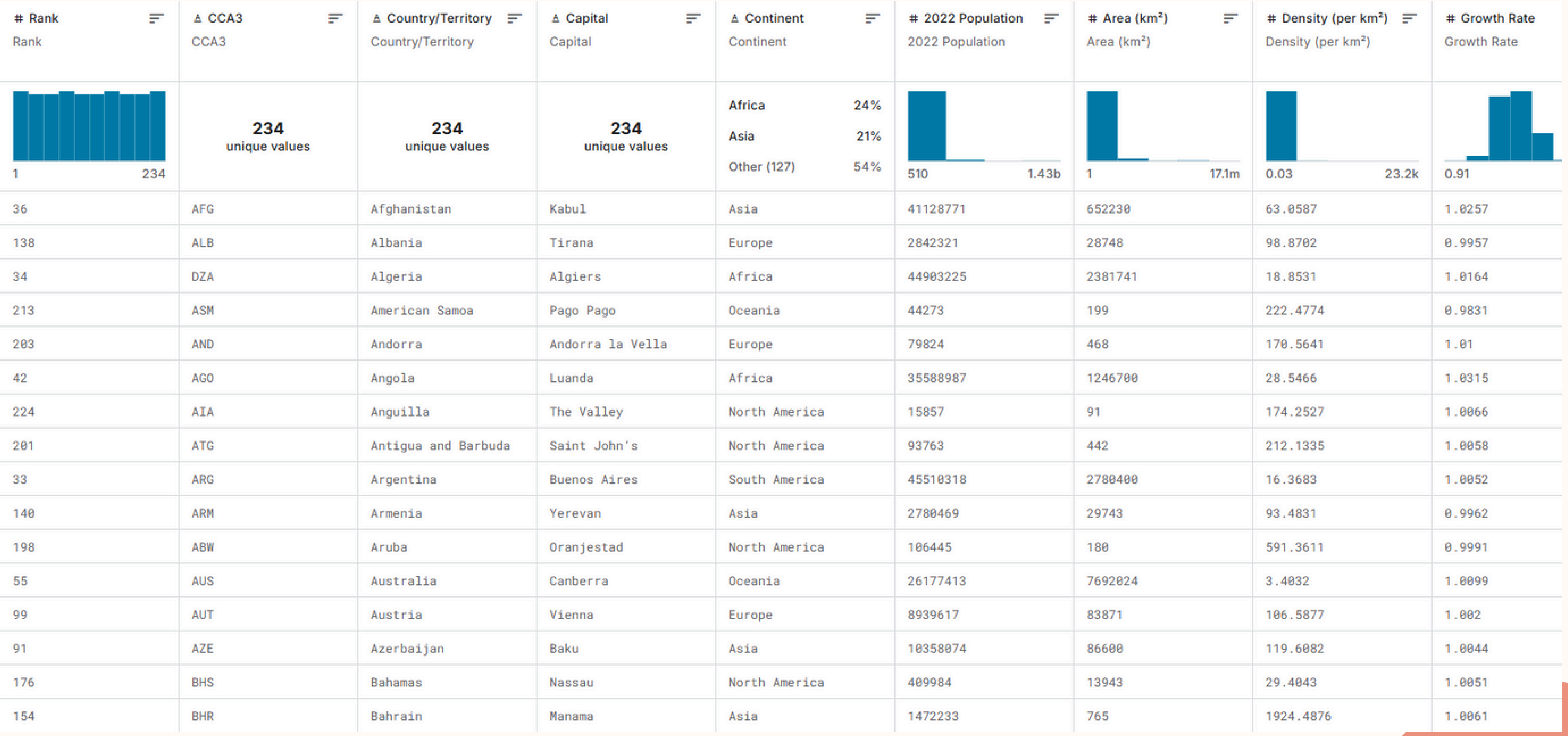
Projet Dana - Gr 6

DATA SET.

Description : Évolution de la population de tout les pays de 1970 a 2022.



# Rank	CCA3	Country/Territory	Capital	Continent	# 2022 Population	# Area (km^2)	# Density (per km^2)	# Growth Rate
Rank	CCA3	Country/Territory	Capital	Continent	2022 Population	Area (km^2)	Density (per km^2)	Growth Rate
1 234	234 unique values	234 unique values	234 unique values	Africa 24% Asia 21% Other (127) 54%	510 1.43b	1 17.1m	0.03 23.2k	0.91
36	AFG	Afghanistan	Kabul	Asia	41128771	652230	63.0587	1.0257
138	ALB	Albania	Tirana	Europe	2842321	28748	98.8702	0.9957
34	DZA	Algeria	Algiers	Africa	44903225	2381741	18.8531	1.0164
213	ASM	American Samoa	Pago Pago	Oceania	44273	199	222.4774	0.9831
203	AND	Andorra	Andorra la Vella	Europe	79824	468	170.5641	1.01
42	AGO	Angola	Luanda	Africa	35588987	1246700	28.5466	1.0315
224	AIA	Anguilla	The Valley	North America	15857	91	174.2527	1.0066
201	ATG	Antigua and Barbuda	Saint John's	North America	93763	442	212.1335	1.0058
33	ARG	Argentina	Buenos Aires	South America	45510318	2780400	16.3683	1.0052
140	ARM	Armenia	Yerevan	Asia	2780469	29743	93.4831	0.9962
198	ABW	Aruba	Oranjestad	North America	106445	180	591.3611	0.9991
55	AUS	Australia	Canberra	Oceania	26177413	7692024	3.4032	1.0099
99	AUT	Austria	Vienna	Europe	8939617	83871	106.5877	1.002
91	AZE	Azerbaijan	Baku	Asia	10358074	86600	119.6082	1.0044
176	BHS	Bahamas	Nassau	North America	409984	13943	29.4043	1.0051
154	BHR	Bahrain	Manama	Asia	1472233	765	1924.4876	1.0061

data set simplifié

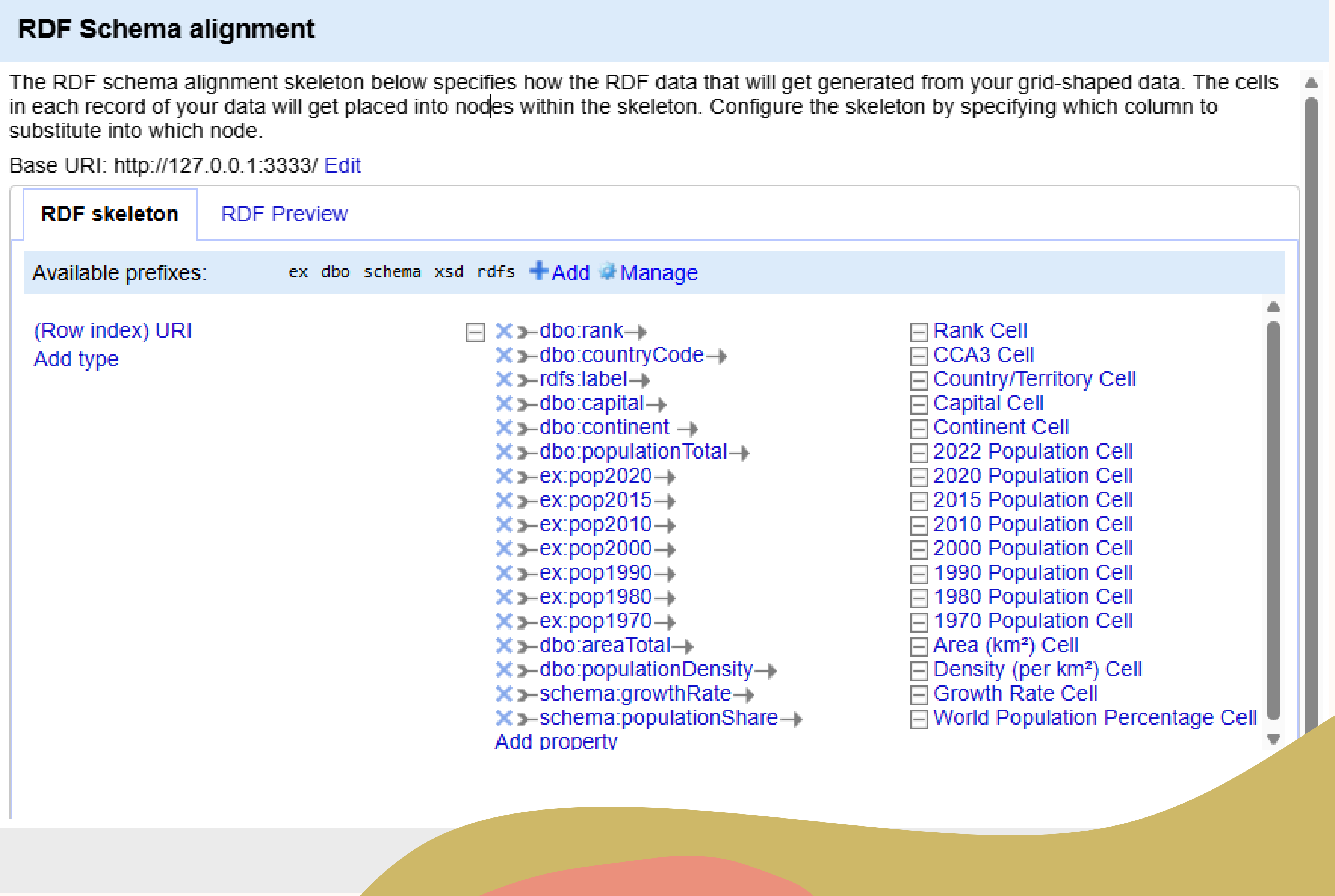
Source : https://www.kaggle.com/datasets/iamsouravbanerjee/world-population-dataset

ÉTAPE 1.

Transformation en RDF.

Nous avons utilisé OpenRefine pour transformer le fichier CSV. Celui-ci étant déjà propre, aucune opération de nettoyage n’a été nécessaire. Nous avons typé les variables selon leur nature afin d’assurer une structure cohérente et adaptée aux différents types de données.

Source: https://openrefine.org/



RDF Schema alignment

The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.

Base URI: http://127.0.0.1:3333/ Edit

RDF skeleton | RDF Preview

Available prefixes: ex dbo schema xsd rdfs +Add Manage

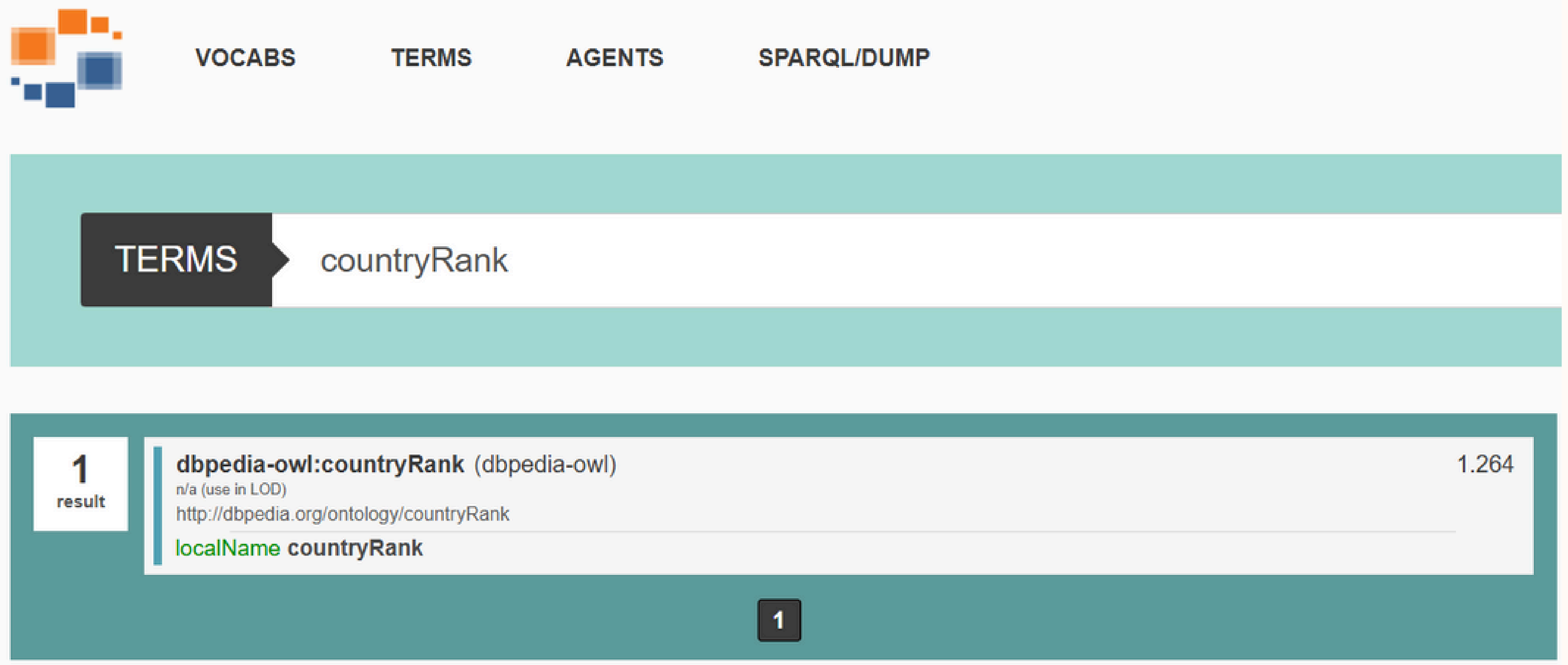
(Row index) URI	dbo:rank→	Rank Cell
Add type	dbo:countryCode→	CCA3 Cell
	rdfs:label→	Country/Territory Cell
	dbo:capital→	Capital Cell
	dbo:continent →	Continent Cell
	dbo:populationTotal→	2022 Population Cell
	ex:pop2020→	2020 Population Cell
	ex:pop2015→	2015 Population Cell
	ex:pop2010→	2010 Population Cell
	ex:pop2000→	2000 Population Cell
	ex:pop1990→	1990 Population Cell
	ex:pop1980→	1980 Population Cell
	ex:pop1970→	1970 Population Cell
	dbo:areaTotal→	Area (km^2) Cell
	dbo:populationDensity→	Density (per km^2) Cell
	schema:growthRate→	Growth Rate Cell
	schema:populationShare→	World Population Percentage Cell
	Add property	

RDF skeleton OpenRefine

ÉTAPE 1.

Transformation en RDF.

Pour la création des prédicats, nous avons sélectionné ceux correspondant à nos besoins principalement dans le vocabulaires dbo. Nous avons également défini nos propres prédicats afin de différencier les années liées aux données de population et pour certaines autres données.



countryRank

1 result

dbpedia-owl:countryRank (dbpedia-owl) 1.264
n/a (use in LOD)
http://dbpedia.org/ontology/countryRank
localName countryRank

Exemple de recherche de prédicat existant

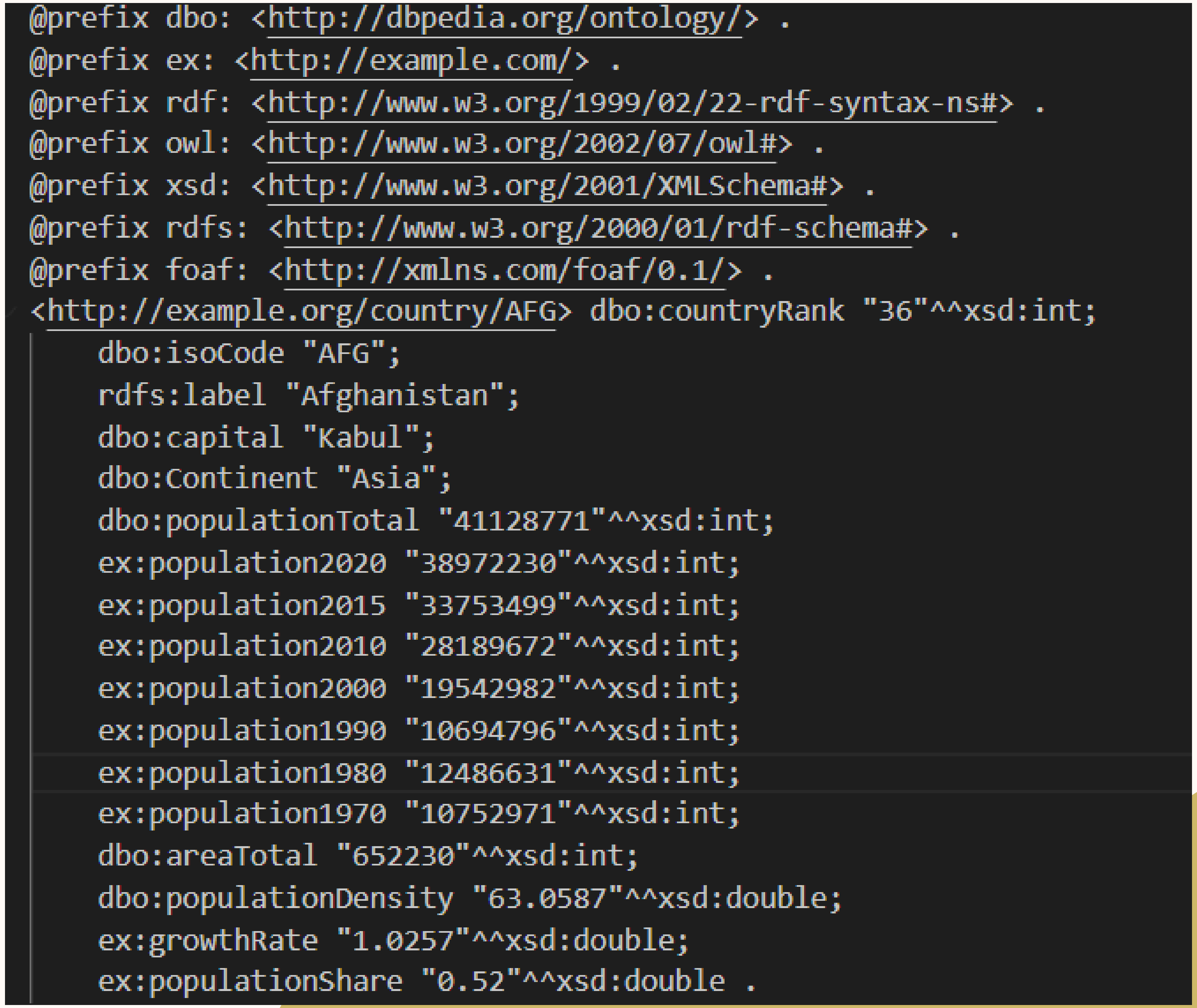
Source: https://lov.linkeddata.es/dataset/lov/

ÉTAPE 1.

Transformation en RDF.

Voici le rendu en Turtle après la transformation par Open Refine :

Extrait du fichier Turtle



```
@prefix dbo: <http://dbpedia.org/ontology/> .
@prefix ex: <http://example.com/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
<http://example.org/country/AFG> dbo:countryRank "36"^^xsd:int;
    dbo:isoCode "AFG";
    rdfs:label "Afghanistan";
    dbo:capital "Kabul";
    dbo:Continent "Asia";
    dbo:populationTotal "41128771"^^xsd:int;
    ex:population2020 "38972230"^^xsd:int;
    ex:population2015 "33753499"^^xsd:int;
    ex:population2010 "28189672"^^xsd:int;
    ex:population2000 "19542982"^^xsd:int;
    ex:population1990 "10694796"^^xsd:int;
    ex:population1980 "12486631"^^xsd:int;
    ex:population1970 "10752971"^^xsd:int;
    dbo:areaTotal "652230"^^xsd:int;
    dbo:populationDensity "63.0587"^^xsd:double;
    ex:growthRate "1.0257"^^xsd:double;
    ex:populationShare "0.52"^^xsd:double .
```

ÉTAPE 1.

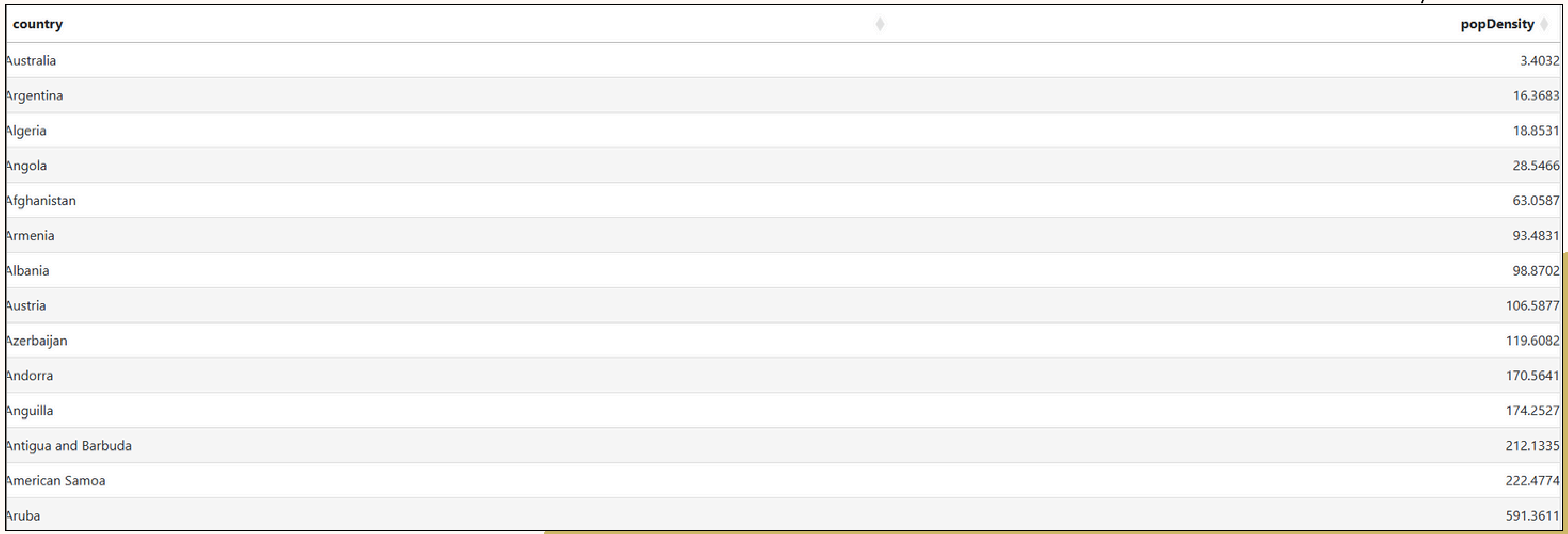
Requete 1.

Nom des pays commençant par la lettre A trier par leur taux de densité de la population.

```
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?country ?popDensity
WHERE {
   ?s rdfs:label ?country ;
      dbo:populationDensity ?popDensity .
   FILTER (regex(?country, "^A"))
}
ORDER BY (?popDensity)
LIMIT 10
```

Résultat de la requête 1 Fuseki



country	popDensity
Australia	3.4032
Argentina	16.3683
Algeria	18.8531
Angola	28.5466
Afghanistan	63.0587
Armenia	93.4831
Albania	98.8702
Austria	106.5877
Azerbaijan	119.6082
Andorra	170.5641
Anguilla	174.2527
Antigua and Barbuda	212.1335
American Samoa	222.4774
Aruba	591.3611

ÉTAPE 1

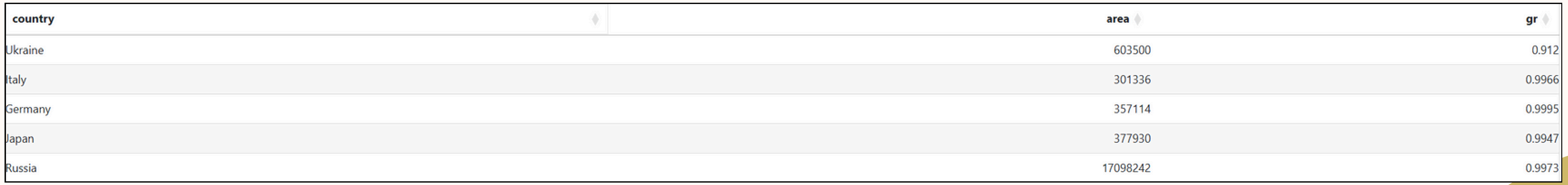
Requete 2

Affiche les pays de superficie > 300 000 km² qui ont un taux de croissance < 1

```
PREFIX ex: <http://example.com/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?country ?area ?gr
WHERE {
  ?s rdfs:label ?country ;
     dbo:areaTotal ?area ;
     ex:growthRate ?gr ;
    dbo:populationTotal ?pop .
  FILTER (?gr < 1 && ?area > 300000)
}
ORDER BY (?pop)
LIMIT 10
```

Résultat de la requête 2 Fuseki



country	area	gr
Ukraine	603500	0.912
Italy	301336	0.9966
Germany	357114	0.9995
Japan	377930	0.9947
Russia	17098242	0.9973

ÉTAPE 2.

Requete 3 .

Affiche le nombre de prix nobel par Million d'habitant
lien avec G2

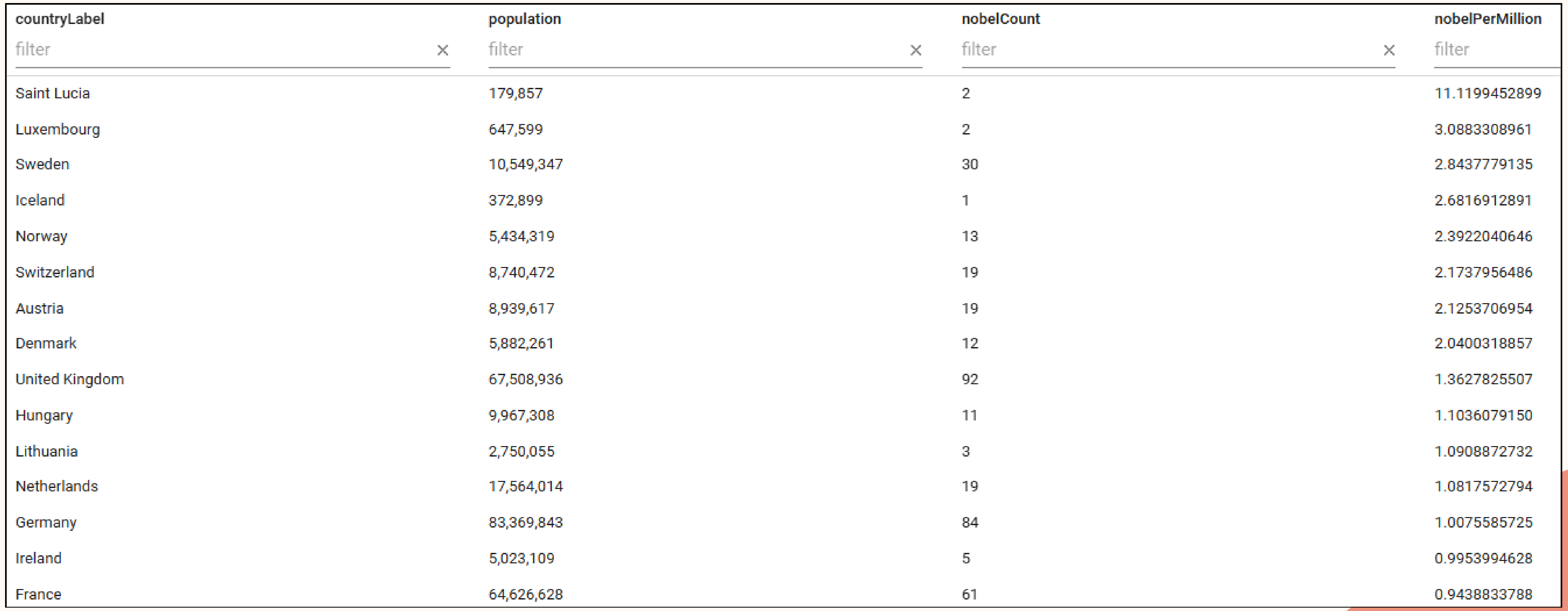
```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX schema: <http://schema.org/>

SELECT ?countryLabel ?population
(COUNT(?nobel) AS ?nobelCount)
((COUNT(?nobel) / ?population) * 1000000 AS ?nobelPerMillion)
WHERE {
  ?s rdfs:label ?countryLabel ;
     dbo:populationTotal ?population .

  SERVICE <https://api.triplydb.com/datasets/Ijjaziad/laureate-nobel/sparql> {
    ?nobel schema:birthPlace/rdfs:label ?countryName .
  }

  FILTER(STR(?countryLabel) = STR(?countryName))
}
GROUP BY ?countryLabel ?population
ORDER BY DESC(?nobelPerMillion)
LIMIT 15
```

Résultat de la requête 3 Triplydb



countryLabel	population	nobelCount	nobelPerMillion
filter ×	filter ×	filter ×	filter
Saint Lucia	179,857	2	11.1199452899
Luxembourg	647,599	2	3.0883308961
Sweden	10,549,347	30	2.8437779135
Iceland	372,899	1	2.6816912891
Norway	5,434,319	13	2.3922040646
Switzerland	8,740,472	19	2.1737956486
Austria	8,939,617	19	2.1253706954
Denmark	5,882,261	12	2.0400318857
United Kingdom	67,508,936	92	1.3627825507
Hungary	9,967,308	11	1.1036079150
Lithuania	2,750,055	3	1.0908872732
Netherlands	17,564,014	19	1.0817572794
Germany	83,369,843	84	1.0075585725
Ireland	5,023,109	5	0.9953994628
France	64,626,628	61	0.9438833788

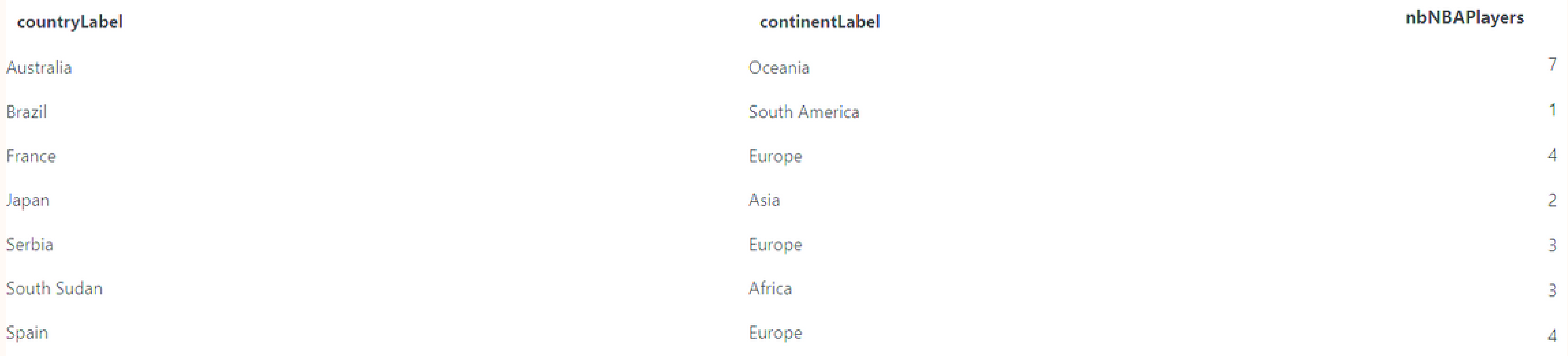
ÉTAPE 2

Requete 3

Nombre de joueurs jouant en NBA en 2023, ayant participé aux JO 2024 et ne jouant pour une équipe nationale nord-américaine

```
SELECT ?countryLabel ?continentLabel (COUNT(DISTINCT ?joAthlete) AS ?
nbNBAPlayers)
WHERE {
 ?country rdf:type dbo:Country ;
        rdfs:label ?countryLabel ;
        dbo:isoCode ?isocode ;
     dbo:continent ?continent.
 ?continent rdfs:label ?continentLabel.
 ?isocode rdfs:label ?isocodeLabel .
 SERVICE <http://localhost:3030/jo> {
    ?joAthlete a sport:Athlete ;
            dbo:sport dbr:Basketball ;
            foaf:name ?joName ;
            schema:alpha3Code ?isocodeLabel .
 }
 BIND(STR(?joName) AS ?rawName)
 BIND(STRBEFORE(?rawName, " ") AS ?lastName)
 BIND(STRAFTER(?rawName, " ") AS ?firstName)
 BIND(LCASE(CONCAT(?firstName, " ", ?lastName)) AS ?joNormalized)
 SERVICE <http://localhost:3030/nba> {
    ?nbaPlayer dcterms:temporal ?seasonData .
    ?seasonData dbo:league dbr:National_Basketball_Association ;
             dbo:season "2022-23" .
    ?nbaPlayer gn:name ?nbaPlayerName .
 }
 BIND(LCASE(STR(?nbaPlayerName)) AS ?nbaNormalized)
 FILTER(?joNormalized = ?nbaNormalized)
 FILTER(?continentLabel != "North America")
}
GROUP BY ?countryLabel ?continentLabel
ORDER BY ?countryLabel
```

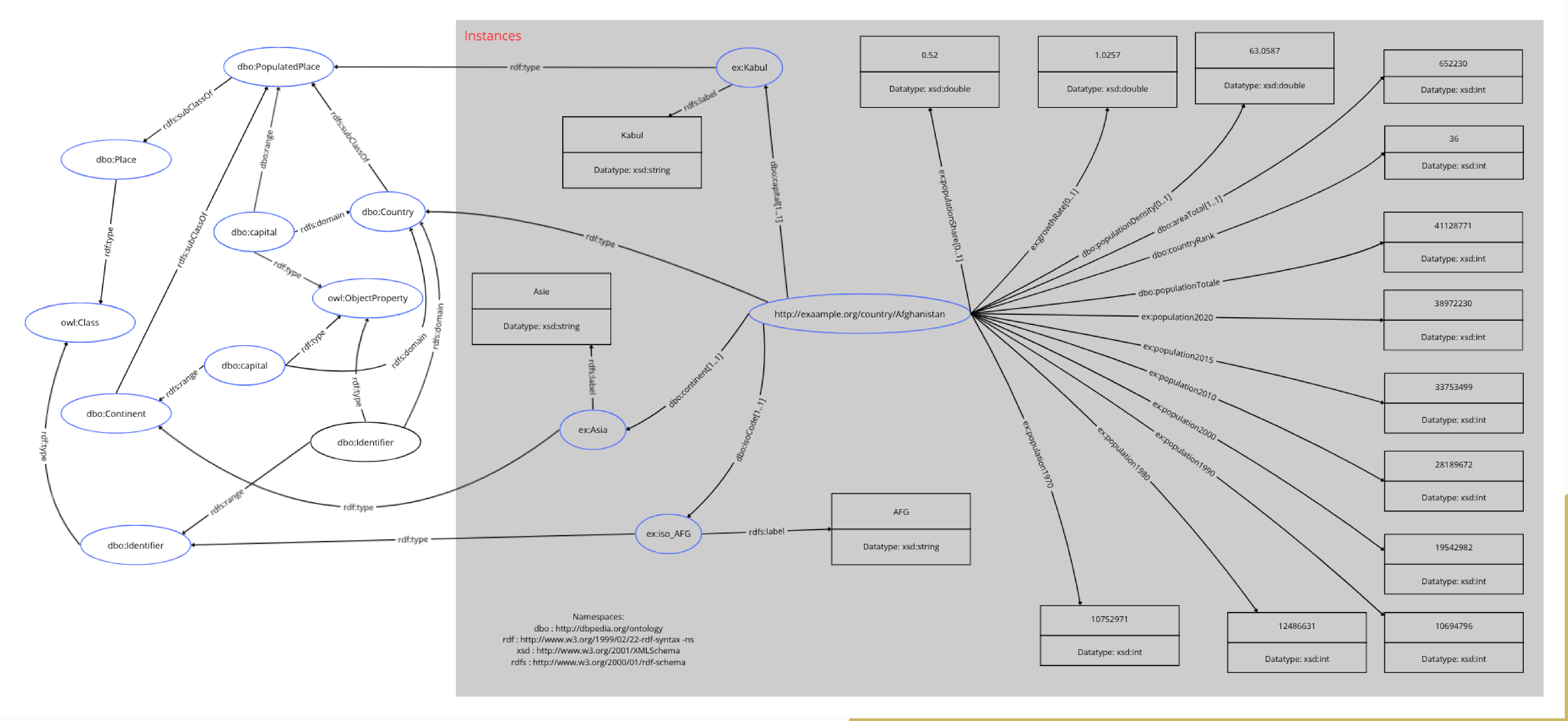
Résultat de la requête 3 Triplydb



countryLabel	continentLabel	nbNBAPlayers
Australia	Oceania	7
Brazil	South America	1
France	Europe	4
Japan	Asia	2
Serbia	Europe	3
South Sudan	Africa	3
Spain	Europe	4

ÉTAPE 3

Ontologie avec RDFS/OWL



Namespaces:
dbo : http://dbpedia.org/ontology
rdf : http://www.w3.org/1999/02/22-rdf-syntax -ns
xsd : http://www.w3.org/2001/XMLSchema
rdfs : http://www.w3.org/2000/01/rdf-schema

ÉTAPE 4.

Liaison au Cloud Linked Data.

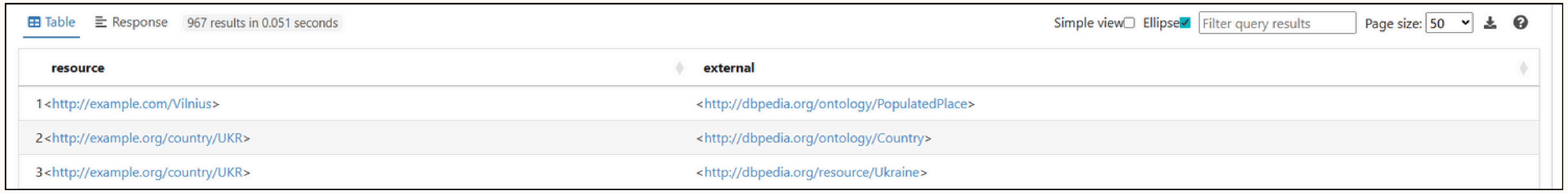
Pour réaliser la liaison de notre dataset au Web de données liées, nous avons enrichi notre modèle avec deux types de liens externes : owl:sameAs vers DBpedia et skos:exactMatch vers EU Vocabularies

```
SELECT DISTINCT ?resource ?external
WHERE {
   ?resource ?p ?external .
   FILTER(isIRI(?external))
   FILTER(
      STRSTARTS(STR(?external), "http://dbpedia.org") )
}
```

Exemple de liaison

```
<http://example.org/country/BOL> a <http://dbpedia.org/ontology/Country> ;
  rdfs:label "Bolivia" ;
  <http://dbpedia.org/ontology/areaTotal> "1098581"^^xsd:int ;
  <http://dbpedia.org/ontology/capital> <http://example.com/Sucre> ;
  <http://dbpedia.org/ontology/continent> <http://example.com/South_America> ;
  <http://dbpedia.org/ontology/countryRank> "80"^^xsd:int ;
  <http://dbpedia.org/ontology/isoCode> <http://example.com/iso_BOL> ;
  <http://dbpedia.org/ontology/populationDensity> 1.11272e+01 ;
  <http://dbpedia.org/ontology/populationTotal> "12224110"^^xsd:int ;
  <http://example.com/growthRate> 1.012e+00 ;
  <http://example.com/population1970> "4585693"^^xsd:int ;
  <http://example.com/population1980> "5736088"^^xsd:int ;
  <http://example.com/population1990> "7096194"^^xsd:int ;
  <http://example.com/population2000> "8592656"^^xsd:int ;
  <http://example.com/population2010> "10223270"^^xsd:int ;
  <http://example.com/population2015> "11090085"^^xsd:int ;
  <http://example.com/population2020> "11936162"^^xsd:int ;
  <http://example.com/populationShare> 1.5e-01 ;
  <http://www.w3.org/2002/07/owl#sameAs> <http://dbpedia.org/resource/Bolivia> ;
  <http://www.w3.org/2004/02/skos/core#exactMatch>
<http://publications.europa.eu/resource/authority/country/BOL> .
```

Échantillon de la requête de test :

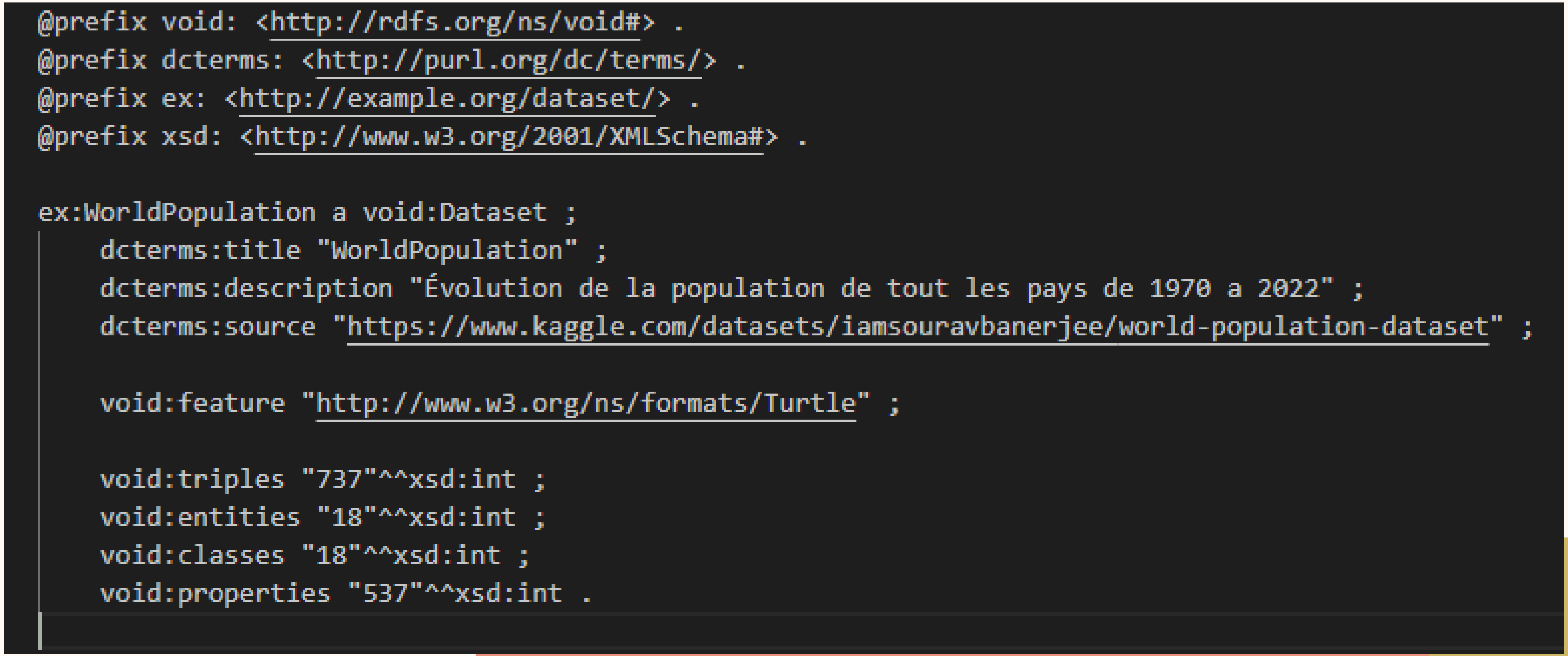


967 results in 0.051 seconds

	resource	external
1	<http://example.com/Vilnius>	<http://dbpedia.org/ontology/PopulatedPlace>
2	<http://example.org/country/UKR>	<http://dbpedia.org/ontology/Country>
3	<http://example.org/country/UKR>	<http://dbpedia.org/resource/Ukraine>

ÉTAPE 5

VOID déscription



```
@prefix void: <http://rdfs.org/ns/void#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix ex: <http://example.org/dataset/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

ex:WorldPopulation a void:Dataset ;
    dcterms:title "WorldPopulation" ;
    dcterms:description "Évolution de la population de tout les pays de 1970 a 2022" ;
    dcterms:source "https://www.kaggle.com/datasets/iamsouravbanerjee/world-population-dataset" ;

    void:feature "http://www.w3.org/ns/formats/Turtle" ;

    void:triples "737"^^xsd:int ;
    void:entities "18"^^xsd:int ;
    void:classes "18"^^xsd:int ;
    void:properties "537"^^xsd:int .
```

Merci pour votre écoute !

Lien GitHub : https://github.com/BraKann/Projet_RDF_WorldPopulation.git

Lien API: https://api.triplydb.com/Asserche/worldPopulation/sparql

Lien triplydb: https://triplydb.com/Asserche/worldPopulation/sparql

Lien annexe autres liaisons avec le G8, G5, G11, G2 et requête avec un graphe diffèrent: https://www.canva.com/design/DAG45j3Q_oE/dcHgPNTOnLMvVxsPEk2sYw/edit?utm_content=DAG45j3Q_oE&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton