HELLENIC REPUBLIC
National and Kapodistrian University of Athens
EST. 1837

Aiteam
NATIONAL & KAPODISTRIAN UNIVERSITY OF ATHENS

# An Introduction to Linked Geospatial Data

December 22, 2025

# Outline

- Examples of geospatial information
- Geospatial knowledge graphs
- GeoSPARQL
- Tutorial for project 3

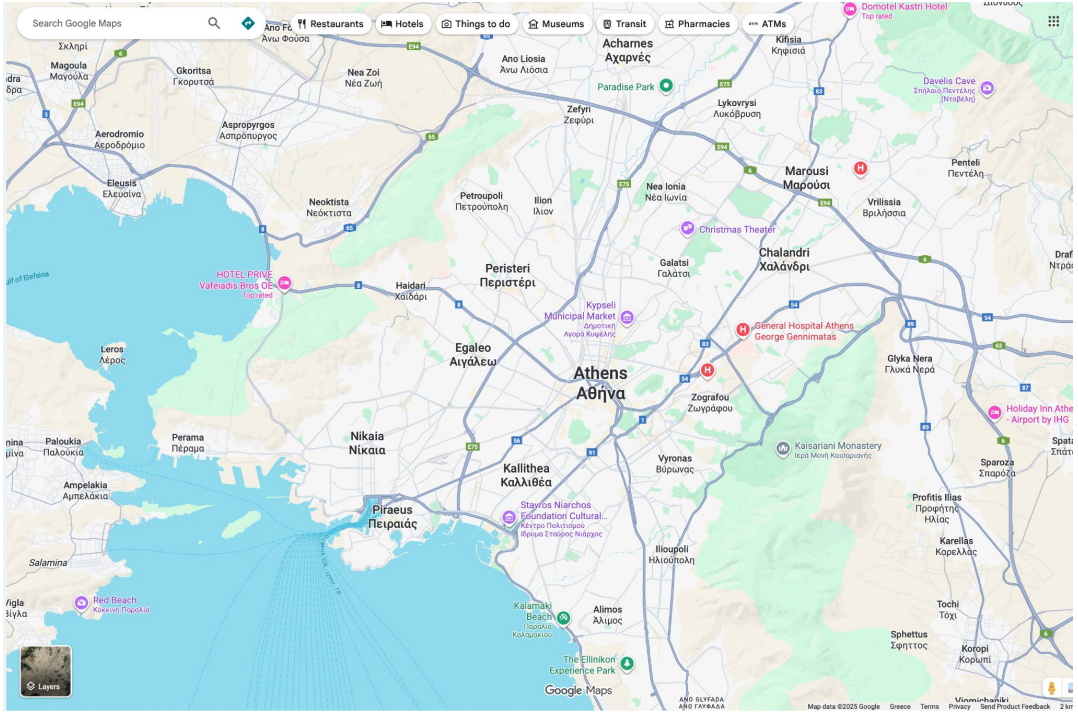
# Examples of geospatial information

# Why Geospatial Information?

- **Geospatial,** and in general **geographical,** information is very important in reality: everything that happens, happens somewhere (**location**).
- **Decision making can be substantially improved** if we know where things take place.

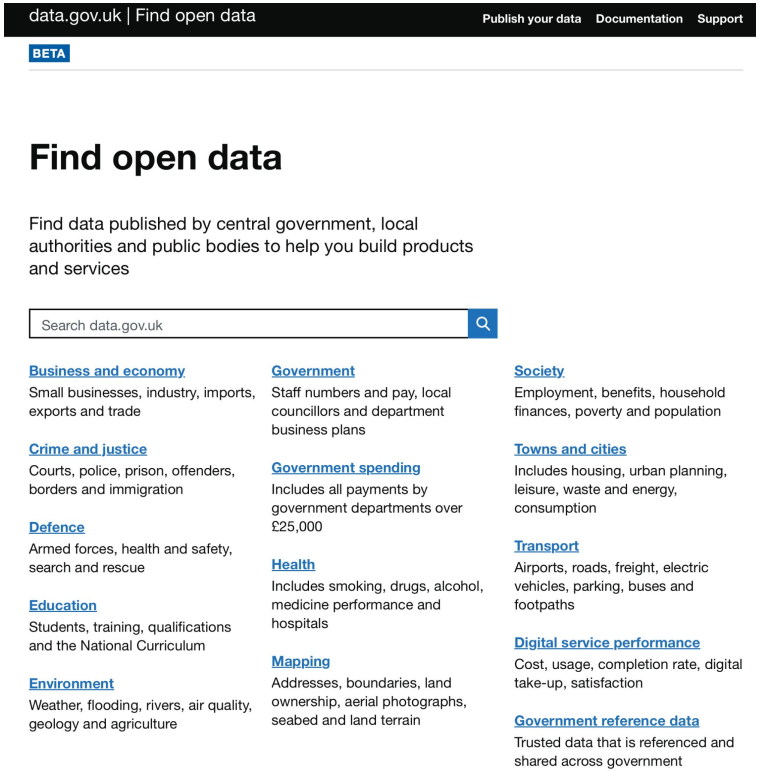
# Geospatial data on the Web

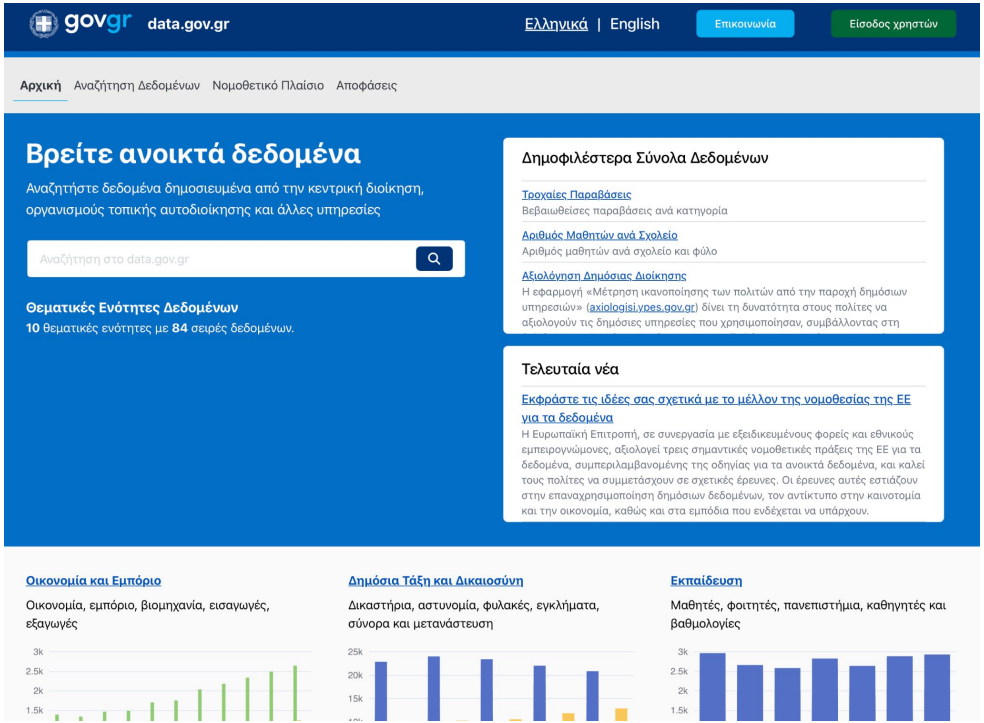
Google Maps



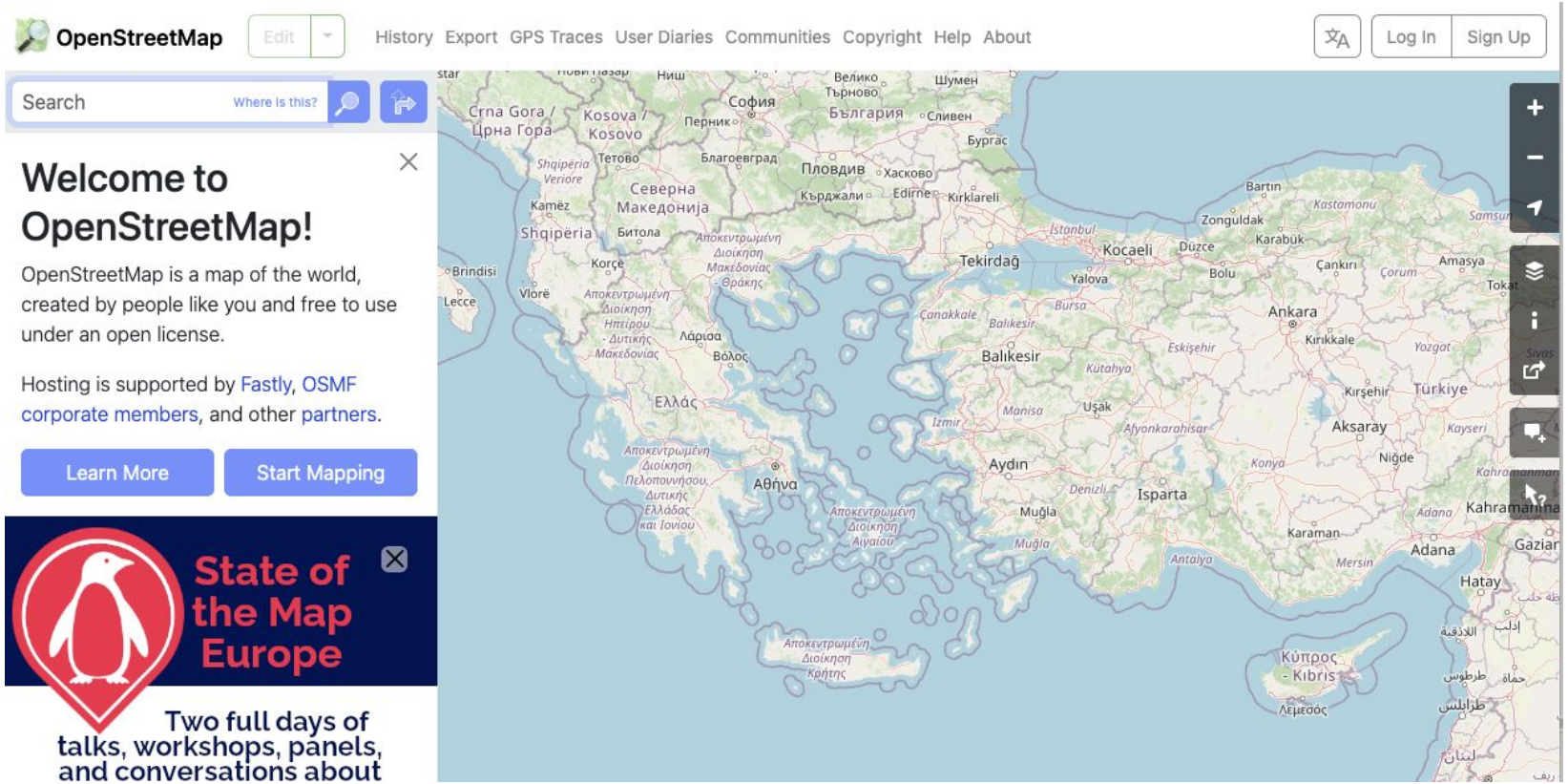

bing maps

# Open Government Data



# OpenStreetMap



# Satellite image of Athens

# Satellite image monitoring of wildfires

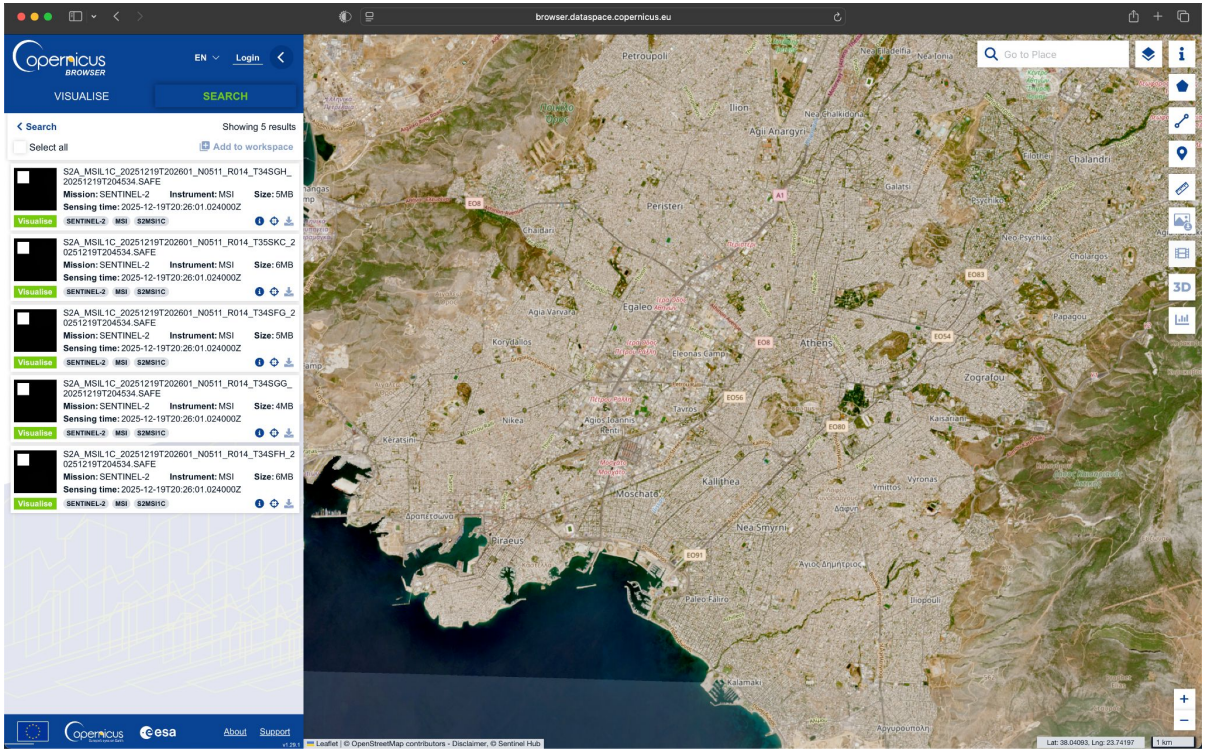
# Earth Observation Programmes

opernicus
Europe's eyes on Earth


JAXA
Japan Aerospace
Exploration Agency

# Copernicus programme

https://browser.dataspace.copernicus.eu/



# Geospatial knowledge graphs

# Geospatial information

- Qualitative
  - Cardinal direction relations (e.g., north of)
  - Topological relations (e.g., within, contains, overlaps, borders etc.)
  - Distance relations (e.g., near, far)
- Quantitative
  - Points
  - Lines
  - Polygons
  - More complex geometries

- Geospatial knowledge graphs are knowledge graphs representing not just thematic (e.g., “Tokyo is the capital of Japan” or “Nara has a population of 367,353 people”) but also geospatial information of the above kinds.

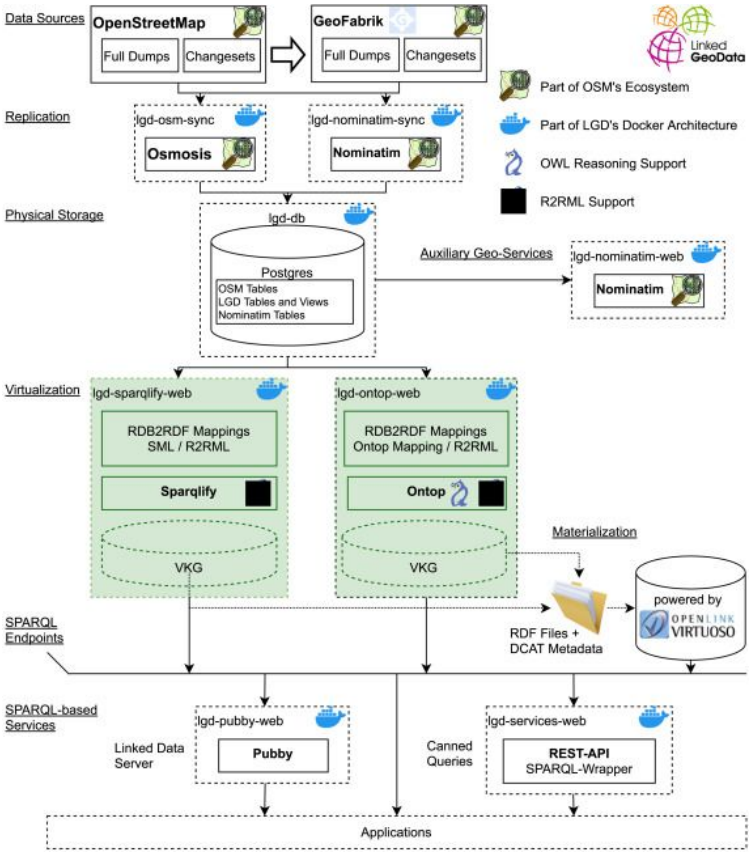
# Geospatial knowledge graphs

- **LinkedGeoData** was the first effort to add a spatial dimension to the Semantic Web. It made OpenStreetMap data available as RDF triples according to the Linked Data principles. In addition, it interlinked this data with other datasets in the Linked Open Data initiative.


Claus Stadler et al. *LinkedGeoData: A Core for a Web of Spatial Open Data*. Semantic Web Journal, 2012.

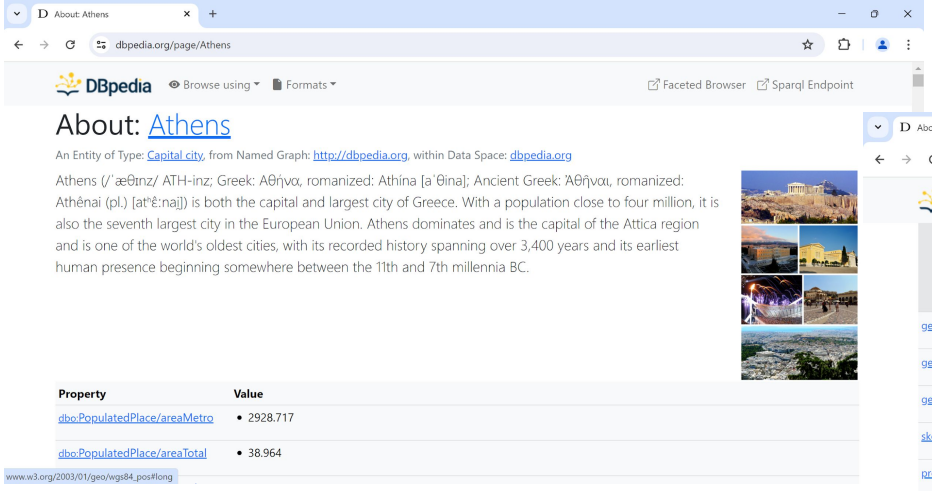
Linfang Ding et al. *Towards the next generation of the LinkedGeoData project using virtual knowledge graphs.* J. Web Semantics, 2021.

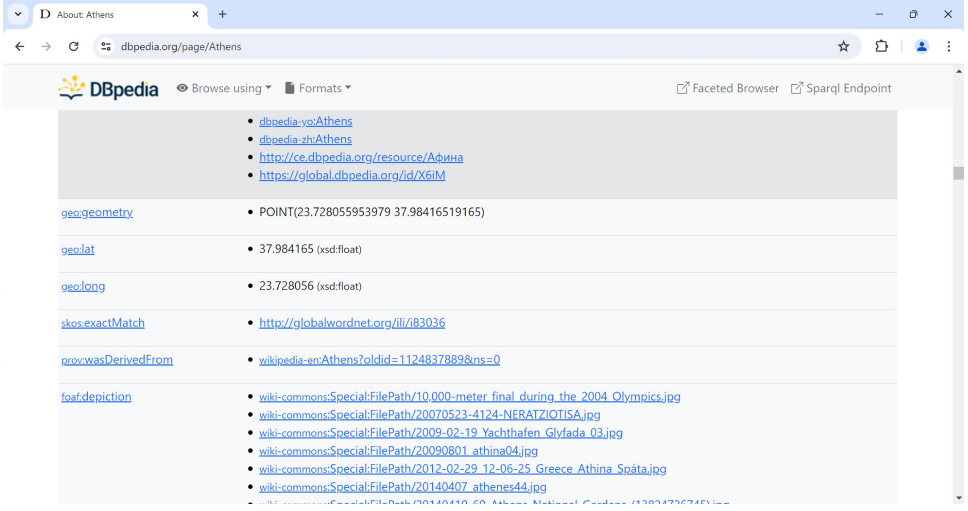
# LinkedGeoData using Virtual Knowledge Graphs



# Geospatial Knowledge Graphs

- DBpedia (lat/long coordinates, some cardinal directions, some topological relations implicitly)



# Number of geoentities in DBpedia

In the English version of DBpedia, **1.53 million entities** have geographic coordinates (latitude and longitude).

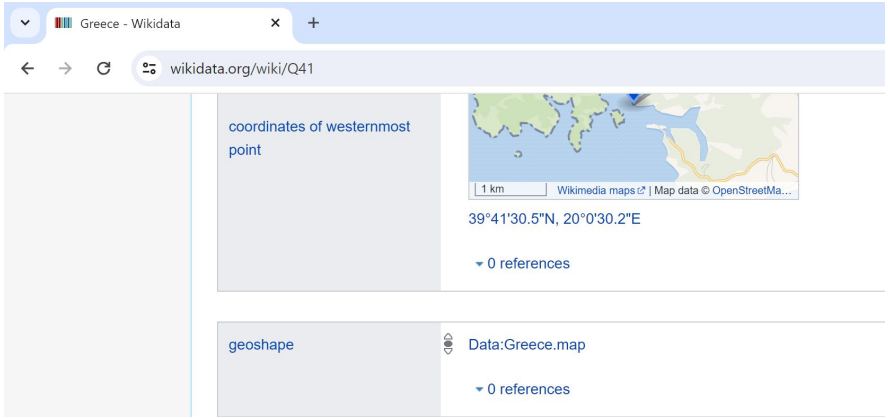
Note: Calculated by Kostas and Gemini 2.5 Pro

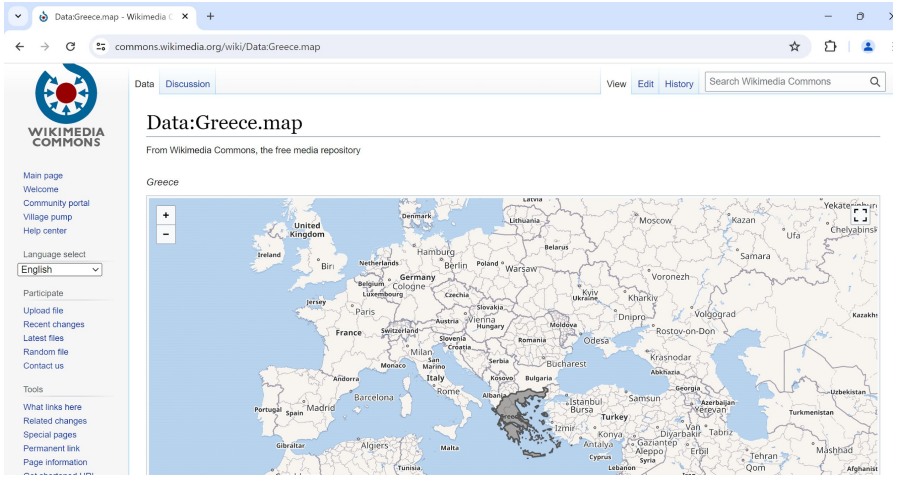
This data is extracted from Wikipedia infoboxes and represents a wide range of geographically located entities, including:

- Populated places: Cities, towns, and villages.
- Other locations: Mountains, rivers, lakes, and other natural features.
- Buildings and structures: Famous landmarks, museums, and stadiums.

# Geospatial Knowledge Graphs

- Wikidata (same as DBpedia but also geoshapes). Geoshapes encode complex geometries e.g.,  multipolygons.



# Number of geoentities in Wikidata

As of the latest data from Wikidata's SPARQL endpoint, the total number of geoentities is **12,968,093**.

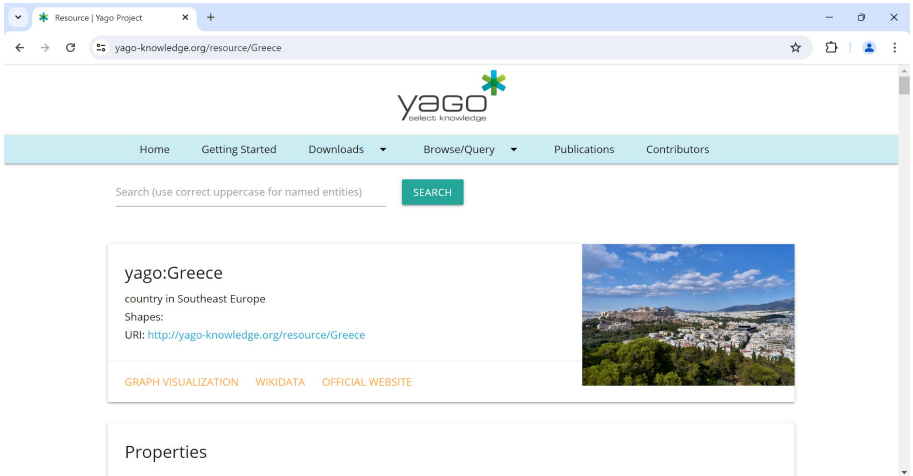
This total is a combination of two types of geospatial data:

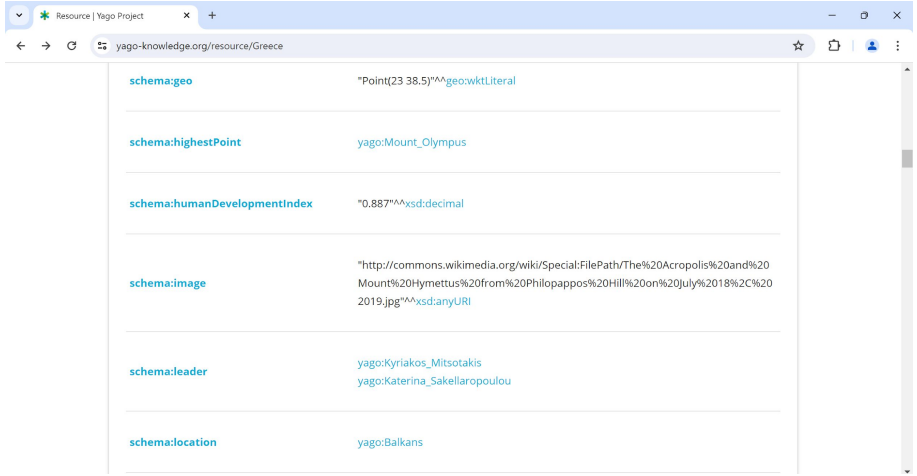
- Entities with coordinate locations: 12,877,759
- Entities with geoshapes: 90,334

Note: Calculated by Gemini 2.5 Pro.

# Geospatial Knowledge Graphs

- YAGO2 (geoentities from Wikipedia and GeoNames, lat/long coordinates only, some topological relations implicitly)

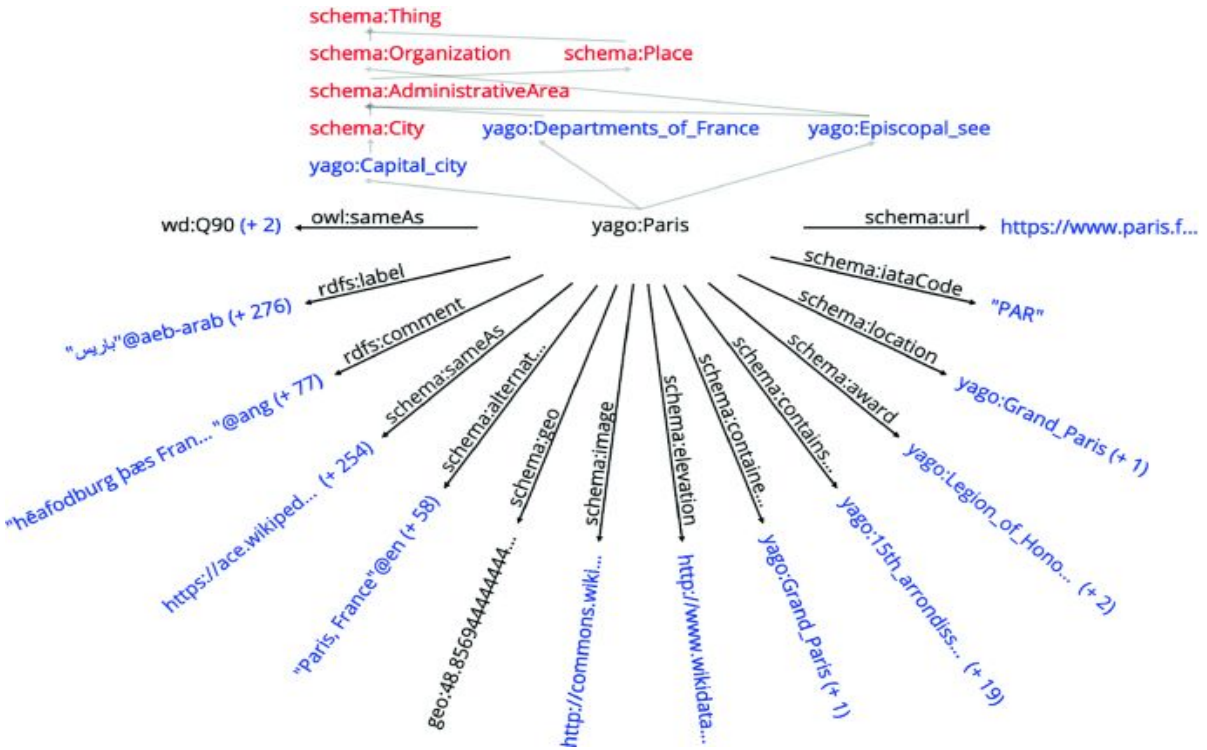



10 million geoentities

Hoffart et al. *YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia.* Artif. Intell. 194: 28-61 (2013)

# An Example from YAGO4.5 (https://yago-knowledge.org/)



# Geospatial information in schema.org

The property **geo** can be used to encode geospatial information in schema.org.

Property `geo` is expected to be used on type `Place` or `Event` and its values can be of types `GeoCoordinates` or `GeoShape`.

An instance of `GeoCoordinates` can be described with various properties including latitude and longitude.

# Example in JSON-LD

```
<script type="application/ld+json">
{
  "@context": "https://schema.org",
  "@type": "Place",
  "geo": {
    "@type": "GeoCoordinates",
    "latitude": "40.75",
    "longitude": "-73.98"
  },
  "name": "Empire State Building"
}
</script>
```

# Geospatial information in schema.org

An instance of `GeoShape` can be described with various properties including line, box and polygon.

A more specific type is `GeoCircle`.

# YAGO 4.5

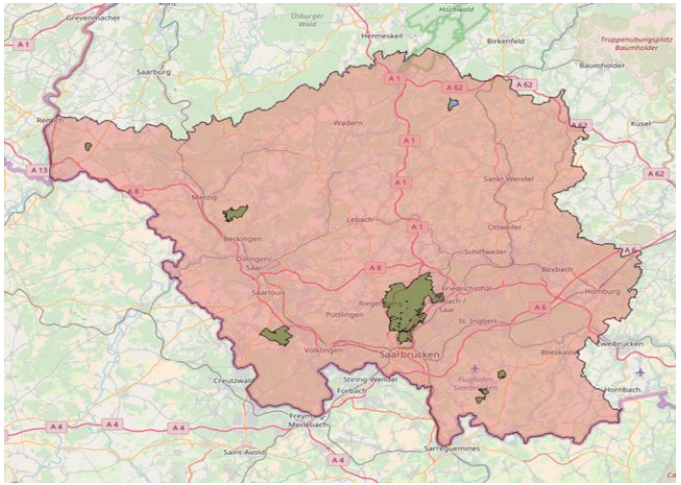
YAGO is now in version 4.5. So far it has not supported more complex geometries than coordinates.


Suchanek, F et al. YAGO 4.5: A Large and Clean Knowledge Base with a Rich Taxonomy. SIGIR 2024.

# Geospatial Knowledge Graphs

- YAGO2geo (it adds to YAGO2 more complex geometries from administrative divisions of four countries and some OpenStreetMap features)

Polygons and multipolygons: 640,000
Linestrings: 137,000




Karalis et al. *Extending the YAGO2 Knowledge Graph with Precise Geospatial Knowledge.* ISWC 2019.

# YAGO2geo data sources

The YAGO2geo knowledge graph is constructed by enriching YAGO2 with data from the following sources:

- The Greek Administrative Geography (GAG) dataset
- Administrative divisions data for the United Kingdom obtained from Ordnance Survey and Ordnance Survey Northern Ireland
- Administrative divisions data for Ireland obtained from Ordnance Survey Ireland
- Administrative divisions data for the USA obtained from the United States Census Bureau
- Administrative divisions data from the Global Administrative Areas dataset (GADM)
- the following classes of OSM features: natural, waterways, landuse and leisure, places

# The Ontology of YAGO2geo

The ontology of YAGO2geo is defined by enriching the ontology of YAGO2 with:

- the ontologies of the administrative geographies of Greece, UK, Ireland and USA
- the ontology of the Database of Global Administrative Areas (GADM)
- an ontology for the part of OSM covered

# The Knowledge Graph

The knowledge graph (instances) is constructed by extending YAGO2 with new triples encoding thematic and geospatial information from the above data sources using their ontologies.

In addition, `owl:sameAs` triples are added for objects in two different data sources that refer to the same real world entity.

**Example**: the geoentity `hellenic_republic_390903` of YAGO2 is identical to the entity `GRC` of GADM since they both represent the country Greece. Therefore the triple

```
hellenic_republic_390903 owl:sameAs GRC
```

is added.

# The Knowledge Graph

The matching phase for identifying identical entities consists of applying two filters:

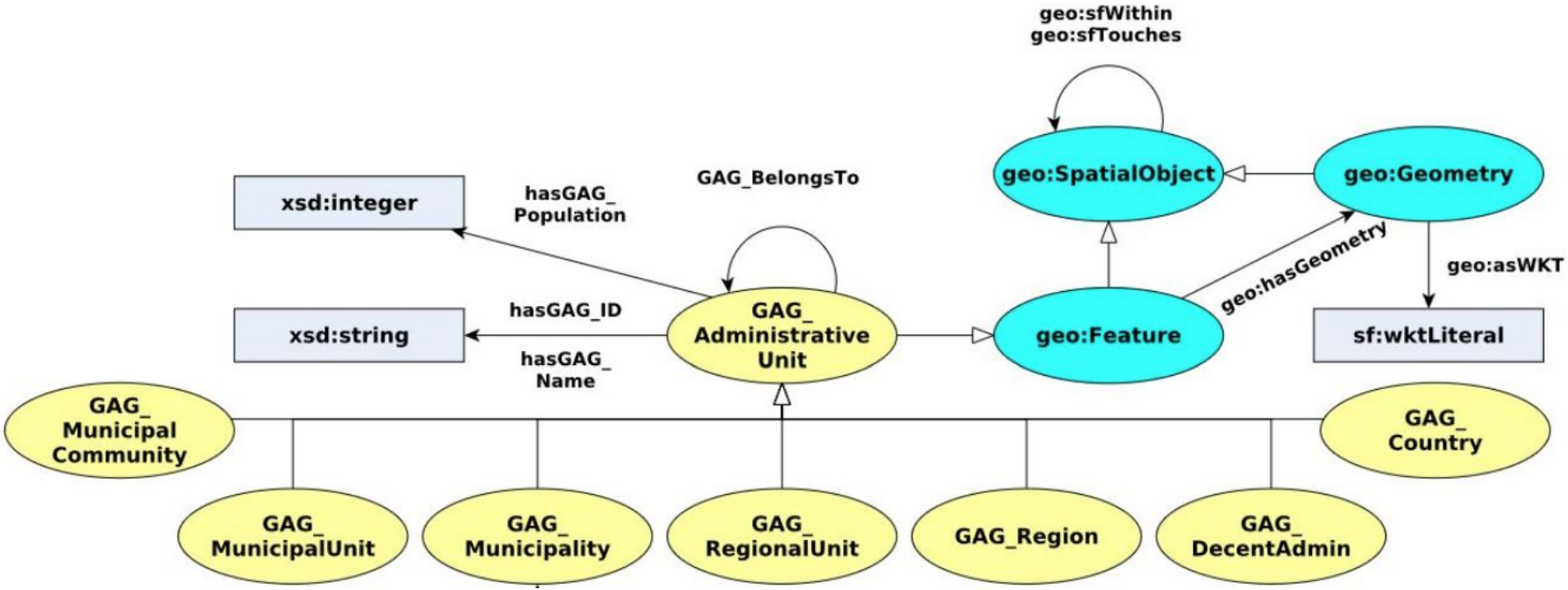
- the label similarity filter (Jaro-Winkler similarity)
- the geometry distance filter (Euclidean distance in the WGS:84 coordinate system smaller than 0.2 degrees).

The methodology is based on the methodology that was used in YAGO2 when integrating information from GeoNames. A similar approach has been used in LinkedGeoData.

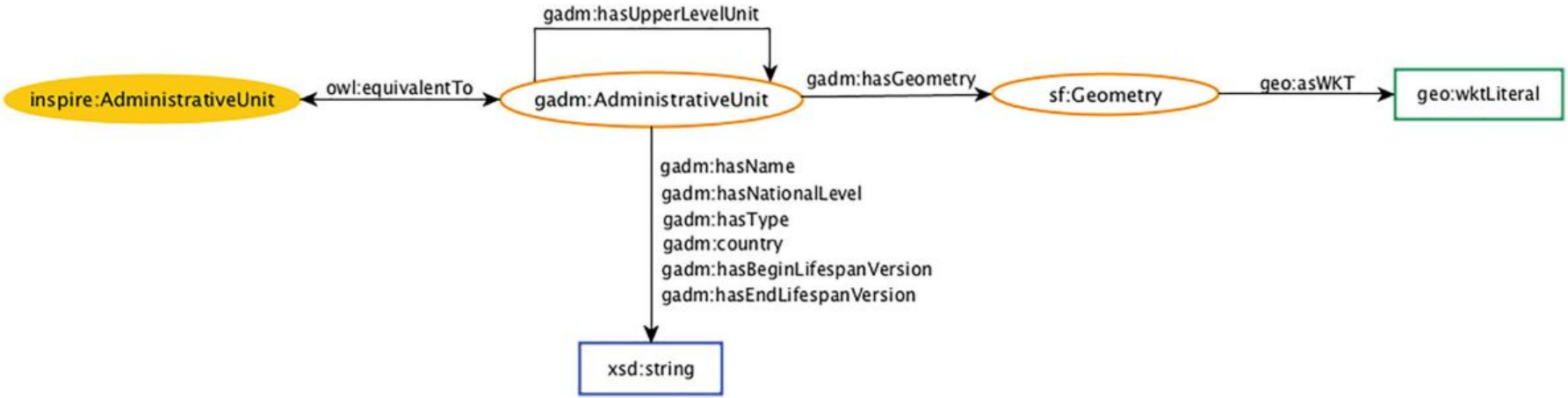
# The Greek Administrative Geography (GAG)

See https://en.wikipedia.org/wiki/Kallikratis_Programme

# The ontology of YAGO2: GAG



# The ontology of YAGO2: GADM



# An Example from YAGO2geo

```
gag:Olympia
   rdf:type gag:MunicipalCommunity;
   gag:hasGAG_Name "Ancient Olympia";
   gag:hasGAG_Population "184"^^xsd:int;
   geo:hasGeometry geo:OlympiaGeometry.
```

Ancient Olympia

# An Example from YAGO2geo

```
geo:OlympiaGeometry
    rdf:type geo:Geometry;
    geo:asWKT "MULTIPOLYGON(((308511.906249999 4201042,308763.8125
    4200714, 308840.09375 4200629,308939.3125 4200545,.....,
    308390.000000001 4201276,308451.593749999 4201167,308467
    4201125,308511.906249999
    4201042)));<http://www.opengis.net/def/crs/EPSG/0/2100>"^^sf:wktLitera
    l.
```

Ancient Olympia

# YAGO4geo

We have shown how to extend YAGO4 with rich geospatial information in the spirit of YAGO2geo.

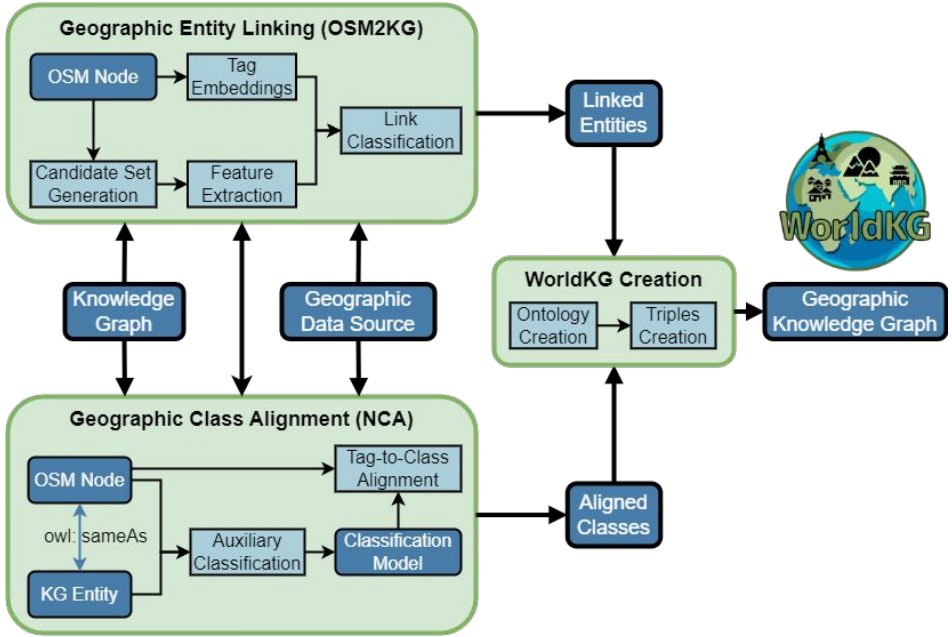
The idea here was to use the schema.org facilities for representing geospatial information since YAGO4 heavily relies on schema.org.

However, the limited geometric types of schema.org (e.g., multipolygons are not supported) did not allow us to come up with an interesting geospatial knowledge graph.

Siampou, M.D et al. *Extending YAGO4 Knowledge Graph with Geospatial Knowledge.* 5th International Workshop on Geospatial Linked Data at ESWC 2022.

# Geospatial Knowledge Graphs

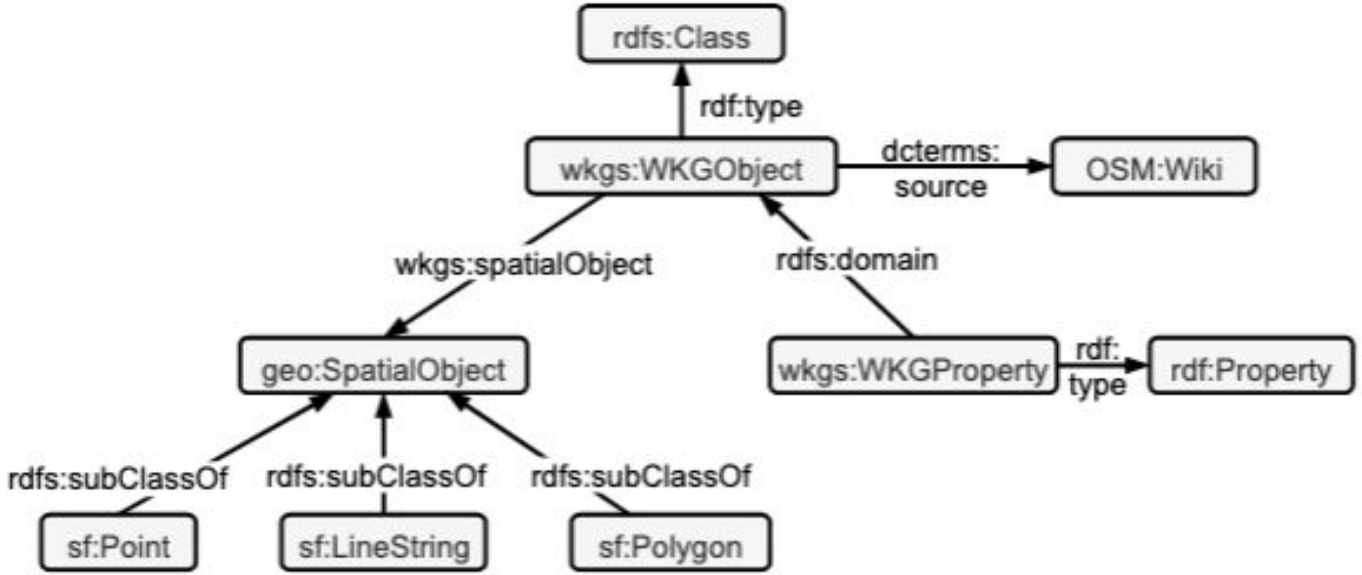
- WorldKG (makes OpenStreetMap data available as a knowledge graph)






Dsouza, et al. *WorldKG: A World-Scale Geographic Knowledge Graph.* CIKM 2021.

# The WorldKG ontology




Only OSM nodes i.e.,
only points

# WorldKG classes

- WorldKG classes are constructed from **map features** of OSM. See https://wiki.openstreetmap.org/wiki/Map_features.
- The map feature list is used to construct a class hierarchy. In particular, the authors consider all keys in the feature map list as top-level classes (e.g., `natural`). Values assigned to the keys are represented as their subclasses.
- For example, from `natural=cave_entrance` becomes a subclass of `natural`.

# WorldKG classes (cont'd)

- Only categorical values of keys are considered as subclasses in WorldKG.
- Other value types, e.g., Boolean or numerical values, are not considered as subclasses. Instead, the top-level class provided by the corresponding key is used.
- For example, an entity with a tag `building=yes` is typed as `wkgs:Building`.

# WorldKG properties

- Properties and their values are created from OSM keys that have a valid English OSM Wiki page and are not mapped to own classes. For example, the value of property `wkgs:addrCountry` may be inferred from the key with Wiki page https://wiki.openstreetmap.org/wiki/Germany .

- Each class and property is linked to an OSM Wiki page via `dcterms:source`.

# Schema Alignment with Existing KGs

- To link the WorldKG ontology to other existing ontologies, the authors determine equivalent OSM tags and classes from the Wikidata and DBpedia knowledge graphs.
- The **Neural Class Alignment** technique from the paper below is used.

Dsouza, et al. *Towards Neural Schema Alignment for OpenStreetMap and Knowledge Graphs.* ISWC 2021.

# WorldKG Statistics

| Quantity | Count |
|---|---|
| Total triples | 828,550,751 |
| Total entities | 113,444,975 |
| Top-level classes | 33 |
| Sub-classes | 1,143 |
| Unique properties | 1,820 |
| Links to Wikidata classes | 40 |
| Links to DBpedia classes | 21 |

Only points in the SPARQL endpoint currently.

13 million

1.53 million

In total, WorldKG covers 113,444,975 geographic entities, clearly more than Wikidata (8,621,058) and DBpedia (8,621,058).

# Geospatial Knowledge Graphs

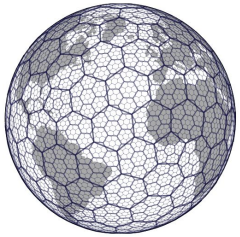
- KnowWhereGraph (integrates thematic and place-centric datasets into a geospatial knowledge graph).
- Pilot scenarios in disaster relief, agricultural land use and food-related supply chains.

| Thematic Datasets | | | | | Place-Centric Datasets | | |
|---|---|---|---|---|---|---|---|
| **Dataset Name/ Theme** | **Source Agency** | **Key Attributes** | **Spatial Coverage** | **Temporal Coverage** | **Place-Centric Dataset** | **Defining Authority** | **Spatial Coverage** |
| Soil Properties | USDA | soil type, farmland class | Targeted regions in US | Current | S2 Cells | Google | Lvl 9 (Global), Lvl 13 (US), |
| Wildfires | USGS, USDA, USFS, NIFC | wildfire type, burn severity, num. acres burned, contained date | US | 1984–current | Global Administrative Regions | University of Berkeley, Museum of Vertebrate Zoology and the International Rice Research Institute | Global |
| Earthquakes | USGS | magnitude, length, width, geometry | Global (mag. over 4.5) | 2011-01-01 to 2022-01-18 | | | |
| Climate Hazards | NOAA | injuries, deaths, property damages | US | 1950–2022 | | | |
| Expert - Covid-19 Mobility | Direct Relief (DR) | name, affiliation, expertise | Global | 2021 | US Federal Judicial District | DoJ, ESRI | US |
| Expert - General | KWG, UC System, DR, Semantic Scholar | name, affiliation, expertise with spatiotemporal scopes | Global | unlimited | National Weather Zones | NOAA | US |
| Cropland Types | USDA | crop types (raster data) | US | 2008-2021 | FIPS Codes | NRCS | US |
| Air Qual. Obs. | U.S. EPA | AQI value, CO concentration | US | 1980–2022 | Designated Market Area | Nielen | US |
| Smoke Plumes | NOAA | daily smoke plumes extent | US | 2010-2022 | ZIP | ZCTA | US |
| Climate Observations | NOAA | temperature, precipitation, PDSI, PHSI | US | 1950 - 2022 | Climate Division | NOAA | US |
| Disaster Declaration | FEMA | designated area, program, amount approved, program designated date | US | 1953 - 2022 | Census Metropolitan Area | US Census | US |
| Smoke Plume Extents | NOAA | Smoke extent | US | 2017 - 2022 | Drought Zone | NDMC, USDA,NOAA | US |
| BlueSky Forecasts | Bluesky | PM10, PM5 | US | 2022-03-07 | Geographic Name Information System | USGS | US |
| Transportation (highway network) | DOT | road type, road length, road sign | US | 2014 | | | |
| Public Health Observations | CDC, US Census, University of Wisconsin Population Health Institute | below poverty level, diabetes, obesity, mental health provider rate, annual mammogram, vaccinated, injury death | US | 2017, 2021 | | | |
| Public Health Infrastructure | HIFLD | pharmacies, hospitals, dialysis centers, public health departments | US | 2017 - 2022 | | | |
| Social Vulnerability | CDC/ATSDR | social vulnerability index | US | 2018 | | | |
| Hurricane Tracks | NOAA | max wind speed, min pressure | US | 1851-2020 | | | |

Janowicz et al. *Know, Know Where, Knowwheregraph: A Densely Connected, Cross-Domain Knowledge Graph and Geo-Enrichment Service Stack for Applications in Environmental Intelligence*. AI Mag. 43(1): 30-39 (2022)

# Geospatial knowledge graphs

- **H3-GeoKG** is geospatial knowledge graph build from OSM data (nodes, ways and relations).
- The developers of H3-GeoKG use the H3 discrete global grid to harmonize the geometries that are extracted from OSM similar to KnowWhereGraph.
- The most important difference with KnowWhereGraph is that the discrete global grid uses a square-based grid cell geometry, whereas H3 uses a hexagonal-based grid cell geometry.
- The benefit is that in a hexagonal grid the distance to all neighboring cells is uniform, which is not the case for square- or triangular-based grids.



Martin Böckling et al. *A Planet Scale Spatial-Temporal Knowledge Graph Based On OpenStreetMap And H3 Grid.* GeoLD 2024.

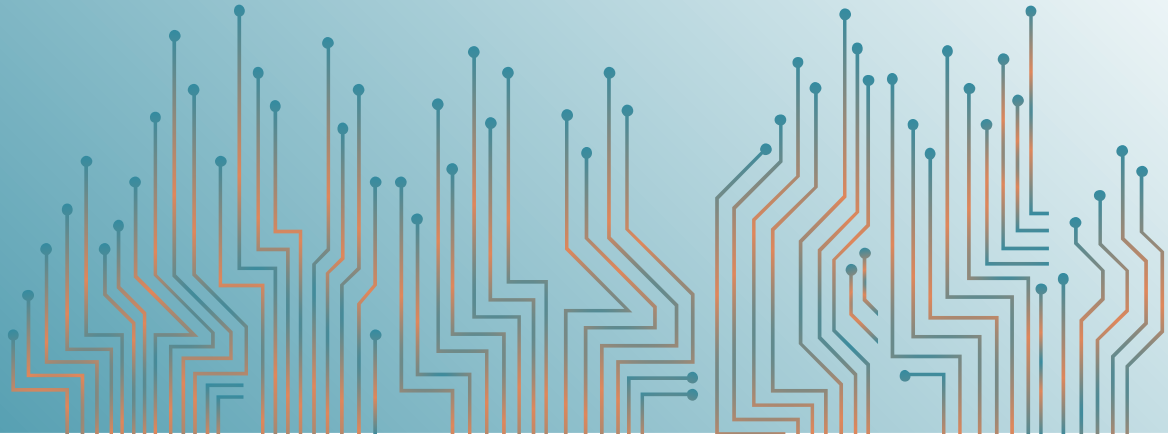
# Other geospatial knowledge graphs

- CrowdGeoKG (Chen et al., 2017)
- HGeoKG (Li et al., 2025)

J. Chen et al. *CrowdGeoKG: Crowdsourced Geo-Knowledge Graph.* CCKS 2017.

Tailong Li et al. *HGeoKG: A Hierarchical Geographic Knowledge Graph for Geographic Knowledge Reasoning.* ISPRS Int. J. Geo-Inf. 2025.

# Thank you!



Aiteam
NATIONAL & KAPODISTRIAN
UNIVERSITY OF ATHENS