Air Pollution in Madrid

Semantic Web
Group 07

Group participants

Luciano García Giordano

José Luis Del Pino García

Jhon Toledo Barreto

Jesús Hormigos Villalba

Daniel José Argota García

Datasets selected

Stations.csv

Column	Type	Comments / Range (Rounded)	Problems
id	number	Unique identifier of the station.	
name	name	Name of the station	
address	name	Address of the station	
lon	number	Longitud of the station, Range: (-3.77, 3.58)	
lat	number	Latitude of the station, Range: (40.35, 40.52)	
elevation	number	Elevation of the station, Range: (590, 730)	

Datasets selected

Yearly data files

Column	Type	Comments / Range (Rounded)	Problems
date	date	Unique identifier of the station.	
BEN	number	Benzene, Range: (0,20)	Blank values
CH4	number	Methane, Range: (1,3.4)	Blank values
CO	number	Carbon Monoxide, Range: (0.1,4.91)	Blank values
EBE	number	Ethylbenzene, Range: (0,39)	Blank values
NMHC	number	Non-methane hydrocarbons, Range: (0,4.41)	Blank values
NO	number	Nitrogen monoxide, Range: (0,980)	Blank values
NO_2	number	Nitrogen dioxide, Range: (0,350)	Blank values
NOx	number	Nitrogen oxides, Range: (0,1800)	Blank values
O_3	number	Ozone, Range: (0,200)	Blank values
PM10	number	Particles <10 µm, Range: (0,310)	Blank values
PM25	number	Particles <25 µm, Range: (0,96)	Blank values
SO_2	number	Sulfur Dioxide, Range: (1,99)	Blank values
TCH	number	Total hydrocarbons(hexane), Range: (0.15,6.68)	Blank values
TOL	number	Toluene, Range: (0,85)	Blank values
station	number	Unique identifier of the station.	

Datasets selected

R1. Data are in one of the selected smart city domains. ✓

R2. Data are available as a CSV file. ✓

R3. Data have an open license so they can be published. ✓

R4. Data can be easily linked with generic real-world entities (e.g., locations). ✓

Naming Strategy

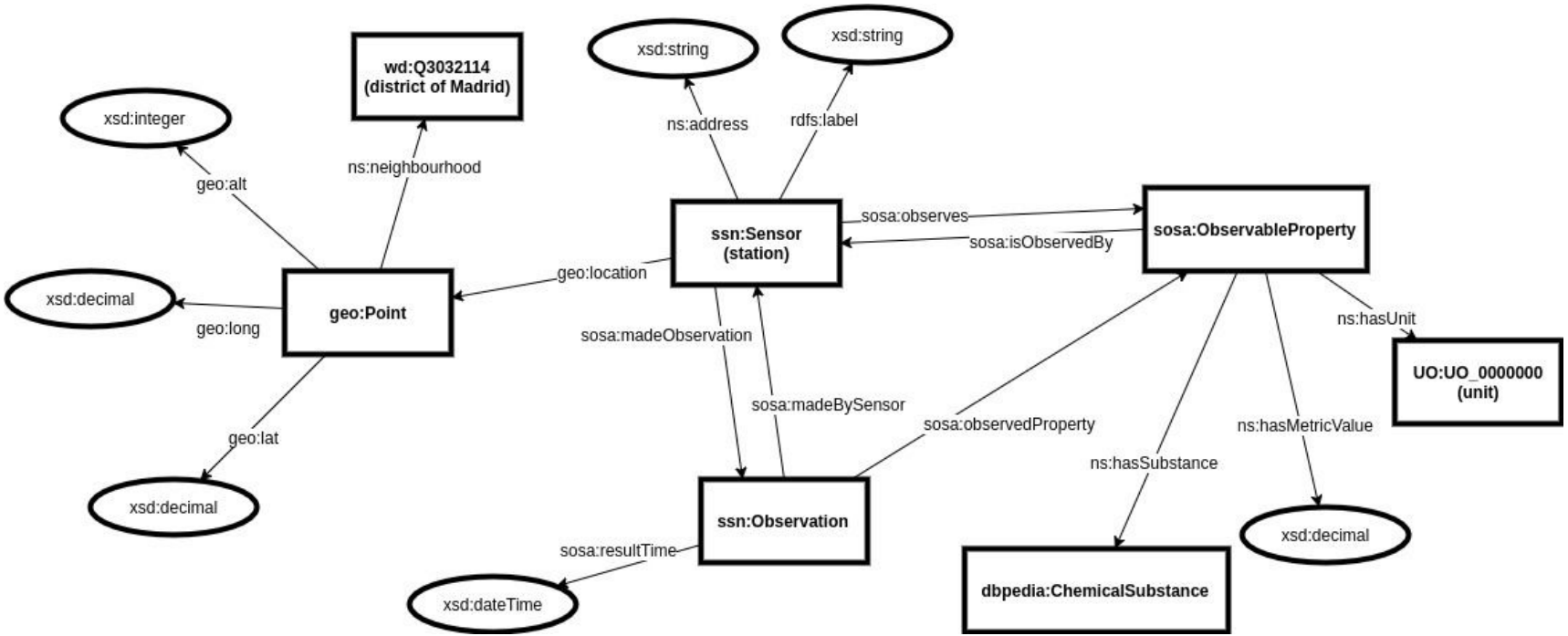
The data to be generated will follow the given patterns:

1. Stations: http://www.group07.linkeddata.org/individual/station#station{id}
2. Measurements: http://www.group07.linkeddata.org/individual/measurement#measurement{id_of_station}
3. Substance Measurements: http://www.group07.linkeddata.org/individual/substanceMeasurement#substanceMeasurement{id_of_station}

CSV to RDF / LODRefine (Filtering and linking)

- Stations
  - Linking neighbourhoods
- Measurements
  - Removal of missing values
  - Transpose for an easier and more straightforward linking

Ontology


wd:Q3032114
(district of Madrid)
xsd:string
xsd:string
xsd:integer
ns:address
rdfs:label
ns:neighbourhood
geo:alt
ssn:Sensor
(station)
sosa:observes
sosa:isObservedBy
sosa:ObservableProperty
xsd:decimal
geo:Point
geo:location
geo:long
sosa:madeObservation
ns:hasUnit
UO:UO_0000000
(unit)
sosa:madeBySensor
sosa:observedProperty
ns:hasMetricValue
geo:lat
ns:hasSubstance
xsd:decimal
ssn:Observation
xsd:decimal
sosa:resultTime
dbpedia:ChemicalSubstance
xsd:dateTime

Final implementation

- Virtuoso as SPARQL endpoint for data
- HTML + JS front end

Demo...

The end.