BARCELONA
LINKED DATA PROJECT

SEMANTIC WEB, LINKED DATA AND KNOWLEDGE GRAPHS – GROUP 08

Zhen Jiang • Huangkai Chen • Laura Nguyen • Giovana Yixinqiu • Hugo Núñez Búa

17 November 2025

HANDS-ON 1 (GOAL + REQUIREMENTS)

Goal:
Find datasets we will use for all later assignments (RDF, linking, publishing)
and define the application we want to build.

Dataset Requirements (R1–R6):
✓ R1 Smart City domain → Barcelona
✓ R2 Available as CSV → OpenDataBCN
✓ R3 Open license (CC BY 4.0)
✓ R4 Linkable to real-world entities
(✓ R5 Documentation exists — optional)
(✓ R6 Multiple data sources — optional)

HANDS-ON 1 (SELECTED DATASETS)

We selected three datasets from Open Data BCN:

Air Quality Data:

- Data of the contaminants measured in the stations of the city of Barcelona

Traffic Road Segments

- Traffic state information in the city of Barcelona by sections.

Bicycle Parking

- Parking service for bikes in the city of Barcelona

HANDS-ON 1 (APP DEFINITION)

Application Name
BCN Mobility & Air Explorer

Goal
Show and compare different smart city datasets on a map
(air quality, traffic, bicycle parking).

Main Features

- Show air quality stations + pollutants
- Display traffic congestion
- Show bike parking by district
- Filter by district or pollutant type

HANDS-ON 2 (DATASET ANALYSIS)

Data Analysis

- Ranges, missing values, inconsistent formats
- Different structures between datasets
- Need for normalization (dates, coordinates, pollutant IDs)

License Analysis

- Publisher: Barcelona City Council
- License: CC BY 4.0 → compatible with Linked Data
- We reused CC BY 4.0 for our generated RDF

Dataset	Main columns	Notes
Air quality	station_id, station_name, district, long/lat, pollutant, value, datetime	Numeric values for pollutant concentration; time-series data.
Traffic	section_id, street, start and end points, district, intensity	Mobility congestion with its start and end points.
Bike parking	id, address, district, lat, lon, type	Static dataset with spatial coordinates.

HANDS-ON 2 (RESOURCE NAMING STRATEGY)

We defined a consistent URI pattern:

Base Domain:
http://bcn-mobility-air.org/

Ontology Namespace:
http://bcn-mobility-air.org/ontology/

Data Namespace:
http://bcn-mobility-air.org/data/

Resource Patterns:

- /AirQualityStation/{id}
- /BicycleParking/{id}
- /RoadSegment/{id}

URIs are stable, readable, and dereferenceable.

The following URI structure will be used for both the ontology and the generated RDF data:

Type	Base URI pattern	Example
Ontology	`http://bcn-mobility-air.org/ontology/`	`http://bcn-mobility-air.org/ontology/`
Resources	`http://bcn-mobility-air.org/data/{class}/{identifier}`	`http://bcn-mobility-air.org/resource/Station/1`

- **Domain:** bcn-mobility-air.org
- **Ontology pattern:** slash URIs (`/`)
- **Data pattern:** slash URIs (`/`)
- **Content negotiation:** `.html` for human view, `.ttl` for RDF data.

HANDS-ON 2 (ONTOLOGY)

Main classes

- AirQualityStation
- AirQualityReading
- Pollutant
- BicycleParking
- RoadSegment
- CongestionMeasurement
- District
- Neighborhood

Main properties

- isLocatedInDistrict
- isLocatedInNeighborhood
- hasAirQualityReading
- measuresPollutant
- hasCongestionMeasurement
- startPoint / endPoint

Ontology implemented in Turtle: ontology.ttl.

HANDS-ON 3 (OPENREFINE CLEANING)

Tasks Performed

- Loaded CSV files into OpenRefine
- Removed empty/duplicate rows
- Standardized date/time formats
- Fixed datetime formats
- Normalized pollutant labels
- Add necessary information about the IDs:
  - AirStation ID: station and pollutant names
  - Road ID: start and end points names

Deliverables:

- *-updated.csv
- *-operations.json

HANDS-ON 4 (RML TO RDF)

RDF Generation Approach

1. RML mappings for each dataset
2. Mapping rules for classes, URIs, properties
3. Used RMLMapper to generate RDF in Turtle syntax

Deliverables:

- mappings/*.rml.ttl
- rdf/*.ttl
- queries.sparql

Outcome: First full RDF version of all datasets.

HANDS-ON 5 (LINKING WITH WIKIDATA)

Identified linkable entities

- Districts
- Neighborhoods
- Pollutants
- Streets

Reconciliation:

- Using OpenRefine Wikidata reconciliation
- Added owl:sameAs links
- All datasets enriched with external semantics

Deliverables:

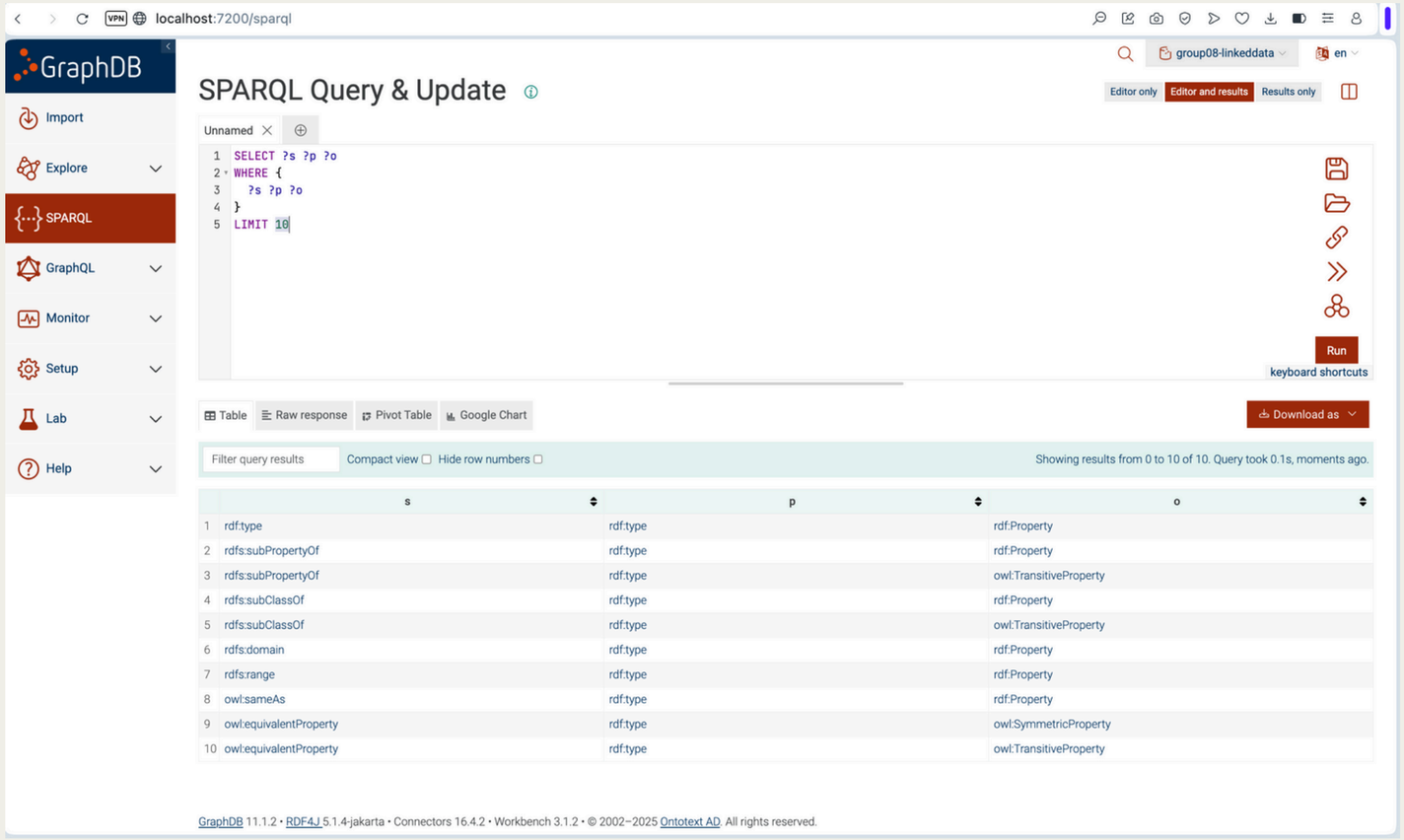
- *-with-links.csv
- *-with-links.json
- *-with-links.rml.ttl
- *-with-links.ttl
- queries-with-links.sparql

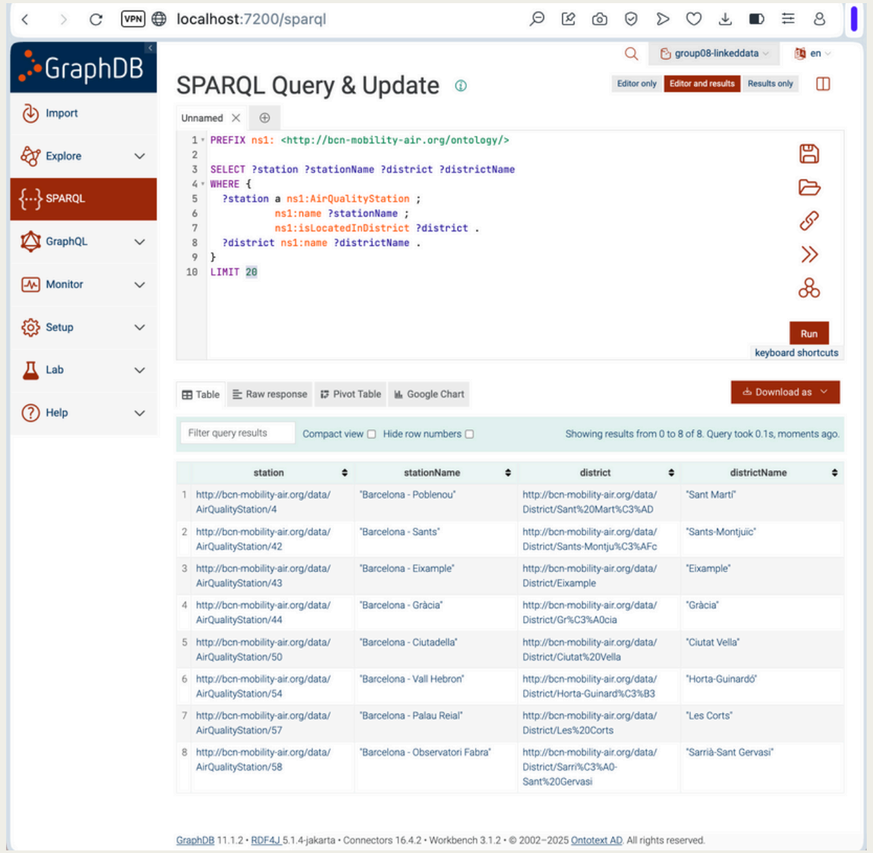
HANDS-ON (PUBLISHING IN GRAPHDB)

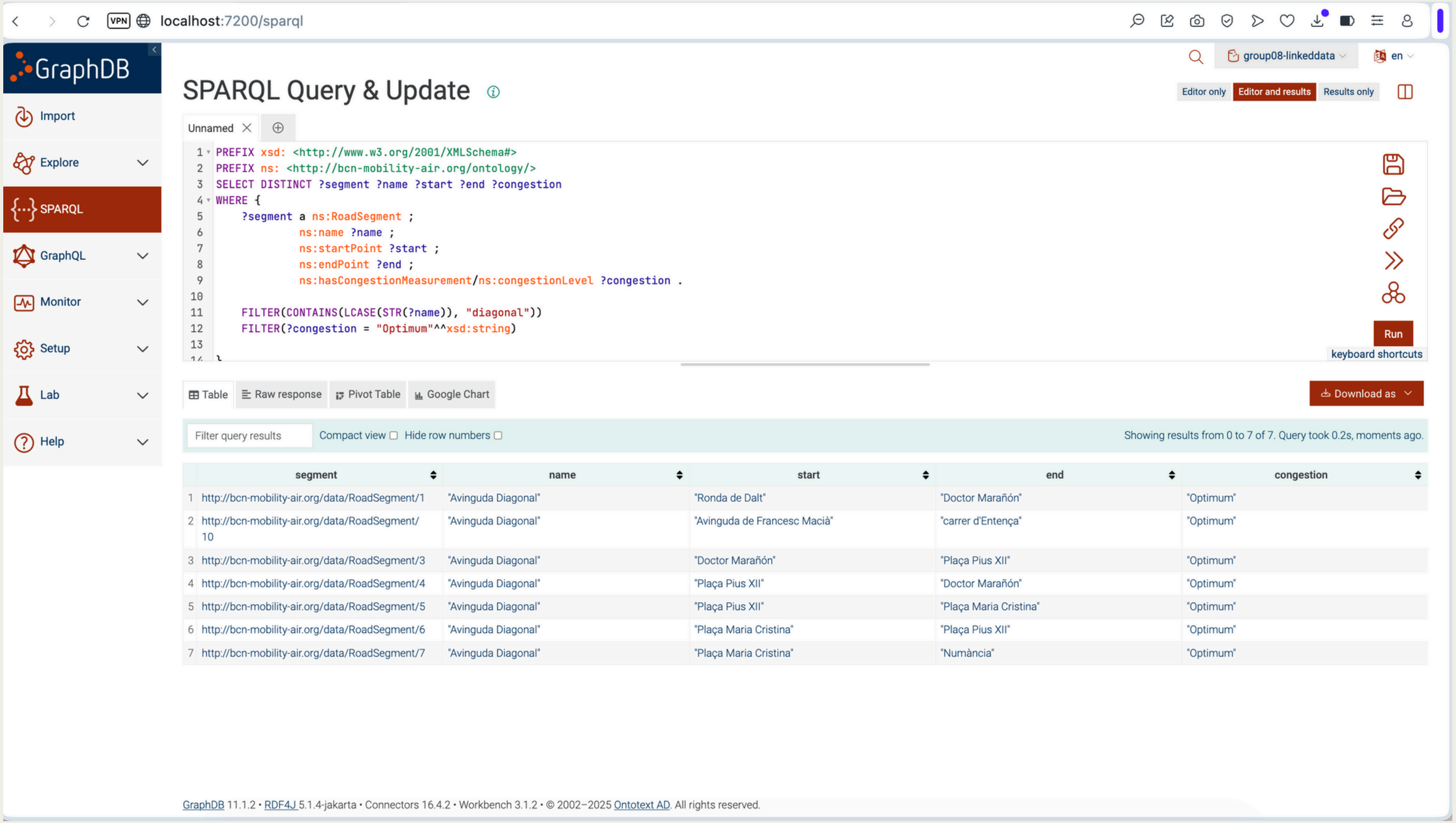
Steps

1. Created repository: group08-linkeddata
2. Imported:
    - output-with-links.ttl
    - ontology.ttl
3. Verified triples using Explore tab
4. Ran SPARQL queries to validate graph

SPARQL EXAMPLES






localhost:7200/sparql
GraphDB
Import
Explore
SPARQL
GraphQL
Monitor
Setup
Lab
Help
group08-linkeddata
en
SPARQL Query & Update
Editor only
Editor and results
Results only
Unnamed
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ns: <http://bcn-mobility-air.org/ontology/>
SELECT DISTINCT ?segment ?name ?start ?end ?congestion
WHERE {
?segment a ns:RoadSegment ;
ns:name ?name ;
ns:startPoint ?start ;
ns:endPoint ?end ;
ns:hasCongestionMeasurement/ns:congestionLevel ?congestion .
FILTER(CONTAINS(LCASE(STR(?name)), "diagonal"))
FILTER(?congestion = "Optimum"^^xsd:string)
}
Run
keyboard shortcuts
Table
Raw response
Pivot Table
Google Chart
Download as
Filter query results
Compact view
Hide row numbers
Showing results from 0 to 7 of 7. Query took 0.2s, moments ago.
segment
name
start
end
congestion
1 http://bcn-mobility-air.org/data/RoadSegment/1 "Avinguda Diagonal" "Ronda de Dalt" "Doctor Marañón" "Optimum"
2 http://bcn-mobility-air.org/data/RoadSegment/10 "Avinguda Diagonal" "Avinguda de Francesc Macià" "carrer d'Entença" "Optimum"
3 http://bcn-mobility-air.org/data/RoadSegment/3 "Avinguda Diagonal" "Doctor Marañón" "Plaça Pius XII" "Optimum"
4 http://bcn-mobility-air.org/data/RoadSegment/4 "Avinguda Diagonal" "Plaça Pius XII" "Doctor Marañón" "Optimum"
5 http://bcn-mobility-air.org/data/RoadSegment/5 "Avinguda Diagonal" "Plaça Pius XII" "Plaça Maria Cristina" "Optimum"
6 http://bcn-mobility-air.org/data/RoadSegment/6 "Avinguda Diagonal" "Plaça Maria Cristina" "Plaça Pius XII" "Optimum"
7 http://bcn-mobility-air.org/data/RoadSegment/7 "Avinguda Diagonal" "Plaça Maria Cristina" "Numància" "Optimum"
GraphDB 11.1.2 • RDF4J 5.1.4-jakarta • Connectors 16.4.2 • Workbench 3.1.2 • © 2002–2025 Ontotext AD. All rights reserved.

RESULTS

Our knowledge graph supports:

- Unified mobility & air quality view
- Interoperability across unrelated sources
- Real semantic linking to Wikidata
- Querying complex relationships
- A reusable ontology

C O N C L U S I O N S

Across the 6 hands-ons, we:

- Identified and analyzed smart city datasets
- Cleaned and reconciled real data
- Designed an ontology
- Generated RDF using RML
- Linked to Wikidata
- Published the knowledge graph in GraphDB
- Validated it with SPARQL queries

This project gave us practical end-to-end experience with the Semantic Web stack.

Thank you!