

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

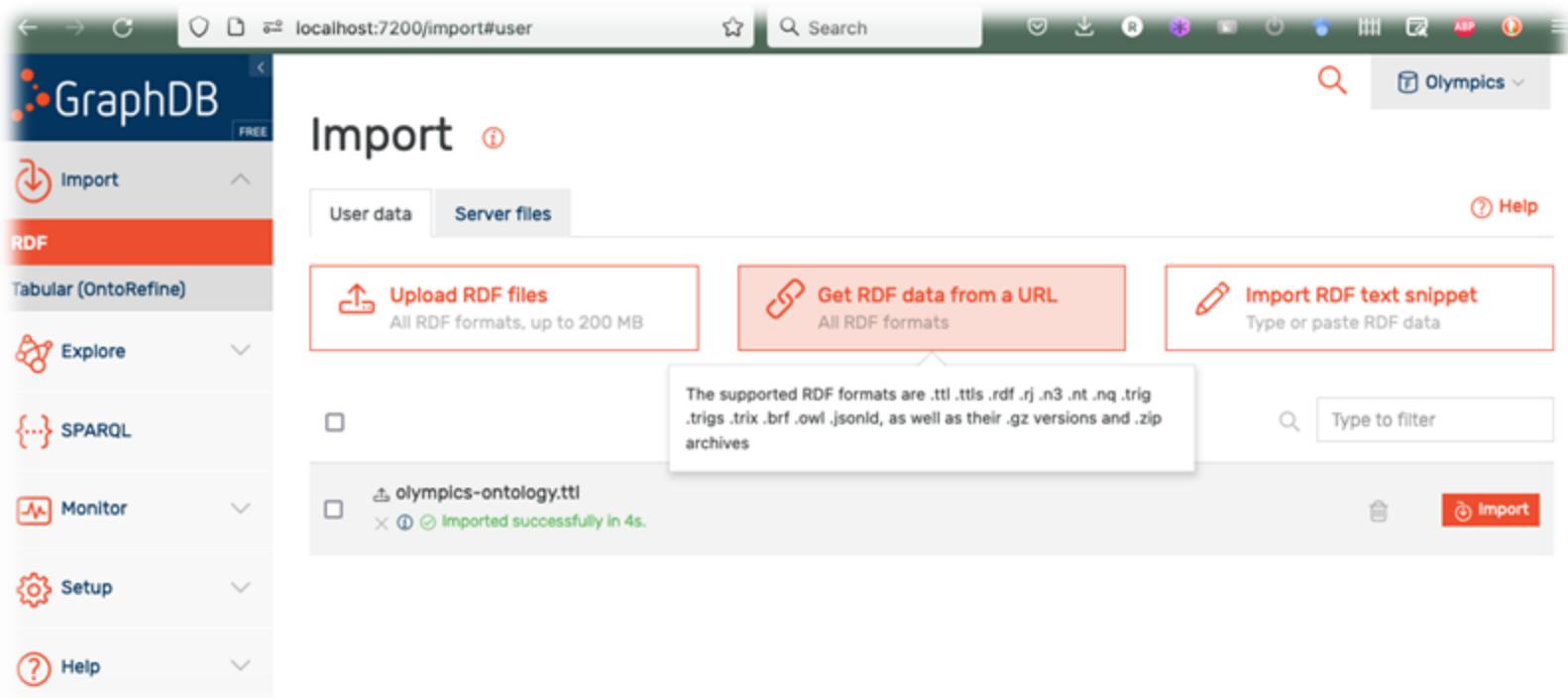
This ontology should have the necessary classes/properties to model the data of the following CSV file (medals_paris2024):

<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.

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? [Watch these screencasts](#)

All	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

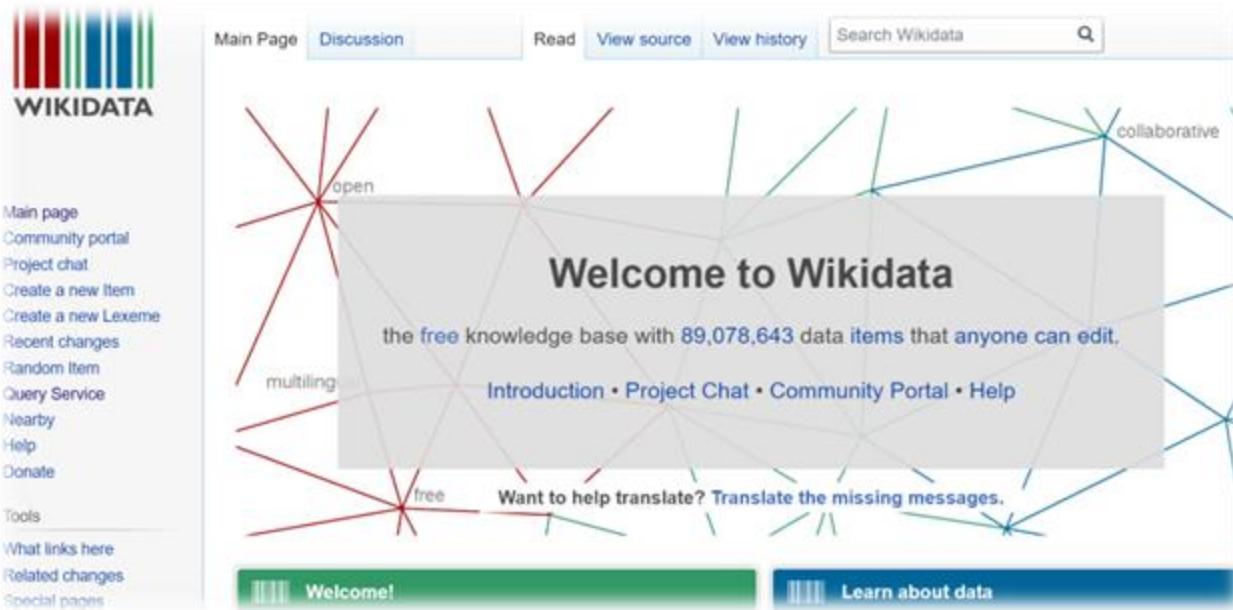
Query 1: List the 10 countries with the highest number of medals in the Paris 2024 olympics

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

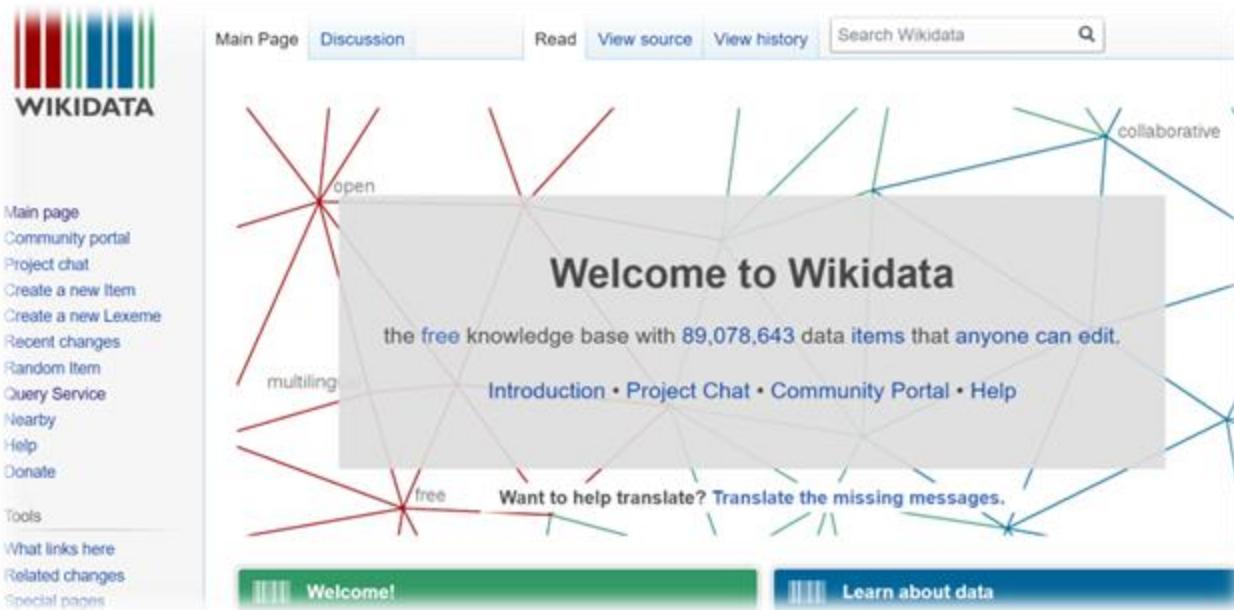
Query 2: List the population of all countries in Wikidata



- 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

The screenshot shows the Wikidata item page for France (Q142). The main content includes:

- Labels:** France (Q142)
- Type:** country in Western Europe
- Also known as:** fr, FR, Republic of France, French Republic, FRA, the Hexagon
- In more languages:** Configure
- Table of labels:** A table showing France's name in various languages.

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 ainée 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

- Wikidata sidebar:** Wikipedia (298 entries) with links to various language versions of France.

IRI vs URL

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

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

wd:Q142

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 ainée de l'Eglise Pays de Molière
Spanish	Francia	país de Europa Occidental	República Francesa República Francesa
German	Frankreich	Staat in Westeuropa mit Überseegebieten	Französische Republik fr FR République française La France L'Hexagone la République française FRA

All entered languages

Wikipedia (298 entries)

- ab Франција
- ace Peurancih
- ady Франције
- af Frankryk
- als Frankreich
- ami France
- am ԱՌԵՖ
- ang Francland
- anp ଫର୍ନେସ
- an Francia
- arc ଫର୍ନେସ
- ar فرنسا
- ary فرانسا
- ast Francia
- as ຜູາກ
- avk Franca
- av Паранс
- awa ଫର୍ନେସ
- ay Phransiya
- azب فرانس
- az Fransa
- ban Perancis
- bar Frankreich
- bat_smg Prancūzéjé
- ba Франција
- 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)



Predicate

Subject

Object

The screenshot shows the Wikidata interface for the entity Q142, which represents France. The 'Statements' section is highlighted with a red border. A red arrow points from the 'Predicate' label to the 'instance of' statement. A purple arrow points from the 'Subject' label to the 'France' entity card. A purple arrow points from the 'Object' label to the 'sovereign state' object. The 'France' entity card shows it is a sovereign state, a republic, and part of Western Europe.

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

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

wd:Q142

Wikipedia (298 entries)

fr	France
bjn	Parancis
bm	Färansli
bn	ফ্রান্স
bo	ফ্রান্স
bpy	ফ্রান্স
br	Bro-C'hall
bs	Francuska
bug	ଫ୍ରାନ୍ସ
bxr	Франци
ca	França
cbk_zam	Francia
cdo	Huák-guók
ceb	Pransiya
ce	Франци
chr	ଫ୍ରାନ୍ସ
ch	Francia
chy	France
ckb	لەندەرەق
co	Francia
crh	Frenkistan
csb	Francējō
cs	Francie
cu	Франкія
cv	Франци
cy	Frainc
dag	France
da	Frankrig
de	Frankreich
dq	Fransa
dsb	Francojska
dty	ଫାନ୍ସ
dv	ଫ୍ରାନ୍ସ
dz	ফ্ৰান୍ସ
ee	Frans
el	Γαλλία

Wikidata with SPARQL

Show direct objects (in green)

The screenshot shows the Wikidata Query Service interface. On the left, the query code is displayed:

```
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
}
```

On the right, the results table shows three rows:

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

At the bottom, it says "3 results in 134 ms".

The screenshot shows a browser window displaying the Wikidata page for France (Q142). The sidebar on the right lists language equivalents for "France". The main content area shows the following statements for France:

Statements	
instance of	sovereign state colonial power Mediterranean country republic country OECD country
part of	Western Europe Pyrenees–Mediterranean Euroregion European Union

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 with SPARQL

Display all statements

The screenshot shows the Wikidata Query Service interface. The top navigation bar includes the service name, a search bar, and various browser controls. Below the header is a toolbar with icons for examples, help, more tools, and a query builder. The main area contains a code editor with the following SPARQL query:

```
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 }
```

Below the code editor is a results table with the following data:

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

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

Western Europe

- + 0 references

Pyrenees-Mediterranean Euroregion

- + 1 reference

European Union

- + 0 references

Wikipedia (298 entries)

- fr France
- bgn Parancis
- bn ফ্রান্সি
- bo ຝຣັ້ງ
- bpy ଫ୍ରାନ୍ସ
- br Bro-Chall
- bs Francuska
- bug ຜິວ ພົມ
- bxr Франци
- ca França
- cck,zam Francia
- cdo Huásk-guók
- cob Pransiya
- ce Franci
- chr ຜິວ ພົມ
- ch Francia
- chy France
- cik لسواد
- co Francia
- crh Frankenistan
- csl Franceljib
- cs Francie
- cu Франкія
- cv Франци
- cy Ffrainc
- dag France
- da Frankrig
- de Frankreich
- dqg Fransa
- dab Francojiska
- dt ພິວ
- dv ජ්‍රාන්ස්
- dz ຜິວ ພົມ
- ee Frans
- el Ελλάδι

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 with SPARQL

Display all statements and their qualifiers

The screenshot shows the Wikidata Query Service interface. On the left, there is a sidebar with various icons for navigation and tools. The main area contains a SPARQL query editor and a results table.

SPARQL Query:

```
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 }
```

Results:

type	typeLabel	startTime
wd:Q7270	republic	4 October 1958

1 result in 5 ms

The screenshot shows a browser window displaying the details of a statement for France. The URL is https://www.wikidata.org/wiki/Q142. The page lists various qualifiers and their values for the statement.

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)
- Western Europe (~ 0 references)
- Pyrenees–Mediterranean Euroregion (1 reference)
- European Union (~ 0 references)

Wikidata (298 entries):

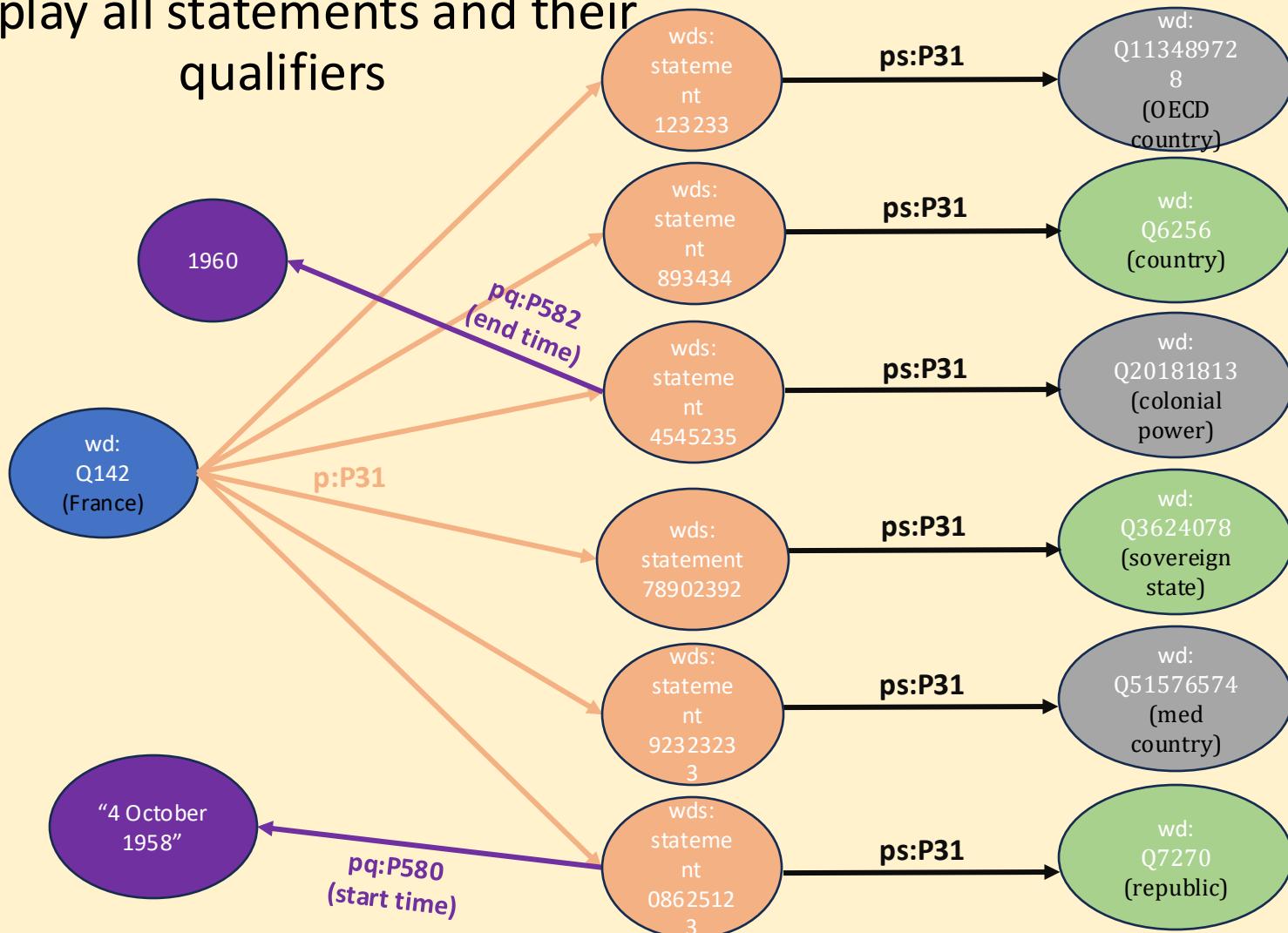
- fr:France
- bjn:Parancis
- bn:ফ্রান্স
- bo:ଫ୍ରାନ୍ସ
- bpy:ଫ୍ରାନ୍ସ
- br:Bro-Chall
- bs:Francuska
- bug:ଫ୍ରାନ୍ସ
- bzr:Франция
- ca:França
- cbk_zam:Francia
- cdo:Huakk-guk
- ceb:Pransiya
- ce:Франција
- chr:Ֆրան୍ସ
- ch:Francia
- chy:France
- ckb:لەھەز
- co:Francia
- crh:Fränkistan
- cab:Franseljö
- cs:Francie
- cu:Франкија
- cv:Франција
- cy:Ffrainc
- dag:France
- da:Frankrig
- de:Frankreich
- dq:Fransa
- dsb:Francoska
- dy:ଫ୍ରାନ୍ସ
- dv:ଫ୍ରାନ୍ସ
- dz:ফ্ৰান୍ସ
- ee:Frans
- el:Γαλλία

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:

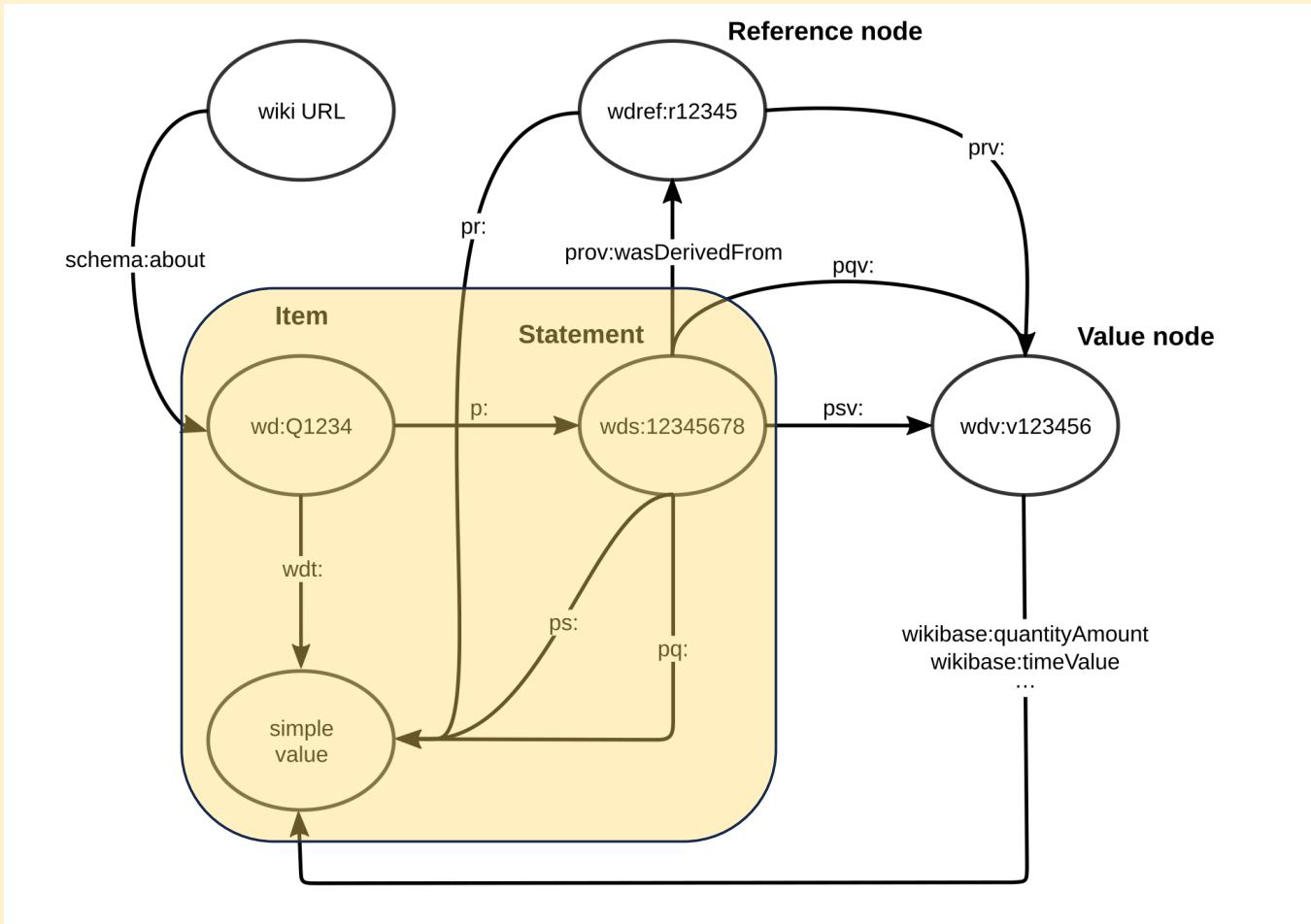
For France, the type **republic** is the only object with the start time qualifier (P580)

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)