

Web of Data

Master 2 - Data Science



Joe Raad

joe.raad@universite-paris-saclay.fr

Plan for this hands-on session

1. Create a simple ontology (using Protégé)
2. Convert a CSV dataset to RDF (using OntoRefine)
3. Learn how to query a large knowledge graph (Wikidata)
4. Answer a complex question over data (using SPARQL)
5. Link entities with SPARQL (using GraphDB)
6. Show results in a Web page (using HTML+Javascript)

<https://github.com/raadjoe/wod-2025-26>

What you need

- Basic knowledge of SPARQL

- GraphDB (triple store)

Installation and Tutorial:

<https://github.com/raadjoe/wod-2025-26>

- Protégé (ontology editor)

https://protege.stanford.edu/download/protege/4.3/installanywhere/Web_Installers/

Or Web Protégé: <https://webprotege.stanford.edu/>

- Any text editor with RDF syntax highlighting (e.g. Visual Studio Code)

1. Create Ontology using Protégé

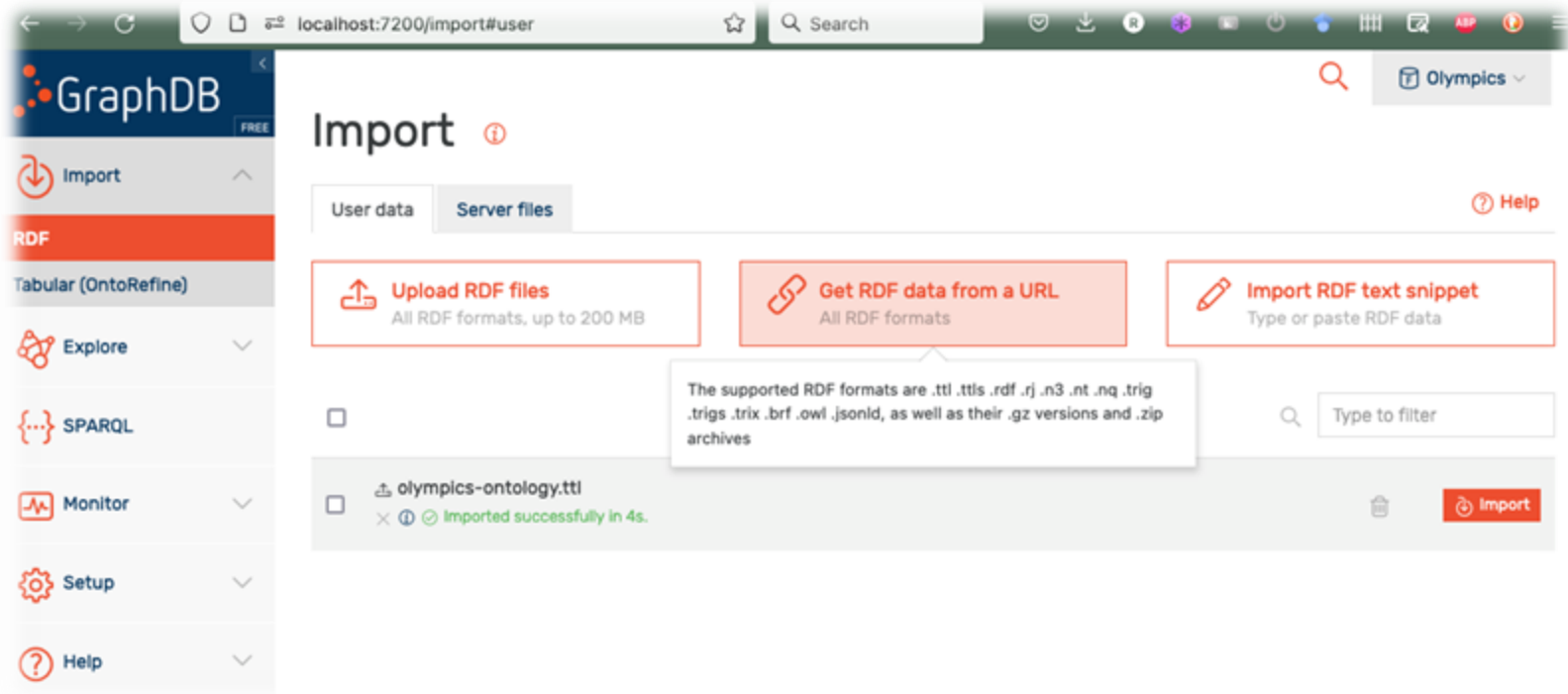
This ontology should have the necessary classes/properties to model the data of the following CSV file (medals_paris2024):

<https://github.com/raadjoe/wod-2025-26>

	country_code	country	country_long	Gold Medal	Silver Medal	Bronze Medal	Total
1	USA	United States	United States of America	40	44	42	126
2	CHN	China	People's Republic of China	40	27	24	91
3	JPN	Japan	Japan	20	12	13	45
4	AUS	Australia	Australia	18	19	16	53
5	FRA	France	France	16	26	22	64
6	NED	Netherlands	Netherlands	15	7	12	34
7	GBR	Great Britain	Great Britain	14	22	29	65
8	KOR	Korea	Republic of Korea	13	9	10	32
9	ITA	Italy	Italy	12	13	15	40
10	GER	Germany	Germany	12	13	8	33

2. Import Ontology using GraphDB

- Create a new repository in GraphDB named “Olympics”
- Import the ontology to this repository (upload RDF file)



3. Convert CSV to RDF using OntoRefine

GraphDB OntoRefine is an upgraded version of the open-source [OpenRefine](#) data transformation tool. It allows the quick mapping of any structured data to RDF.

The visual interface is optimized to guide you in choosing the proper predicates and types, defining the datatype to RDF mappings, and implementing complex transformation using OpenRefine's GREL language.

The screenshot shows the GraphDB OntoRefine web interface. The browser address bar indicates the URL is localhost:7200/ontorefine. The interface features a sidebar on the left with navigation options: Import, RDF, Tabular (OntoRefine), Explore, SPARQL, Monitor, Setup, and Help. The main area displays a table of data for a file named 'olympics-csv'. The table has 138 rows and columns for 'total_summer', 'no_games', 'no_hosted', and 'team'. A tooltip titled 'Using facets and filters' is visible, explaining how to use facets and filters to select subsets of data. The table data is as follows:

	total_summer	no_games	no_hosted	team
1.	2399	48	4	United States
2.	1413	32	1	Russia
3.	780	49	3	Great Britain
4.	671	49	2	France
5.	1304	54	2	Germany
6.	549	48	1	Italy
7.	483	48	1	Sweden
8.	476	47	0	Hungary
9.	473	19	1	China
10.	468	43	2	Australia
11.	398	41	1	Japan
12.	302	46	1	Finland
13.	301	40	0	Romania
14.	279	47	1	Canada
15.	271	42	0	Poland

4. Query Local Data using SPARQL

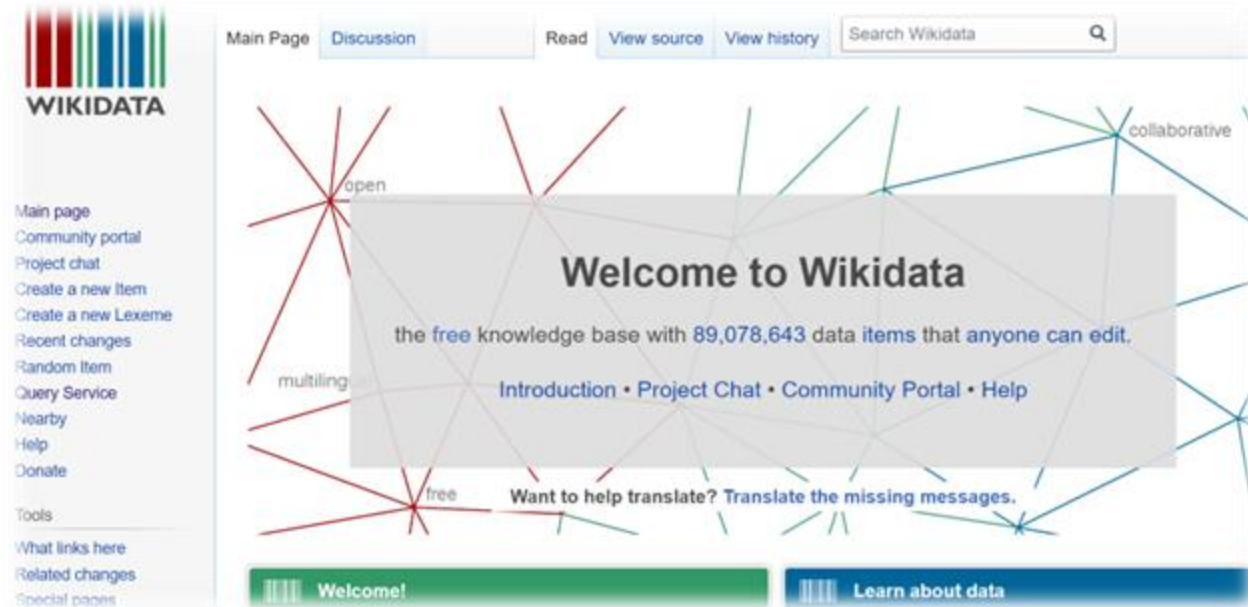
Query 1: List the 10 countries with the highest number of medals in the Paris 2024 olympics

when same number of medals, prioritise the number of gold medals, then silver and bronze

SPARQL documentation: <https://www.w3.org/TR/sparql11-query/>

4. Query Wikidata using SPARQL

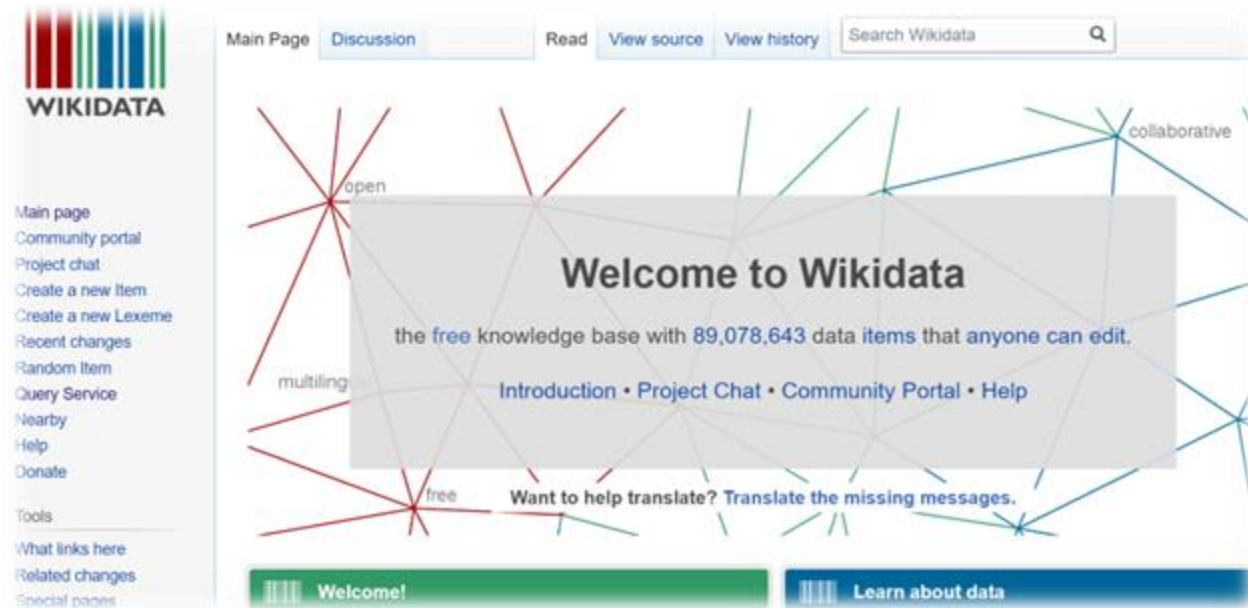
Query 2: List the population of all countries in Wikidata



- Wikidata SPARQL interface : <https://query.wikidata.org/>
- Wikidata SPARQL endpoint : <https://query.wikidata.org/bigdata/namespace/wdq/sparql>

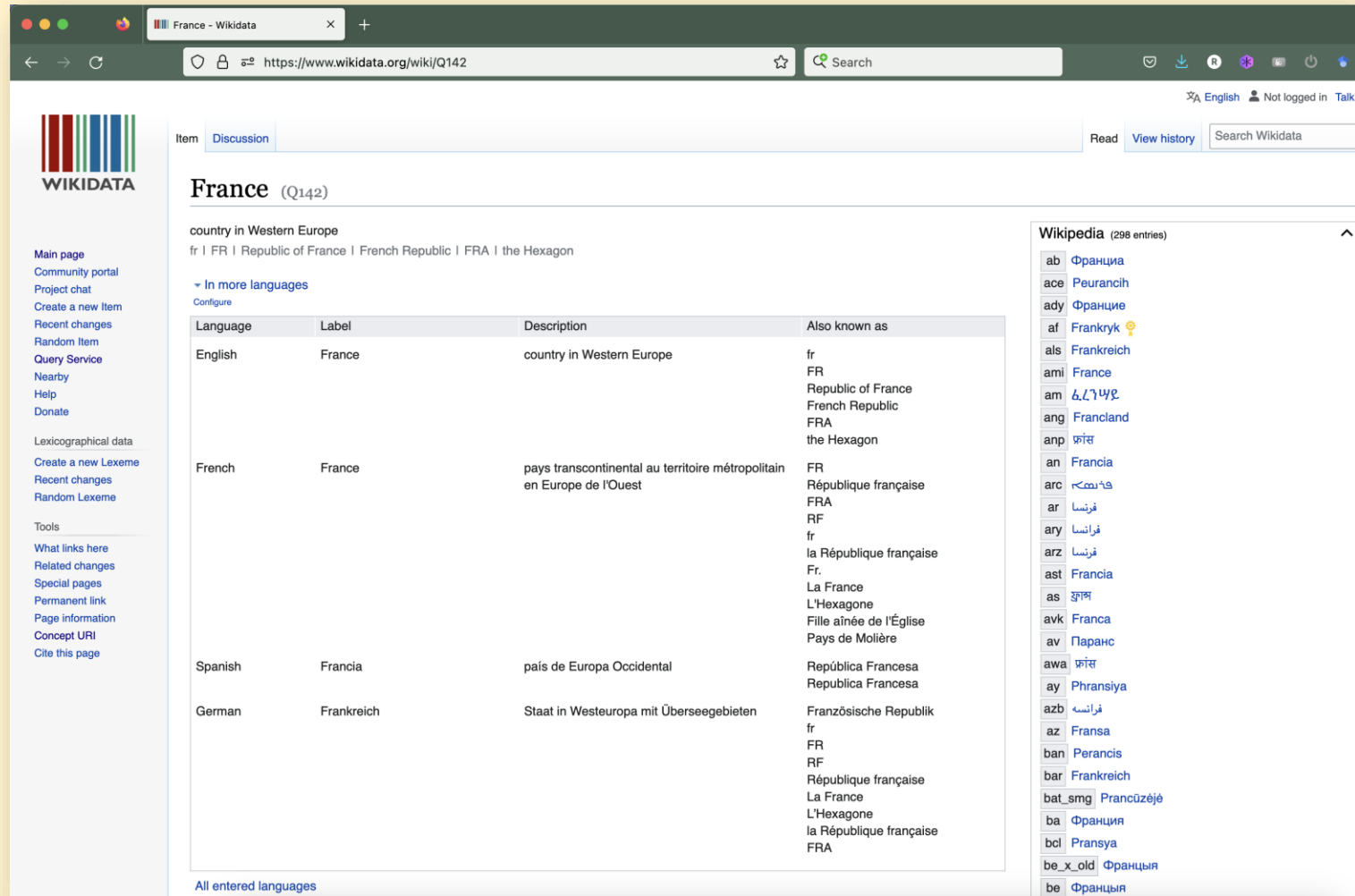
4. Query Wikidata using SPARQL

Query 3: List the population of all countries in Wikidata in 2016



- Wikidata SPARQL interface : <https://query.wikidata.org/>
- Wikidata SPARQL endpoint : <https://query.wikidata.org/bigdata/namespace/wdq/sparql>

Example : The entity France in Wikidata



France (Q142)

country in Western Europe

fr | FR | Republic of France | French Republic | FRA | the Hexagon

[In more languages](#)

[Configure](#)

Language	Label	Description	Also known as
English	France	country in Western Europe	fr FR Republic of France French Republic FRA the Hexagon
French	France	pays transcontinental au territoire métropolitain en Europe de l'Ouest	FR République française FRA RF fr la République française Fr. La France L'Hexagone Fille aînée de l'Église Pays de Molière
Spanish	Francia	país de Europa Occidental	República Francesa República Francesa
German	Frankreich	Staat in Westeuropa mit Überseegebieten	Französische Republik fr FR RF République française La France L'Hexagone la République française FRA

[All entered languages](#)

Wikipedia (298 entries)

- ab Франция
- ace Peuranchih
- ady Францие
- af Frankryk
- als Frankreich
- ami France
- am ፌዴራላዊ ዲሞክራሲያዊ ሪፐብሊክ
- ang Francland
- anp फ्रांस
- an Francia
- arc ܦܪܢܨܝܐ
- ar فرنسا
- ary فرانس
- arz فرنسا
- ast Francia
- as ফ্রান্স
- avk Franca
- av Паранс
- awa फ्रांस
- ay Phransiya
- azb فرانسه
- az Fransa
- ban Perancis
- bar Frankreich
- bat_smg Prancūzėjė
- ba Франция
- bcl Pransya
- be_x_old Францыя
- be Францыя

IRI vs URL

The screenshot shows the Wikidata page for France (Q142). The browser address bar displays the URL `https://www.wikidata.org/wiki/Q142`. An arrow points from the text "URL of the Web page of the entity France" to this address bar. Another arrow points from the text "IRI of the entity France in the Wikidata dataset" to the label "France (Q142)" in the main content area. Above the main content, the text "PREFIX wd: <<http://www.wikidata.org/entity/>> wd:Q142" is shown, with an arrow pointing to the "Q142" part of the label.

URL of the Web page of the entity France

IRI of the entity France in the Wikidata dataset

Attention!

The IRI for the France entity in the Wikidata graph is different from the URL of the web page displaying the description of the France entity (entity vs wiki)

We use the IRI in our SPARQL queries and not the URL

Language	Label	Description	Also known as
English	France	country in Western Europe	fr FR Republic of France French Republic FRA the Hexagon
French	France	pays transcontinental au territoire métropolitain en Europe de l'Ouest	FR République française FRA RF fr la République française Fr. La France L'Hexagone Fille aînée de l'Église Pays de Molière
Spanish	Francia	país de Europa Occidental	República Francesa República Francesa
German	Frankreich	Staat in Westeuropa mit Überseegebieten	Französische Republik fr FR RF République française La France L'Hexagone la République française FRA

All entered languages

Wikipedia (298 entries)

- ab Франция
- ace Peurancih
- ady Францие
- af Frankryk
- als Frankreich
- ami France
- am ፌዴራላዊ ዲሞክራሲያዊ ሪፐብሊክ
- ang Francland
- anp फ्रांस
- an Francia
- arc ܦܪܢܨܝܐ
- ar فرنسا
- ary فرانس
- arz فرنسا
- ast Francia
- as ফ্রান্স
- avk Franca
- av Паранс
- awa फ्रांस
- ay Phransiya
- azb فرانسه
- az Fransa
- ban Perancis
- bar Frankreich
- bat_smg Prancūzėjė
- ba Франция
- bcl Pransya
- be_x_old Францыя
- be Францыя

Wikidata is a set of triples

PREFIX wdt: <<http://www.wikidata.org/prop/direct/>>

wdt:P31

(equivalent to
rdf:type in Wikidata)

The screenshot shows the Wikidata page for France (Q142). The page is divided into a left sidebar with 'Statements' and a main content area. The 'Statements' section lists various categories for France, including 'instance of', 'colonial power', 'Mediterranean country', 'republic', 'country', 'OECD country', 'Western Europe', 'Pyrenees-Mediterranean Euroregion', and 'European Union'. The main content area displays details for each statement, such as 'sovereign state' with 1 reference, 'colonial power' with an end time of 1960, 'republic' with a start time of 4 October 1958, and 'country' with 0 references. A red arrow points from the text 'wdt:P31 (equivalent to rdf:type in Wikidata)' to the 'instance of' statement in the 'Statements' section. A purple arrow points from the word 'Predicate' to the 'instance of' statement. A purple arrow points from the word 'Object' to the '4 October 1958' date in the 'republic' statement. A purple arrow points from the word 'Subject' to the 'France' label in the 'Statements' section.

Statement	Value	References
instance of	sovereign state	1 reference
colonial power	end time: 1960	0 references
Mediterranean country		0 references
republic	start time: 4 October 1958	0 references
country		0 references
OECD country		0 references
Western Europe		0 references
Pyrenees-Mediterranean Euroregion		1 reference
European Union		0 references

PREFIX wd: <<http://www.wikidata.org/entity/>>

wd:Q142

Wikidata with SPARQL

Show direct objects (in green)

Statements

- instance of
 - sovereign state
 - 1 reference
 - colonial power
 - end time 1960
 - 0 references
 - Mediterranean country
 - 0 references
 - republic
 - start time 4 October 1958
 - 0 references
 - country
 - 0 references
 - OECD country
 - 0 references- part of
 - Western Europe
 - 0 references
 - Pyrenees-Mediterranean Euroregion
 - 1 reference
 - European Union
 - 0 references

Only the three direct objects appear as objects of wdt:P31 despite the presence of three other objects in this page (colonial power, Mediterranean country and OECD country)

To obtain the complete list of objects you will need to access first all the statements then access their values (next slide)

Wikidata Query Service

https://query.wikidata.org/#PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>

SELECT ?type ?typeLabel WHERE {
wd:Q142 wdt:P31 ?type .

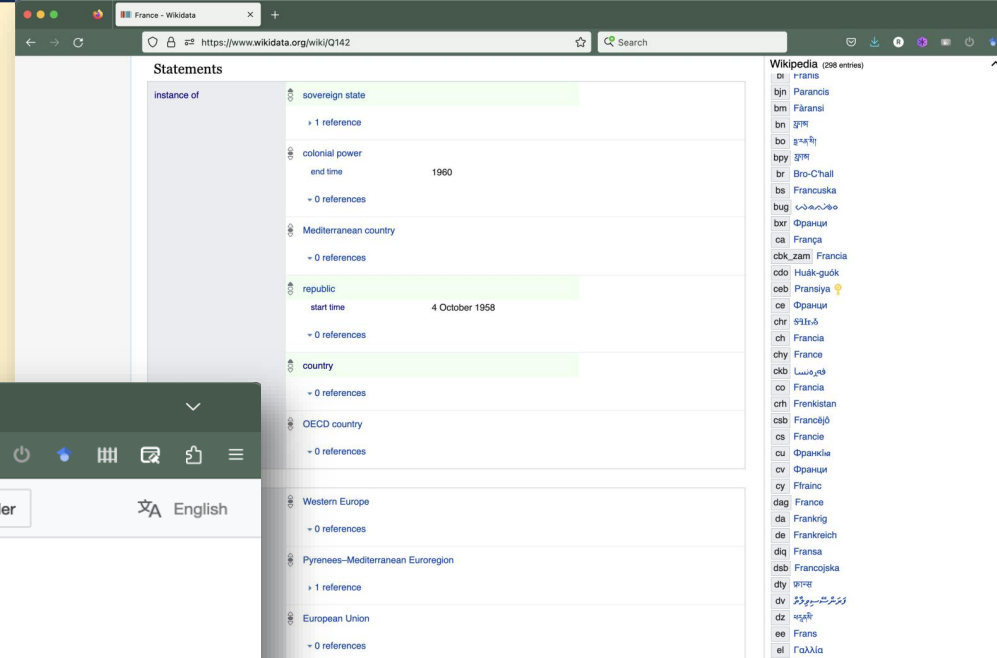
SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
SERVICE wikibase:label permet d'afficher le label d'une variable
en ajoutant "Label" après le nom de la variable dans la partie SELECT
ce SERVICE de gestion de label est spécifique à Wikidata
}

3 results in 134 ms

type	typeLabel
wd:Q6256	country
wd:Q7270	republic
wd:Q3624078	sovereign state

Wikidata with SPARQL

Display all statements



Replace **wdt:P31** with **p:P31** to get all statements.
For each statement, access its values with **ps:P31**

This is the general method for all properties in Wikidata (not just P31)

Wikidata Query Service

```
1 PREFIX ps: <http://www.wikidata.org/prop/statement/>
2 PREFIX p: <http://www.wikidata.org/prop/>
3 PREFIX wd: <http://www.wikidata.org/entity/>
4 PREFIX wdt: <http://www.wikidata.org/prop/direct/>
5
6 SELECT ?type ?typeLabel WHERE {
7   wd:Q142 p:P31 ?statement .
8   ?statement ps:P31 ?type .
9
10  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
11 }
```

6 results in 5 ms

type	typeLabel
wd:Q113489728	OECD country
wd:Q6256	country
wd:Q20181813	colonial power
wd:Q3624078	sovereign state
wd:Q51576574	Mediterranean country
wd:Q7270	republic

Wikidata with SPARQL

Display all statements and their qualifiers

The screenshot shows the Wikidata Query Service interface. The top bar includes the Wikidata logo, a search bar, and navigation links. The main area displays a SPARQL query with the following content:

```
1 PREFIX pq: <http://www.wikidata.org/prop/qualifier/>
2 PREFIX ps: <http://www.wikidata.org/prop/statement/>
3 PREFIX p: <http://www.wikidata.org/prop/>
4 PREFIX wd: <http://www.wikidata.org/entity/>
5 PREFIX wdt: <http://www.wikidata.org/prop/direct/>
6
7 SELECT ?type ?typeLabel ?startTime WHERE {
8   wd:Q142 p:P31 ?statement .
9   ?statement ps:P31 ?type ;
10             pq:P580 ?startTime # P580 = la propriété start time
11
12   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
13 }
```

Below the query, the results are displayed in a table format. The table has three columns: **type**, **typeLabel**, and **startTime**. The results show one entry for the type **republic** with the start time **4 October 1958**.

type	typeLabel	startTime
wd:Q7270	republic	4 October 1958

The screenshot shows the Wikidata page for France (Q142). The page displays various statements and their qualifiers. The statements are listed in a table with columns for the statement, the value, and the number of references. The statements include:

- sovereign state** (1 reference)
- colonial power** (end time: 1960, 0 references)
- Mediterranean country** (0 references)
- republic** (start time: 4 October 1958, 0 references)
- country** (0 references)
- OECD country** (0 references)
- Western Europe** (0 references)
- Pyrenees-Mediterranean Euroregion** (1 reference)
- European Union** (0 references)

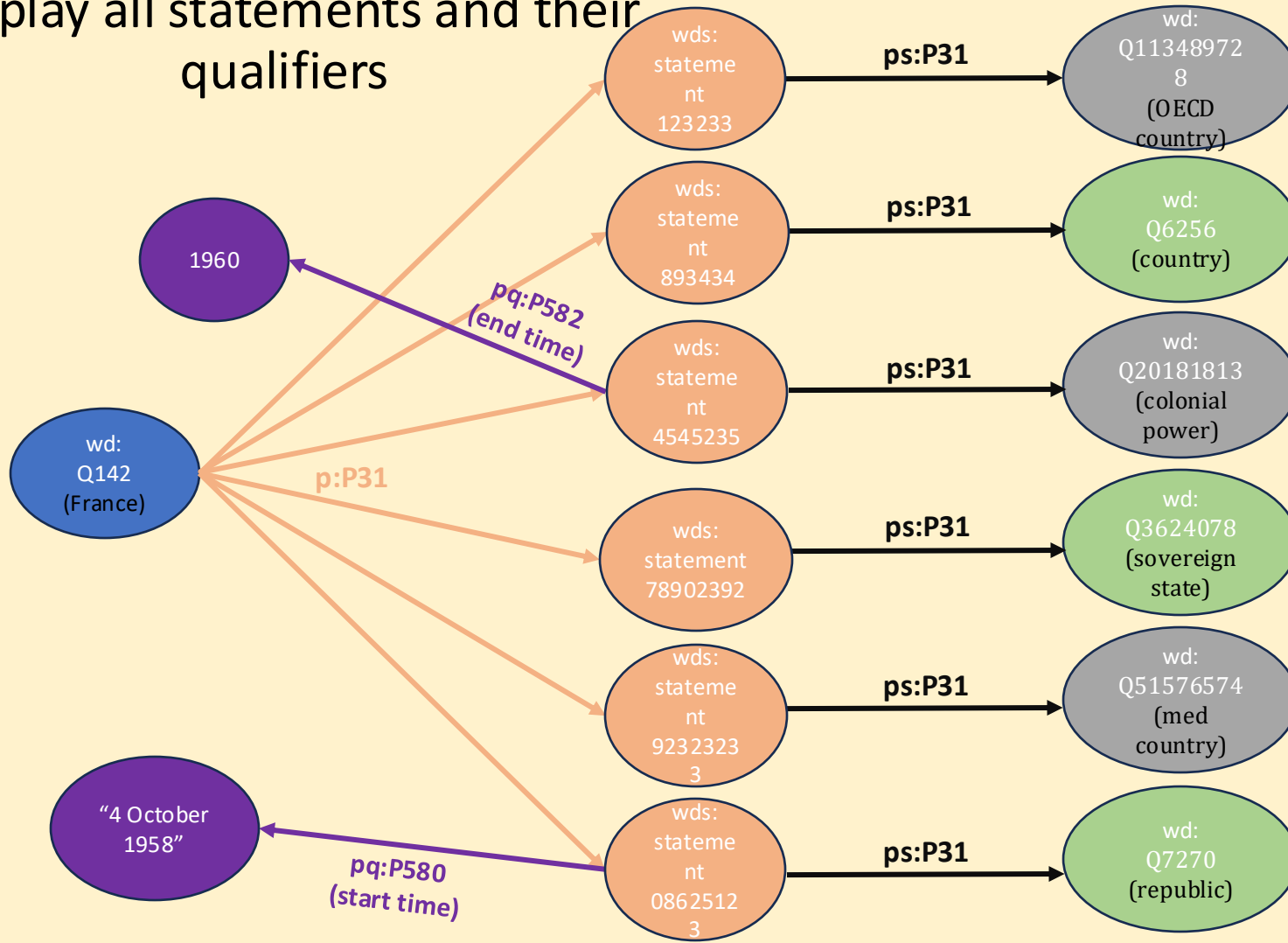
The right sidebar shows the language list for the page, including English, French, and many others.

Replace **wdt:P31** with **p:P31** to get all statements.
For each statement, access its value with **ps:P31**
and access its qualifiers with the **pq** prefix:

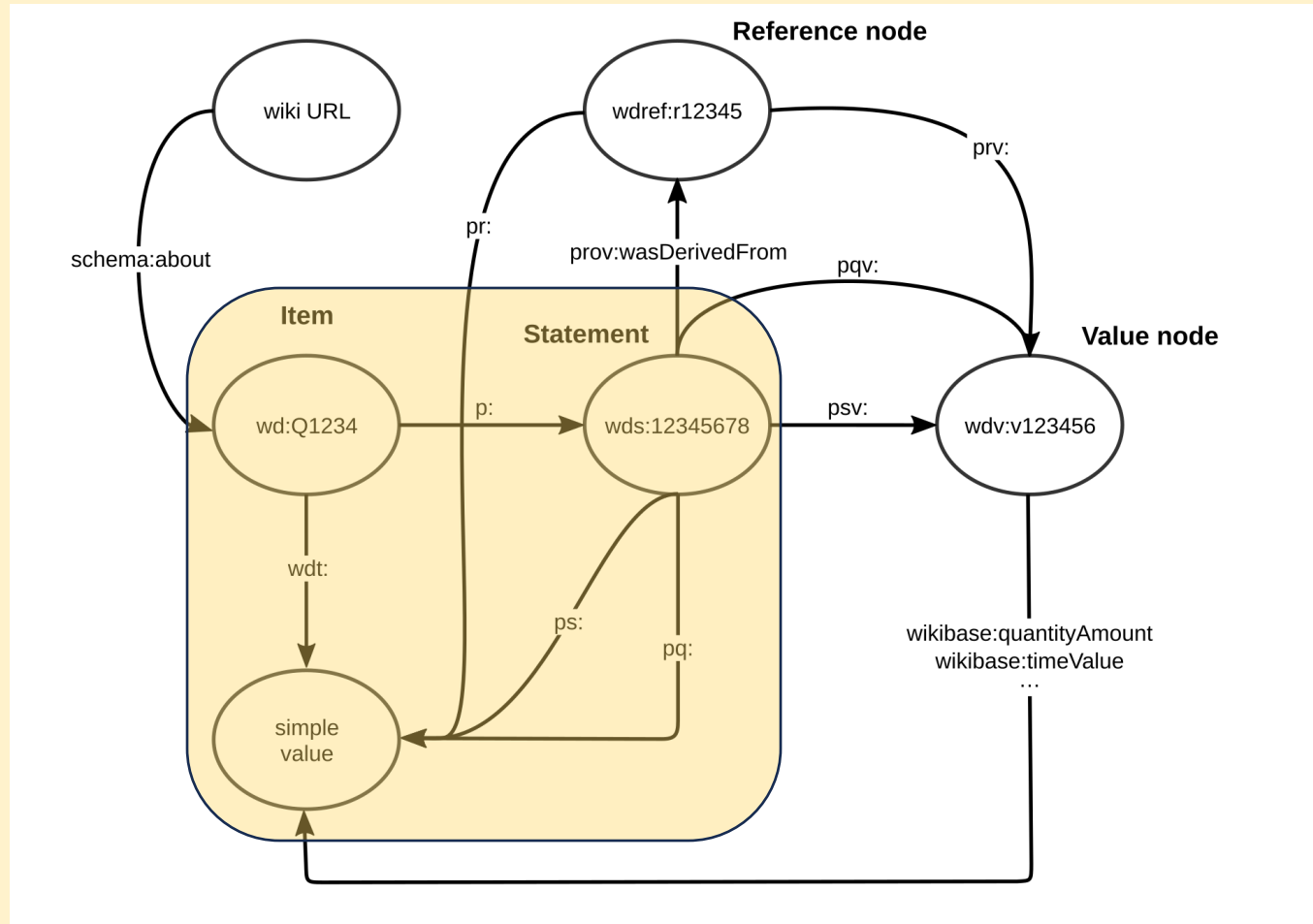
For France, the type **republic** is the only object with
the start time qualifier (P580)

Wikidata with SPARQL

Display all statements and their qualifiers



Wikidata model



5. Link entities to Wikidata

using GraphDB and SPARQL

List the 10 countries that average the highest number of medals per capita
(ratio = population / number of medals)

But information on the population of each country is missing from our dataset.

We need to retrieve the population from Wikidata by linking its countries to the countries in our local dataset

Steps using the GraphDB SPARQL editor:

1. Retrieve the population of each country in Wikidata
(federated SPARQL query using the keyword SERVICE and Wikidata's SPARQL endpoint)
2. Link entities from our KG to entities in Wikidata
(CONSTRUCT query)
3. Save links and the population to our KG
(INSERT query)
4. Query the updated KG to order countries by the ratio described above
(SELECT query)