OPEN DATA AND KNOWLEDGE GRAPHS

- Carlos de la Cruz Ayala Vargas
- Elías Herrero Lázaro
- Rafael Orihuela Brindis
- Pedro Pedregal Pérez

Objetive

- We aim to show the measurements of air quality within a spatial context, using a Madrid City map and taking advantage of the information.
  - A user should be able to:
    - Select to see data from all stations or from only one.
    - Select whether data should belong to a specific day (showing that day's value) or to a time window (showing the average value for that period).
    - Select to see data pertaining to a specific parameter; only one parameter at a time.

Data set

- The Data Set was obtained from the Madrid's City Hall web page.
- The data can be reused under certain guidelines that include citing the source, not distorting the information, and prohibiting the re-identification of personal data
- The information shown is about the city's air quality, measured every day in different stations located in Madrid.
- The observations were recorded from January 1st 2023 to August 20th 2024 and are displayed as in the following table:

Provincia	Municipio	Estacion	Magnitud	Punto_muestreo	Año	Mes	D01	V01	D02	V02
28	79	4	1	28079004_1_38	2023	1	18	V	20	V

Data set

- Additionally, we have another data set that explains the stations information.
- The Data Set was obtained from the Madrid's City Hall web page.
- The data can be reused under certain guidelines that include citing the source, not distorting the information, and prohibiting the re-identification of personal data
- The observations were recorded from January 1st 2023 to August 20th 2024 and are displayed as in the following table:

Codigo	Corto	Estacion	Longitud	Latitud	Altitud	NO2	SO2	PM10	Via_Clase	Via_PAR	Via_Nombre
28079004	4	Plaza de España	3°42'43.91"O	40°25'25.98"N	637	X	X		PLAZA	DE	ESPAÑA

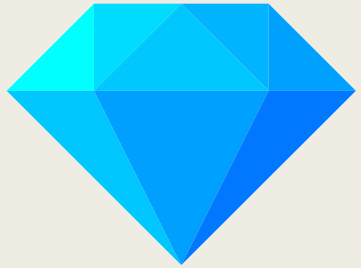
Ontology

- Our ontology “mAIRres” was created based on Ontologys already developed by the World Wide Web Consortium (W3C) and the Wikipedia Foundation, as:
  - SOSA (Sensor, Observation, Sample, and Actuator)
  - SSN (Semantic Sensor Network)
  - RDF (Resource Description Framework)
  - Wikidata (wd)

- The ontology is designed to incorporate standards concepts widely used in environmental monitoring.

Data processing

- We used **OpenRefine** to adecuate our original data set to be able to analize it
  - *The measures dataset had **one row with 30 values.** Changed to **1 row per value** (valid columns were treated)*
  - ***Representative magnitudes***
  - *Date joined in one column **(yyyy-mm-dd)***

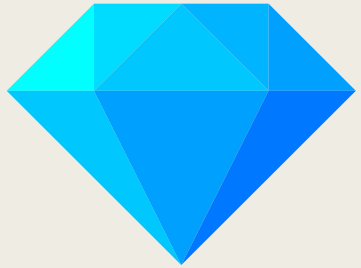


Punto_muestreo	Año	Mes	D01	V01	D02	V02	D03	V03
28079004_1_38	2023	1	18	V	20	V	14	F

Magnitud	Punto_muestreo	ID_TIEMPO	MEDIDA
SO2	28079004_1_38	2023-01-01	20

Data processing

- The stations dataset had **many redundant columns**
  - Deleted “Via” and “Corto” columns
  - **Coordinates** to decimal number
  - **Measure type** to true or false



Codigo	Corto	Estacion	Longitud	Latitud	Altitud	NO2	SO2	PM10	Via_Clase	Via_PAR	Via_Nombre
28079004	4	Plaza de España	3°42'43.91"O	40°25'25.98"N	637	X	X		PLAZA	DE	ESPAÑA

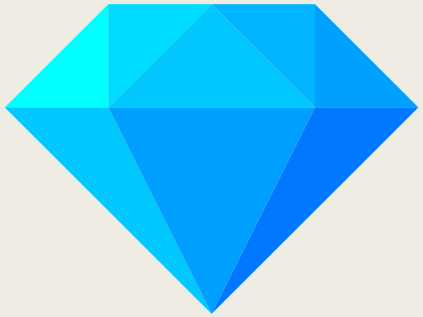
Codigo	Estacion	Longitud	Latitud	Altitud	NO2	SO2	PM10
28079004	Plaza de España	-3.71	40.42	637	true	true	false

RDF Generation

- We build our mappings in **YARRRML**
- So many problems with **Matey**. Changed to **Morph**
- The ontology generation process has been automated with a **Python script** (cleaning, mapping and serializing)

morph

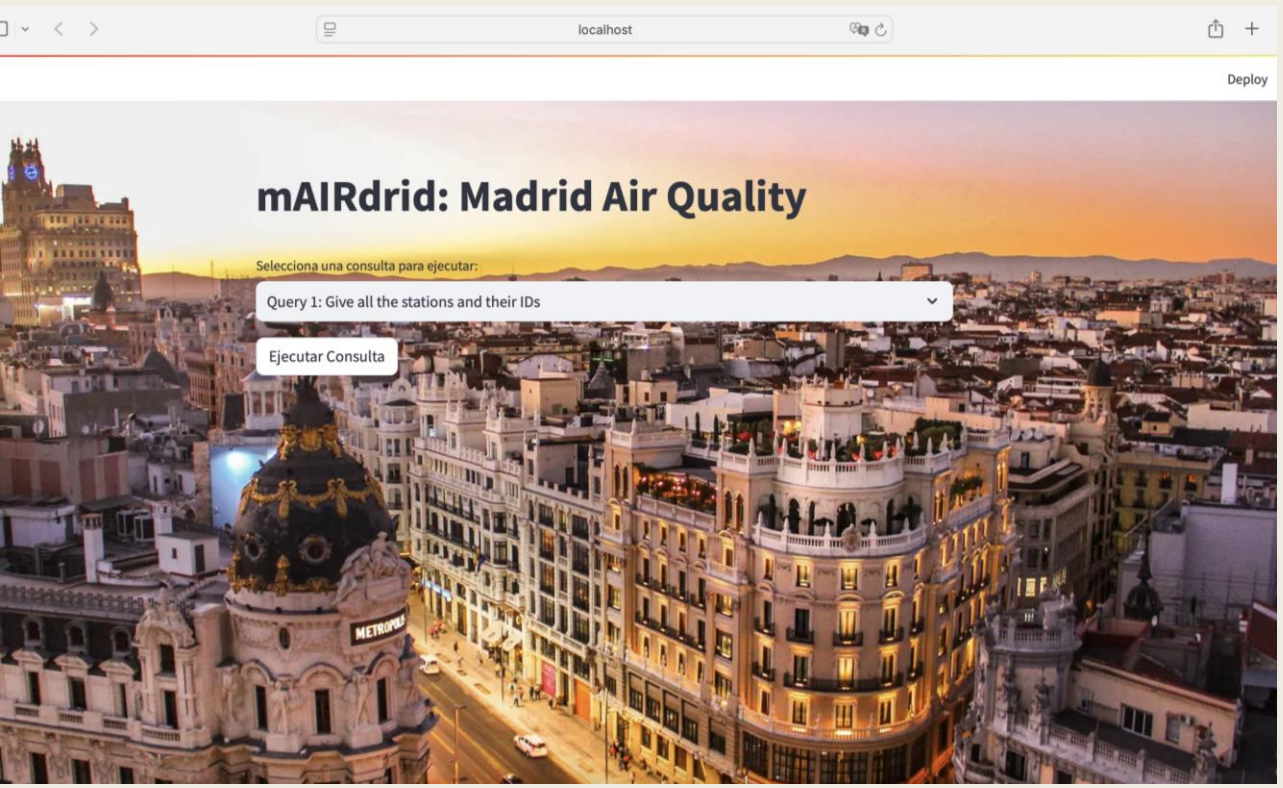
Data Linking

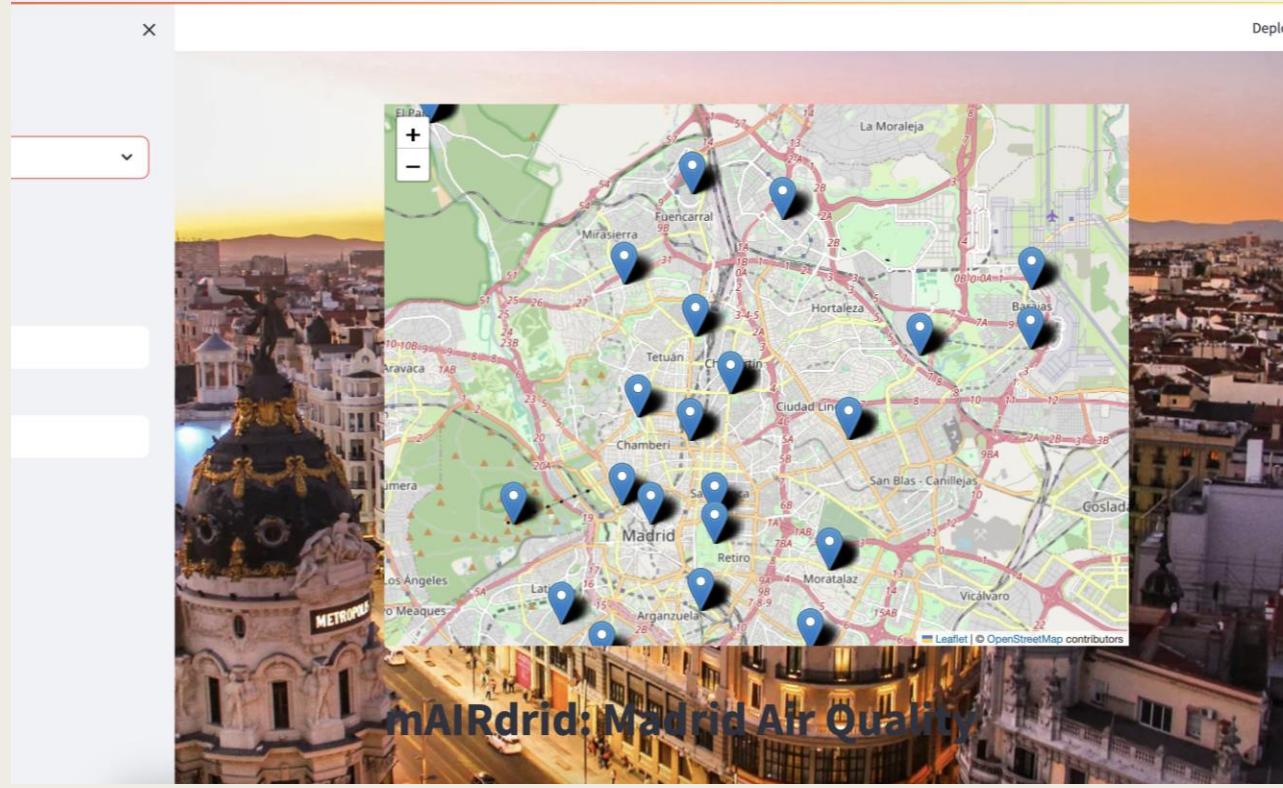


- We build our mappings in **YARRRML**
- So many problems with **Matey**. Changed to Morph
- The ontology generation process has been automated with a Python script (cleaning, mapping and serializing)

Linked Data Publishing

- We used **GraphDB** to create a virtual server, where we published our final ontology.
- This allowed us to accomplish three main goals:
  - Upload and store our final ontology.
  - Consult, using a SPARQL query, the information uploaded.
  - Published the final ontology and create an endpoint SPARQL for anyone who wants to access to the server.





App

- We connected our app code with the URL endpoint to remotely access to the data base, through **SPARQLWrapper**.
- Using the python's library **Streamlit** we created an interactive web application which allows the user to search specific information about Madrid's air quality.