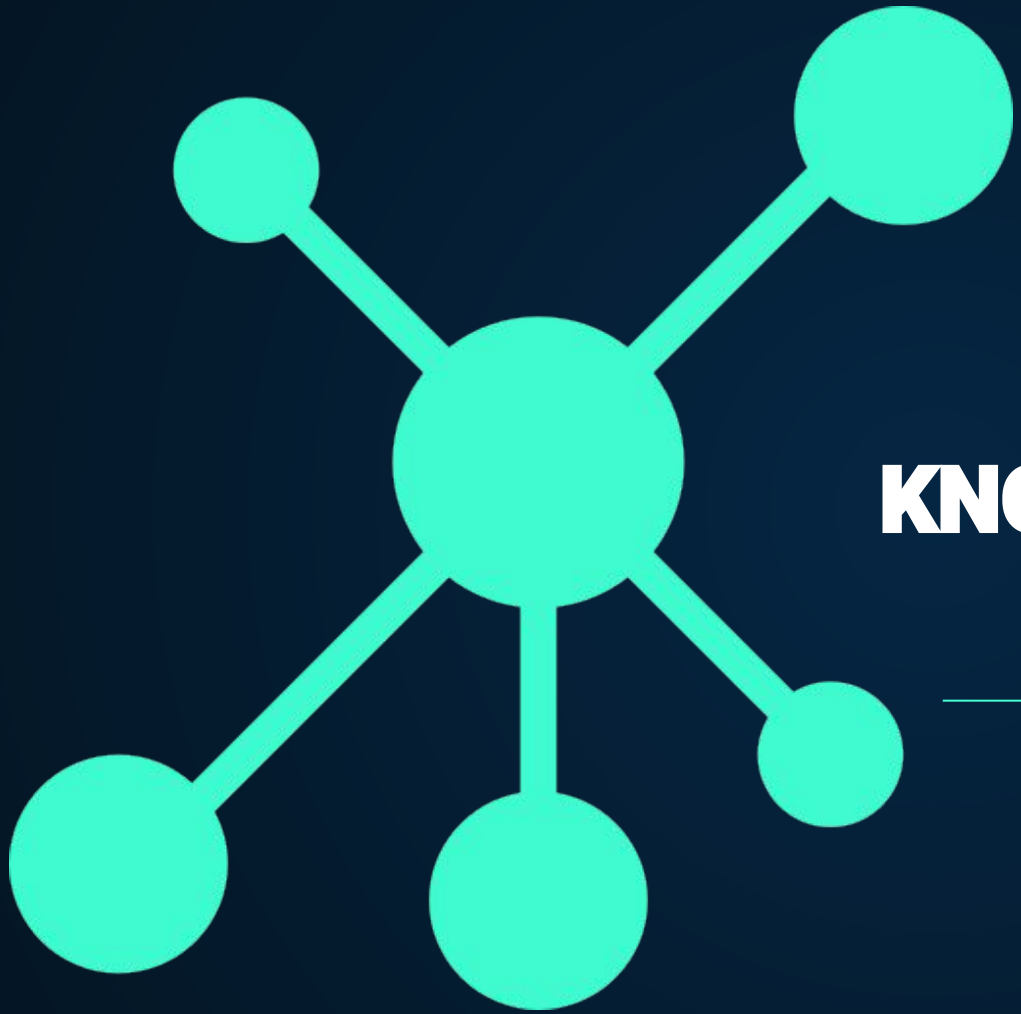


OPEN DATA AND KNOWLEDGE GRAPHS PROJECT

Group 7

AGENDA

01 Goal

02 Dataset

03 Resource Naming Strategy

04 Ontology

05 Translation to RDF

06 Reconciliation

07 Web application

01 GOAL

SMART CITIES → AIR QUALITY

CITY → MADRID

APPLICATION → WEB APP

02 DATASET

```
In [7]: dataset1.head()
```

Out[7]:

	measurement_date	measurement_id	BEN	CH4	CO	EBE	NMHC	NO	NO_2	NOx	...	PM25	SO_2	TCH	TOL	station_id	station_name	station_addı
0	2018-03-01T01:00:00Z	measurement_0	NaN	NaN	0.3	NaN	NaN	1.0	29.0	31.0	...	NaN	2.0	NaN	NaN	28079004	Pza_de_Espana	Plaza de Esp
1	2018-03-01T01:00:00Z	measurement_1	0.5	1.39	0.3	0.2	0.02	6.0	40.0	49.0	...	4.0	3.0	1.41	0.8	28079008	Escuelas_Aguirre	Entre C/ Alca C/ O' Do
2	2018-03-01T01:00:00Z	measurement_2	0.4	NaN	NaN	0.2	NaN	4.0	41.0	47.0	...	NaN	NaN	NaN	1.1	28079011	Avda_Ramon_y_Cajal	Avda. Ramo Cajal esc Príncipe Verg
3	2018-03-01T01:00:00Z	measurement_3	NaN	NaN	0.3	NaN	NaN	1.0	35.0	37.0	...	NaN	NaN	NaN	NaN	28079016	Arturo_Soria	C/ Arturo S esc Vizconde de As
4	2018-03-01T01:00:00Z	measurement_4	NaN	NaN	NaN	NaN	NaN	1.0	27.0	29.0	...	NaN	3.0	NaN	NaN	28079017	Villaverde	C/. J Peña

5 rows × 22 columns

03 RESOURCE NAMING STRATEGY

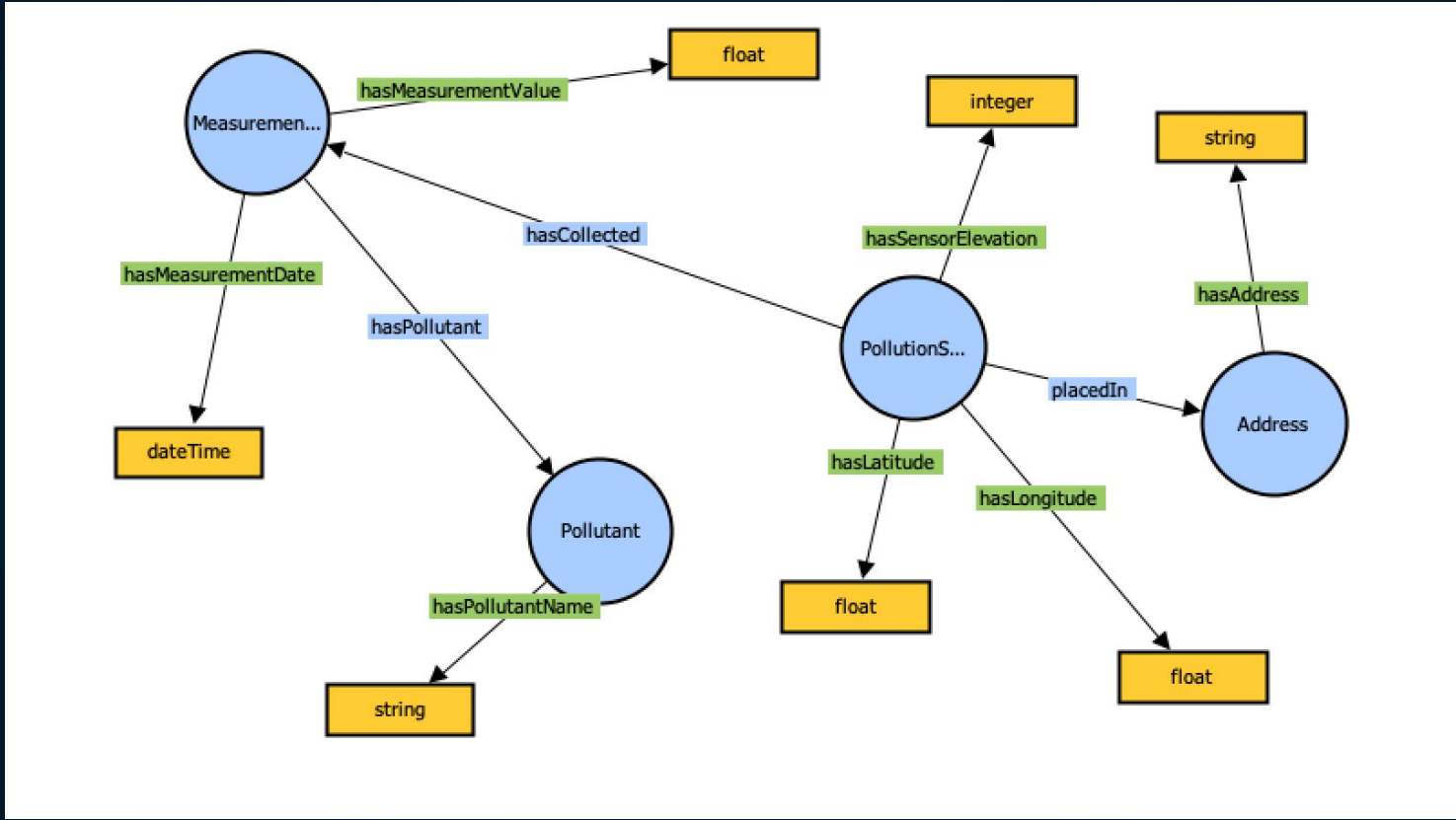
ONTOLOGY → **hash URI #**

http://www.airqualitymadrid.es/ontologies/<ontologyName>#<ClassName>

RESOURCE → **slash URI /**

http://www.airqualitymadrid.es/resource/<className>/<identifier>

04 ONTOLOGY


float
hasMeasurementValue
Measuremen...
integer
string
hasCollected
hasSensorElevation
hasMeasurementDate
hasAddress
hasPollutant
PollutionS...
placedIn
Address
dateTime
hasLatitude
hasLongitude
Pollutant
float
hasPollutantName
float
string

05 TRANSLATION TO RDF

Problem

In [7]:
```
dataset1.head()
```

Out[7]:

	measurement_date	measurement_id	BEN	CH4	CO	EBE	NMHC	NO	NO_2	NOx	...	PM25	SO_2	TCH	TOL	station_id	station_name	station_addr
0	2018-03-01T01:00:00Z	measurement_0	NaN	NaN	0.3	NaN	NaN	1.0	29.0	31.0	...	NaN	2.0	NaN	NaN	28079004	Pza_de_Espana	Plaza de Esp
1	2018-03-01T01:00:00Z	measurement_1	0.5	1.39	0.3	0.2	0.02	6.0	40.0	49.0	...	4.0	3.0	1.41	0.8	28079008	Escuelas_Aguirre	Entre C/ Alca C/ O' Do
2	2018-03-01T01:00:00Z	measurement_2	0.4	NaN	NaN	0.2	NaN	4.0	41.0	47.0	...	NaN	NaN	NaN	1.1	28079011	Avda_Ramon_y_Cajal	Avda. Ramo Cajal esq Príncipe Verg
3	2018-03-01T01:00:00Z	measurement_3	NaN	NaN	0.3	NaN	NaN	1.0	35.0	37.0	...	NaN	NaN	NaN	NaN	28079016	Arturo_Soria	C/ Arturo So esq Vizconde de As
4	2018-03-01T01:00:00Z	measurement_4	NaN	NaN	NaN	NaN	NaN	1.0	27.0	29.0	...	NaN	3.0	NaN	NaN	28079017	Villaverde	C/. J Peña

5 rows × 22 columns

05 TRANSLATION TO RDF

Solution

```
In [17]: dataset2.head()
```

Out[17]:

	measurement_id	date	station	station_name	station_address	station_lon	station_lat	station_elevation	pollutant	pollutant_name	value
0	Measurement_1	2018-01-01 01:00:00	28079008	Escuelas Aguirre	Entre C/ Alcalá y C/ O' Donell	-3.682319	40.421564	67	BEN	Benzene	0.4
1	Measurement_2	2018-01-01 01:00:00	28079011	Avda. Ramón y Cajal	Avda. Ramón y Cajal esq. C/ Príncipe de Vergara	-3.677356	40.451475	7	BEN	Benzene	0.2
2	Measurement_3	2018-01-01 01:00:00	28079018	Farolillo	Calle Farolillo - C/Ervigio	-3.731853	40.394781	6	BEN	Benzene	1.0
3	Measurement_4	2018-01-01 01:00:00	28079024	Casa de Campo	Casa de Campo (Terminal del Teleférico)	-3.747347	40.419356	64	BEN	Benzene	0.3
4	Measurement_5	2018-01-01 01:00:00	28079038	Cuatro Caminos	Avda. Pablo Iglesias esq. C/ Marqués de Lema	-3.707128	40.445544	698	BEN	Benzene	0.2

06 RECONCILIATION

Problem

street ?????

```
In [17]: dataset2.head()
```

Out[17]:

	measurement_id	date	station	station_name	station_address	station_lon	station_lat	station_elevation	pollutant	pollutant_name	value
0	Measurement_1	2018-01-01 01:00:00	28079008	Escuelas Aguirre	Entre C/ Alcalá y C/ O' Donell	-3.682319	40.421564	670	BEN	Benzene	0.4
1	Measurement_2	2018-01-01 01:00:00	28079011	Avda. Ramón y Cajal	Avda. Ramón y Cajal esq. C/ Príncipe de Vergara	-3.677356	40.451475	708	BEN	Benzene	0.2
2	Measurement_3	2018-01-01 01:00:00	28079018	Farolillo	Calle Farolillo - C/Ervigio	-3.731853	40.394781	630	BEN	Benzene	1.0
3	Measurement_4	2018-01-01 01:00:00	28079024	Casa de Campo	Casa de Campo (Terminal del Teleférico)	-3.747347	40.419356	642	BEN	Benzene	0.3
4	Measurement_5	2018-01-01 01:00:00	28079038	Cuatro Caminos	Avda. Pablo Iglesias esq. C/ Marqués de Lema	-3.707128	40.445544	698	BEN	Benzene	0.2

corner ?????

square ?????

06 RECONCILIATION

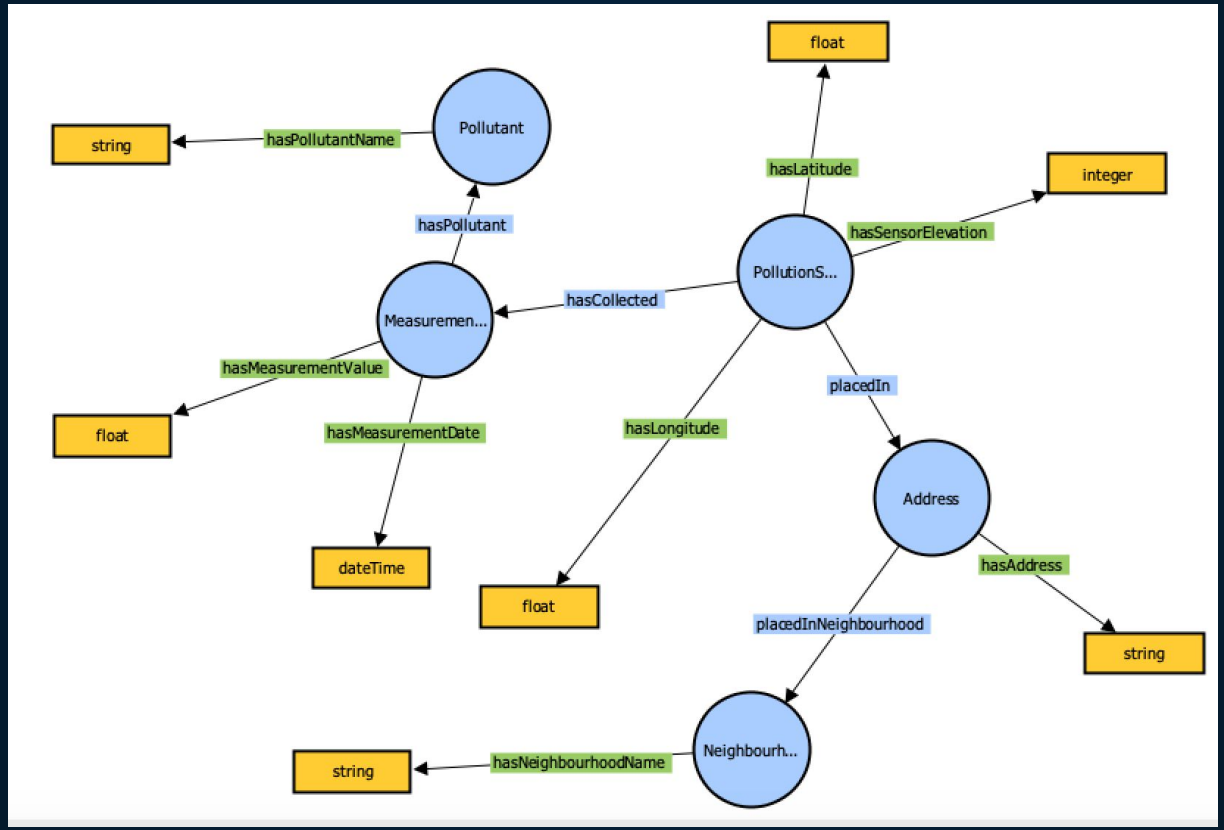
Solution

```
In [15]: dataset3.head()
```

Out[15]:

measurement_id	date	station	station_name	station_address	neighbourhood_name	station_lon	station_lat	station_elevation	pollutant	pollutant_name	value
Measurement_1	2018-01-01 01:00:00	28079008	Escuelas Aguirre	Entre C/ Alcalá y C/ O' Donell	Recoletas	-3.682319	40.421564	670	BEN	Benzene	0.4
Measurement_2	2018-01-01 01:00:00	28079011	Avda. Ramón y Cajal	Avda. Ramón y Cajal esq. C/ Príncipe de Vergara	Hispanoamerica	-3.677356	40.451475	708	BEN	Benzene	0.2
Measurement_3	2018-01-01 01:00:00	28079018	Farolillo	Calle Farolillo - C/Erv igio	Isidro	-3.731853	40.394781	630	BEN	Benzene	1.0
Measurement_4	2018-01-01 01:00:00	28079024	Casa de Campo	Casa de Campo (Terminal del Teleférico)	Case_de_Campo	-3.747347	40.419356	642	BEN	Benzene	0.3
Measurement_5	2018-01-01 01:00:00	28079038	Cuatro Caminos	Avda. Pablo Iglesias esq. C/ Marqués de Lema	Cuatro_Caminos	-3.707128	40.445544	698	BEN	Benzene	0.2

FINAL ONTOLOGY



07 WEB APPLICATION

Query using rdflib

```
def get_stations(self):
    res = self.data.query("""
        PREFIX aqm: <http://www.airqualitymadrid.es/ontologies/airqualitymadrid#>
        select ?station_id ?station_name ?latitude ?longitude ?address ?elevation ?neighbourhood_name
        WHERE {
            ?station_id a aqm:PollutionSensor ;
                aqm:placedIn ?station_name ;
                aqm:hasSensorElevation ?elevation ;
                aqm:hasLatitude ?latitude ;
                aqm:hasLongitude ?longitude .
            ?station_name aqm:hasAddress ?address ;
                aqm:placedInNeighbourhood ?neigh_entity .
            ?neigh_entity aqm:hasNeighbourhoodName ?neighbourhood_name .
        }
        """)
```

07 WEB APPLICATION

Query using rdflib

```
res = self.data.query("""
    PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
    PREFIX sens: <http://www.airqualitymadrid.es/resource/pollution_sensor/>
    PREFIX aqm: <http://www.airqualitymadrid.es/ontologies/airqualitymadrid#>
    select ?pollutant_id ?pollutant_name (MIN(xsd:float(xsd:string(?value))) as ?min)
    (avg(xsd:float(xsd:string(?value))) as ?avg)
    (MAX(xsd:float(xsd:string(?value))) as ?max)
    WHERE {
        sens:""" + station_id + """ a aqm:PollutionSensor ;
            aqm:hasCollected ?measurement .
        ?measurement a aqm:Measurement ;
            aqm:hasPollutant ?pollutant_id ;
            aqm:hasMeasurementDate ?date ;
            aqm:hasMeasurementValue ?value .
        ?pollutant_id a aqm:Pollutant ;
            aqm:hasPollutantName ?pollutant_name .
        FILTER(?date >= \""""" + start_time + """\"^^xsd:dateTime)
        FILTER(?date <= \""""" + end_time + """\"^^xsd:dateTime)
    }
    GROUP BY ?pollutant_id
    """)
```

DEMO

28054

BEN (Benzene)

avg: 0.30 / max: 0.40 / min: 0.30

3DAIRQ

Mouse over the 3D model shows the different city areas and displays the air quality data for each one (each area is covered by one measure device)

You can choose the date by using the bottom slider