Wikidata with SPARQL

David Arroyo Menéndez

September 30, 2019

Wikidata Definition

Wikidata is a collaboratively edited knowledge base hosted by the Wikimedia Foundation. It is a common source of open data that Wikimedia projects such as Wikipedia can use, and anyone else, under a public domain license. This is similar to the way Wikimedia Commons provides storage for media files and access to those files for all Wikimedia projects, and which are also freely available for reuse. Wikidata is powered by the software Wikibase.
See: `https://www.wikidata.org`

SPARQL definition

SPARQL is a recursive acronym for SPARQL Protocol and RDF Query
Language) is an RDF query language—that is, a semantic query language
for databases—able to retrieve and manipulate data stored in Resource
Description Framework.

SPARQL as Language

If you take a look at Germany (Q183), then you can see a whole host of properties like population (P1082), median income (P3529) or even images with the image (P18) property.

```
SELECT
  ?country ?countryLabel ?population ?area ?medianIncome
WHERE {
  ?country wdt:P463 wd:Q458.
  ?country wdt:P1082 ?population.
  ?country wdt:P2046 ?area.
  ?country wdt:P3529 ?medianIncome.

  SERVICE wikibase:label { bd:serviceParam wikibase:language "e
}
```

SPARQL in Python

```
import requests
url = 'https://query.wikidata.org/sparql'
query = """
SELECT
  ?country ?countryLabel ?population ?area ?medianIncome
WHERE {
  ?country wdt:P463 wd:Q458.
  ?country wdt:P1082 ?population.
  ?country wdt:P2046 ?area.
  ?country wdt:P3529 ?medianIncome.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "e
}
"""
r = requests.get(url, params = {'format': 'json', 'query': quer
data = r.json()
print(data)
```