) or micrograms per cubic meter ( $\mu\text{g}/\text{m}^3$ ).
PM25_value_GID-12017	The current concentration of fine particulate matter (PM2.5) in the air, which are the following elements Si, Ca, Al, Na, Mg, K, and Fe, typically measured in micrograms per cubic meter ( $\mu\text{g}/\text{m}^3$ ).
PM10_value_GID-12018	The current concentration of coarse particulate matter (PM10) in the air, which are the following elements Fe, K, Ca, Na and Al, typically measured in micrograms per cubic meter ( $\mu\text{g}/\text{m}^3$ ).
Station_GID-12019	A physical location equipped with monitoring equipment to collect and report environmental data
Location_GID-12020	Geographical coordinates represented by latitude and longitude, along with elevation data, providing the precise geographic position associated with monitoring stations and facilitating accurate location identification.
id_GID-12021	A unique identifier for a geographical location.
latitude_GID-46263	the angular distance between an imaginary line around a heavenly body parallel to its equator and the equator itself
longitude_GID-46270	the angular distance between a point on any meridian and the prime meridian at Greenwich
elevation_GID-49932	a raised or elevated geological formation
HistoryWeatherReport_GID-12022	A report that contains historical data on past weather conditions, including average and extreme temperatures, humidity, visibility, wind characteristics, atmospheric pressure, and precipitation recorded at a specific point in time in the past.
station_id_GID-12023	An identifier, typically in integer format, used to associate a report or data entry with a specific monitoring station within the system, indicating the source or origin of the data.
datetime_GID-12024	A textual representation of the date and time when a report or data entry was generated or recorded, typically in a specific format to convey temporal information.
HistoryPollutionReport_GID-12025	A report that contains historical data on past pollution levels, including measurements of various pollutants such as sulfur dioxide (SO2), nitrogen dioxide (NO2), ozone (O3), fine particulate matter (PM2.5), and coarse particulate matter (PM10) recorded at a specific point in time in the past.

# Teleontology



Our Teleontology from OWL Visulization Tool

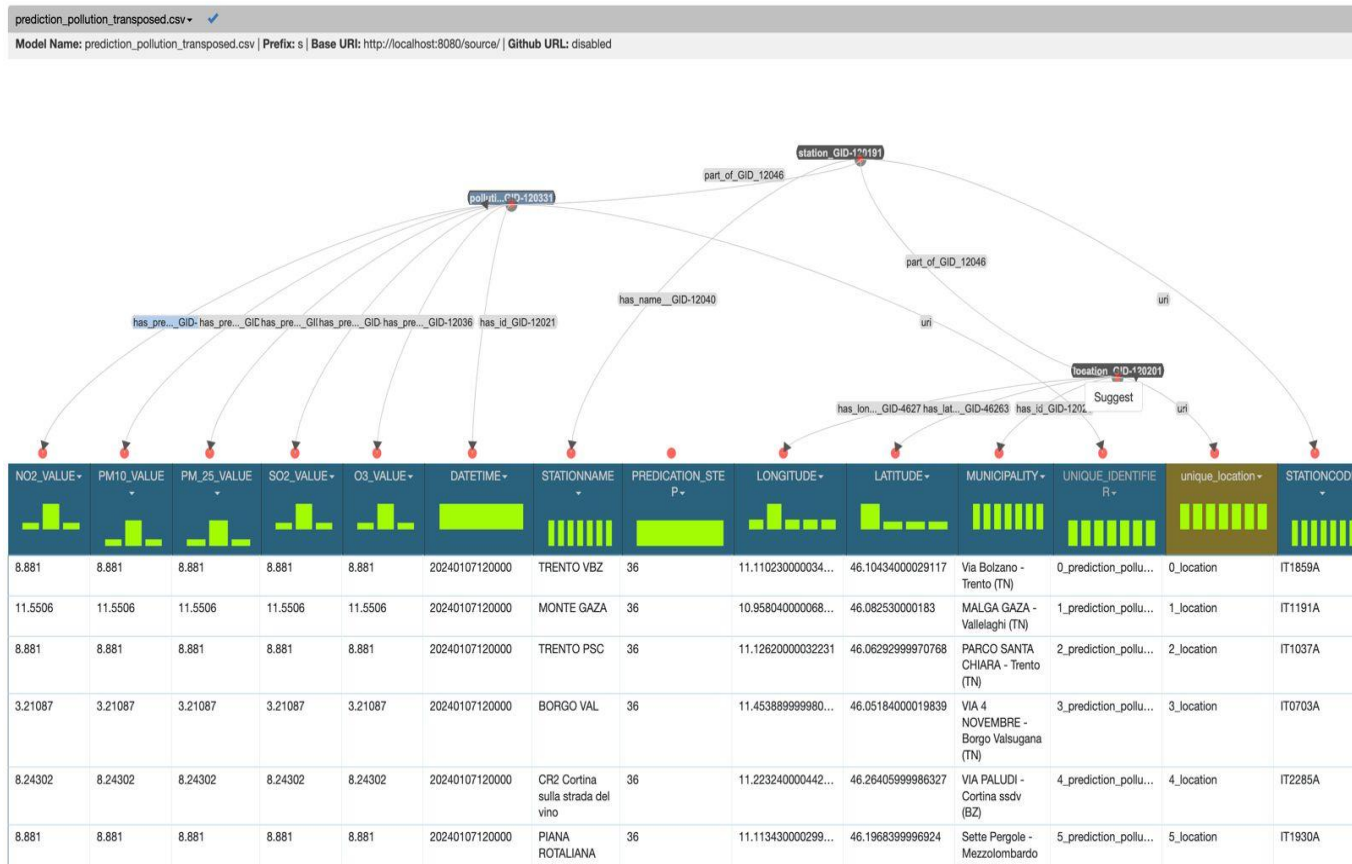
# Entity Identification and Data Merging

- **Schema Layer** : Addition of URI as Entity identifier

Concept	Identifier
location	n_location
daily pollution	n_daily_pollution
historical_pollution_report	n_historical_pollution
daily Weather	n_daily_weather

- **Data Layer** : Merging multiple dataset to reflect on one etype.
- Example: etype: Pollution prediction
  - SO2 value
  - PM10 value
  - PM25 value

# Data Mapping



Data Mapping to The Corresponding Entities

# Final Knowledge Graph

- The objective of constructing a KG encapsulate
  - The climate change in Trentino region reflects the main purpose to have
    - Weather Measurements
    - Pollution Measurements and Prediction
- This KG can be used as a tool for gaining insights into various weather and pollution prediction data, capturing, visiting interesting places trips and conducting events.
- The outcomes of our project are realistic. However, there is some limitations and always room for improvement.

# Final Knowledge Graph

- Etypes and properties coverage – Teleontology and Reference Ontology

$$COV_E(CQ_E) = \frac{|CQ_E \cap T_E|}{CQ_E} = \frac{7}{7} = 1,$$

$$COV_P(RO_P) = \frac{|RO_P \cap T_P|}{RO_P} = \frac{38}{0} = \textit{undefined},$$

$$COV_E(RO_E) = \frac{|RO_E \cap T_E|}{RO_E} = \frac{14}{607} = 0.01.$$

$$COV_P(CQ_P) = \frac{|CQ_P \cap T_P|}{CQ_P} = \frac{32}{32} = 1,$$

# Final Knowledge Graph Exploitation

**CQ1:** What is the current situation of concentration of NO2, SO2, O3, PM10 and PM2.5 in Borgo Valsugana?

```
1  PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2  PREFIX disi: <http://knowdive.disi.unitn.it/etype#>
3
4  SELECT ?no2value ?so2value ?so3value ?pm10value ?pm25value ?name ?Datetime
5  WHERE {
6      ?daily_pollution rdf:type disi:currentpollutionreport_GID-12013 .
7
8      ?daily_pollution disi:schema:has_NO2_value_GID-12015 ?no2value .
9      ?daily_pollution disi:schema:has_SO2_value_GID-12014 ?so2value .
10     ?daily_pollution disi:schema:has_O3_value_GID-12016 ?so3value .
11     ?daily_pollution disi:schema:has_PM10_value_GID-12018 ?pm10value .
12     ?daily_pollution disi:schema:has_PM25_value_GID-12017 ?pm25value .
13     ?daily_pollution disi:schema:has_datetime_GID-12024 ?Datetime .
14
15     ?daily_pollution <http://schema.org/part_of_GID_12046> ?station .
16     ?location <http://schema.org/part_of_GID_12046> ?station .
17
18     ?station disi:schema:has_name__GID-12040 ?name .
19
20     FILTER(STRSTARTS(?name, "BORGO")).
21
22 }
```



# Final Knowledge Graph Exploitation

**CQ2:** What is the prediction for the next day's noon concentration of pollutants: NO2, SO2, O3, PM10 and PM2.5 for LS1 Laives station?

```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX disi: <http://knowdive.disi.unitn.it/etype#>
3
4 SELECT ?no2value ?so2value ?o3value ?PM10value ?name
5 WHERE {
6     ?prediction_pollution rdf:type disi:pollution_prediction_GID-12033 .
7
8     ?prediction_pollution disi:schema:has_pred_NO2_value_GID-12035 ?no2value .
9     ?prediction_pollution disi:schema:has_pred_SO2_value_GID-12034 ?so2value .
10    ?prediction_pollution disi:schema:has_pred_O3_value_GID-12036 ?o3value .
11    ?prediction_pollution disi:schema:has_pred_PM10_value_GID-12038 ?PM10value .
12
13    ?prediction_pollution <http://schema.org/part_of_GID_12046> ?station .
14
15    ?station disi:schema:has_name__GID-12040 ?name .
16
17    FILTER(STRSTARTS(?name, "LS1")).
18 }
```



# Final Knowledge Graph Exploitation

**CQ3:** Query the list of historical weather values: average temperature, average pressure, max wind speed, precipitation and humidity for Trentino area.

```
1  PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2  PREFIX disi: <http://knowdive.disi.unitn.it/etype#>
3
4  SELECT ?weather_history ?temp_avg ?pressure_avg ?wind_max ?rain_precipitation ?
      humidity_perc ?datetime
5  WHERE {
6      ?weather_history rdf:type disi:history_weather_report_GID-12022 .
7
8      ?weather_history disi:schema:has_temp_avg_GID-12041 ?temp_avg .
9      ?weather_history disi:schema:has_pressure_avg_GID-12011 ?pressure_avg .
10     ?weather_history disi:schema:has_wind_max_GID-12009 ?wind_max .
11     ?weather_history disi:schema:has_rain_precipitation_GID-12012 ?rain_precipitation
      .
12     ?weather_history disi:schema:has_humidity_perc_GID-12007 ?humidity_perc .
13     ?weather_history disi:schema:has_datetime_GID-12024 ?datetime .
```

# Final Knowledge Graph Exploitation

**CQ3:** What are the cities with the lowest average values of O3 and NO2 in Trentino area?

```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
3 PREFIX disi: <http://knowdive.disi.unitn.it/etype#>
4
5 SELECT ?name (AVG(xsd:float(?o3value)) AS ?avgo3value)
6 WHERE {
7
8     ?history_pollution rdf:type disi:history_pollution_report_GID-12025 .
9     ?history_pollution disi:schema:has_O3_value_GID-12016 ?o3value .
10
11     ?history_pollution <http://schema.org/part_of_GID_12046> ?station .
12     ?station disi:schema:has_name__GID-12040 ?name .
13
14     FILTER(xsd:float(?o3value) > 0).
15 }
16
17 GROUP BY ?name
18 ORDER BY ?avgo3value
19
20 ### second query
21
22 SELECT ?name (AVG(xsd:float(?no2value)) AS ?avgNo2Value)
23 WHERE {
24
25     ?history_pollution rdf:type disi:history_pollution_report_GID-12025 .
26
27     ?history_pollution disi:schema:has_NO2_value_GID-12015 ?no2value .
28     ?history_pollution <http://schema.org/part_of_GID_12046> ?station .
29
30     ?station disi:schema:has_name__GID-12040 ?name .
31
32     FILTER(xsd:float(?no2value) > 0).
33 }
34
35 GROUP BY ?name
36 ORDER BY ?avgNo2Value
```

# Conclusion

- Successful construction of Knowledge Graph
- Incorporation of pollution and weather data
- Necessary data transformation, mappings and annotation
- Successful modeling of data *consumer* & *producer* role
- Answered competency questions
- KG exploit and data queried via SPARQL queries

# Open issues

- Data in KG are static
- Pollution data without proper coverage (7 pollution stations)
- Better incorporation of weather data
- Improvement of pollution history data ~ payed data sources?
- Possible better incorporation of referenced ontologies

**Thank you for your attention!**

# Resources

- <https://atmosphere.copernicus.eu/european-air-quality-forecast-plots>
- [https://jclock98.github.io/Weather\\_Trentino/](https://jclock98.github.io/Weather_Trentino/)
- <https://www.eea.europa.eu/data-and-maps/explore-interactive-maps/up-to-date-air-quality-data>
- <https://www.meteotrentino.it/index.html#!/home>
- [https://link.springer.com/content/pdf/10.1007/978-1-4419-0221-4\\_17.pdf](https://link.springer.com/content/pdf/10.1007/978-1-4419-0221-4_17.pdf)