

**KnowGraphs WinterSchool 2022**

# Serving and Querying Open Knowledge Graphs on the Web - Part 1



**Axel Polleres**

JANUARY 2022



# What I've planned for today:

- **Part 1:**
  - Interlude
  - Practical examples on querying Open KGs with SPARQL
  - Challenges/limitations of SPARQL over public endpoints
- **Part 2:**
  - Serve and query KGs for local processing – HDT
  - Addressing the SPARQL endpoint bottleneck – where are we?
    - Linked Data Fragments
    - Smart-KG
    - Wise-KG

# Standard format (RDF) & Standard Query language (SPARQL) for Graph Data



- Data representation
  - RDF (= **R**esource **D**escription **F**ramework)
    - a standard Format for publishing Graph Data on the Web.
    - Can be seen as a labeled graph
- **Querying**
  - **SPARQL**
    - a query language (similar to SQL) for RDF data



# RDF was invented for the Web, for annotating Webpages with "typed links"

← → ↻ 🏠 polleres.net/ 133%

[Curriculum Vitae](#) | [Research](#) | [Publications](#) | [Teaching](#) | [Presentations](#) | [Supervised Theses](#) | [Links](#)

**Univ.-Prof. Dr. Axel Polleres**  

Head of the [Institute for Information Business](#) at [WU Wien](#) (Vienna University of Economics & Business).

**Contact Information**

- 📞 +43-1-31336/5200 (landline)
- ✉️ sip:AxelPolleres@ekiga.net (VoIP)
- 📠 +43-1-31336/905297
- 📧 axel dot polleres arroba wu dot ac dot at
- 📍 [Vienna University of Economics and Business](#)  
Institut für Informationswirtschaft  
Gebäude D2, 3. OG  
Welthandelsplatz 1, 1020 Wien, Austria

Axel Polleres

isA

<http://xmlns.com/foaf/0.1/Person>

<http://xmlns.com/foaf/0.1/workplaceHomepage>

← → ↻ 🏠 https://www.wu.ac.at

STUDIERENDE | ALUMNI | PRESSE | MITARBEITENDE | CORPORATE RELATIONS

STUDIUM ▾ FORSCHUNG ▾

**WU**  
WIRTSCHAFTS  
UNIVERSITÄT  
WIEN VIENNA  
UNIVERSITY OF  
ECONOMICS  
AND BUSINESS

RDF uses URIs for the links to define their **Semantics**, i.e., when you look up these links, you can find out what <http://xmlns.com/foaf/0.1/workplaceHomepage> means!

You find such RDF links on my Homepage!  
Click on  on <http://polleres.net>

# RDF Triple:

- A triple of URLs

```
<http://www.polleres.net#me> <http://xmlns.com/foaf/0.1/workplaceHomepage> <http://www.wu.ac.at> .
```

- ... can be seen as an edge in a Graph:



# RDF vocabularies 1/2:

Vocabularies (collections of URIs to define meaning for Links) are identified by a common **URI prefix**:

The

**RDF Core (rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>)** and

**RDFS Schema (rdfs: <http://www.w3.org/2000/01/rdf-schema#>)**

vocabularies define basic meaning for relations such as is-A, subclasses/subproperties, (human-readable) labels, etc. according to the [RDF specification](#):

- Important URIs that used for links (in many KGs):
  - <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> (or short **rdf:type**)
  - <http://www.w3.org/2000/01/rdf-schema#label> (or short **rdfs:label**)
  - <http://www.w3.org/2000/01/rdf-schema#subPropertyOf> (or short **rdfs:subClassOf**)
  - <http://www.w3.org/2000/01/rdf-schema#subClassOf> (or short **rdfs:subPropertyOf**)
  - <http://www.w3.org/2000/01/rdf-schema#domain> (or short **rdfs:domain**)
  - <http://www.w3.org/2000/01/rdf-schema#range> (or short **rdfs:range**)

# RDF vocabularies 2/2:

## ■ Other vocabularies:

- **foaf:** Prefix: <http://xmlns.com/foaf/0.1/> ... The "Friend-of-a-friend" vocabulary models common properties of and classes relating to Persons and social relationships. **E.g.:**

### Properties:

- name
- nickname
- workplaceHomepage
- knows

### Classes:

- Agent
- Person
- Document
- Image

FOAF Vocabulary Specification 0.99

Namespace Document 14 January 2014 - *Paddington Edition*

**This version:** <http://xmlns.com/foaf/spec/20140114.html> (rdf)

**Latest version:** <http://xmlns.com/foaf/spec/> (rdf)

**Previous version:** <http://xmlns.com/foaf/spec/20100809.html> (rdf)

**Authors:** Dan Brickley, Libby Miller

**Contributors:** Members of the FOAF mailing list ([foaf-dev@lists.foaf-project.org](mailto:foaf-dev@lists.foaf-project.org)) and the wider [RDF and Semantic Web developer community](#). See [acknowledgements](#).

Copyright © 2000-2014 Dan Brickley and Libby Miller

This work is licensed under a [Creative Commons Attribution License](#). This copyright applies to the FOAF Vocabulary Specification and accompanying documentation in RDF. Regarding underlying technology, FOAF uses W3C's [RDF](#) technology, an open Web standard that can be freely used by anyone.

---

**Abstract**

This specification describes the FOAF language, defined as a dictionary of named properties and classes using W3C's RDF technology.

FOAF is a project devoted to linking people and information using the Web. Regardless of whether information is in people's heads, in physical or digital documents, or in the form of factual data, it can be linked. FOAF integrates three kinds of network: social networks of human collaboration, friendship and association; representational networks that describe a simplified view of a cartoon universe in factual terms; and *information networks* that use Web-based linking to share independently published descriptions of this interconnected world. FOAF does not compete with socially-oriented Web sites; rather it provides an approach in which different sites can tell different parts of the larger story, and by which users can retain some control over their information in a non-proprietary format.

- **schema:** Prefix: <http://schema.org/> ...
  - Classes and properties important for search engines
  - (founded by Google, Microsoft, Yahoo and Yandex)
- or domain/KG-specific vocabularies, eg.
  - **dbo:** (*DBpedia* Ontology)
  - **wd:**, **wdt:** (*Wikidata* entities and properties)



### Welcome to Schema.org

Schema.org is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond.

Schema.org vocabulary can be used with many different encodings, including RDFa, Microdata and JSON-LD. These vocabularies cover entities, relationships between entities and actions, and can easily be extended through a well-documented extension model. Over 10 million sites use Schema.org to markup their web pages and email messages. Many applications from Google, Microsoft, Pinterest, Yandex and others already use these vocabularies to power rich, extensible experiences.

Founded by Google, Microsoft, Yahoo and Yandex, Schema.org vocabularies are developed by an open **community** process, using the [public-schemaorg@w3.org](mailto:public-schemaorg@w3.org) mailing list and through [GitHub](#).

A shared vocabulary makes it easier for webmasters and developers to decide on a schema and get the maximum benefit for their efforts. It is in this spirit that the founders, together with the larger community have come together – to provide a shared collection of schemas.

# RDF Syntaxes – A simple RDF file:

simple1.nt in NTriples Syntax :

```
<http://www.example.org/klaus> <http://xmlns.com/foaf/0.1/knowns> <http://www.example.org/karl> .  
<http://www.example.org/klaus> <http://xmlns.com/foaf/0.1/nickname> "Niki" .  
<http://www.example.org/alice> <http://xmlns.com/foaf/0.1/knowns> <http://www.example.org/bob> .  
<http://www.example.org/alice> <http://xmlns.com/foaf/0.1/knowns> <http://www.example.org/karl> .  
<http://www.example.org/alice> <http://xmlns.com/foaf/0.1/name> "Alice Wonderland" .  
<http://www.example.org/karl> <http://xmlns.com/foaf/0.1/name> "Karl Mustermann" .  
<http://www.example.org/karl> <http://xmlns.com/foaf/0.1/knowns> <http://www.example.org/joan> .  
<http://www.example.org/bob> <http://xmlns.com/foaf/0.1/name> "Robert Mustermann" .  
<http://www.example.org/bob> <http://xmlns.com/foaf/0.1/nickname> "Bobby" .
```



# RDF Syntaxes – A simple RDF file:

[simple1.ttl](#) in [Turtle](#) (Terse RDF Language) Syntax is a bit more readable:

```
# using the FOAF vocabulary, see http://xmlns.com/foaf/spec/
```

```
@prefix : <http://www.example.org/> .
```

```
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
```

```
:klaus foaf:knows :karl .
```

```
:klaus foaf:nickname "Niki".
```

```
:alice foaf:knows :bob .
```

```
:alice foaf:knows :karl .
```

```
:alice foaf:name "Alice Wonderland" .
```

```
:karl foaf:name "Karl Mustermann" .
```

```
:karl foaf:knows :joan.
```

```
:bob foaf:name "Robert Mustermann" .
```

```
:bob foaf:nickname "Bobby" .
```

# RDF Syntaxes – A simple RDF file:

[simple1.ttl](#) in [Turtle](#) (Terse RDF Language) Syntax is a bit more readable –

Turtle Syntax also allows some **shortcuts** to **group Triples with common subjects**:

```
# using the FOAF vocabulary, see http://xmlns.com/foaf/spec/
```

```
@prefix : <http://www.example.org/> .
```

```
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
```

```
:klaus foaf:knows :karl ;
```

```
    foaf:nickname "Niki".
```

```
:alice foaf:knows :bob , :karl ; foaf:name "Alice Wonderland" .
```

```
:karl foaf:name "Karl Mustermann" ; foaf:knows :joan.
```

```
:bob foaf:name "Robert Mustermann" ; foaf:nickname "Bobby" .
```

***Note: We will need Turtle Syntax for querying RDF data!***

# Standards have lead to big open KGs...

- ... some of which available on the Web
- ... queryable via SPARQL endpoints!



1,101,215,718 triples/edges

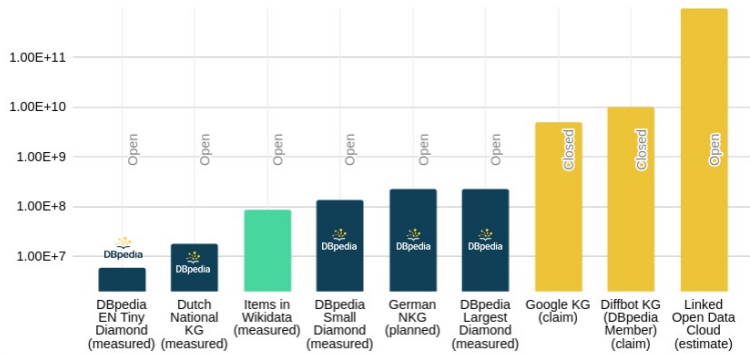


13,602,048,837 triples/edges

- plus useful convenience tools:
  - <http://prefix.cc/> ... find out common URI prefixes for formulating queries
  - <http://yasgui.triply.cc/> ... really nice frontend for querying SPARQL endpoints, e.g. DBpedia
  - <https://query.wikidata.org/> ... really nice frontend specifically for querying Wikidata
  - Plus tons of APIs (e.g. Python, R packages, etc.)

# Standards have lead to big open KGs... ... plus (even bigger?) closed KGs

- ... plus growing interest of Enterprises in KGs
- Success stories of mainly monolithic (but huge) Knowledge Graphs rather than a network of Linked small KGs:
  - <https://www.slideshare.net/Frank.van.Harmelen/adoption-of-knowledge-graphs-late-2019>



<https://www.dbpedia.org> 2021



Open KGs (April 2021)	
DBpedia	~4.58m ent
Yago4	~50m ent
Wikidata	~93m ent


N. Noy, Y. Gao, A. Jain, A. Narayanan, A. Hogan et al.: Knowledge Graphs. Co.

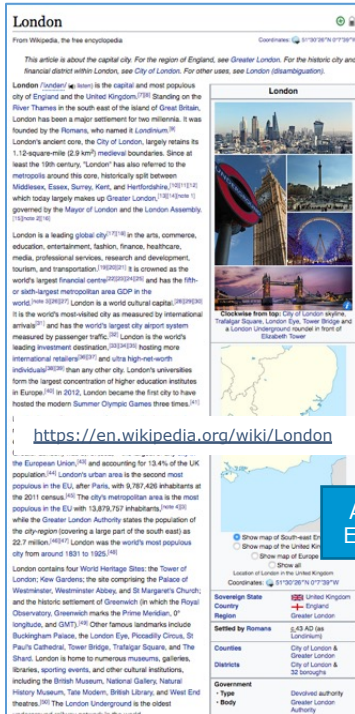


# How do we query these KGs in practice?

- ... SPARQL!

# RDF used in practice on the Web: DBpedia - a "Database-version" of Wikipedia:

- E.g. from 



**London**  
From Wikipedia, the free encyclopedia

This article is about the capital city. For the region of England, see Greater London. For the historic city and financial district within London, see City of London. For other uses, see London (disambiguation).

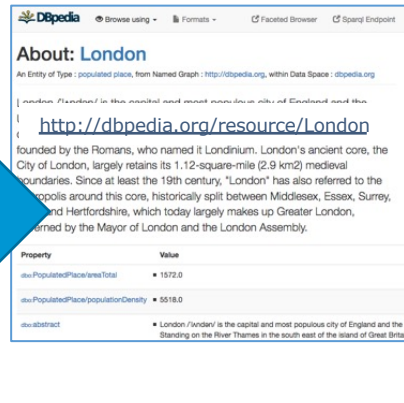
London (/ˈlɒndən/) is the capital and most populous city of England and the United Kingdom.<sup>[1][2]</sup> Standing on the River Thames in the south east of the island of Great Britain, London has been a major settlement for two millennia. It was founded by the Romans, who named it Londinium.<sup>[3]</sup> London's ancient core, the City of London, largely retains its 1.12-square-mile (2.9 km<sup>2</sup>) medieval boundaries. Since at least the 19th century, "London" has also referred to the metropolis around this core, historically split between Middlesex, Essex, Surrey, Kent, and Hertfordshire,<sup>[4][5][6][7]</sup> which today largely makes up Greater London,<sup>[8]</sup> which is governed by the Mayor of London and the London Assembly.<sup>[9][10]</sup>

London is a leading global city<sup>[11][12]</sup> in the arts, commerce, education, entertainment, fashion, finance, healthcare, media, professional services, research and development, tourism, and transportation.<sup>[13][14][15]</sup> It is crowned as the world's largest financial centre<sup>[16][17][18]</sup> and has the fifth- or sixth-largest metropolitan area GDP in the world.<sup>[19][20][21]</sup> London is a world cultural capital,<sup>[22][23]</sup> hosting more international relations<sup>[24][25]</sup> and ultra-high-net-worth individuals<sup>[26][27]</sup> than any other city. London universities form the largest concentration of higher education institutes in Europe.<sup>[28]</sup> In 2012, London became the first city to have hosted the modern Summer Olympic Games three times.<sup>[29]</sup>

<https://en.wikipedia.org/wiki/London>

Automatic Extractors

- One of the central datasets of the Linked Open Data-Cloud
- RDF extracted from Wikipedia-Infoboxes
- You can use a language called SPARQL endpoint (roughly: SQL for RDF) to do **structured queries** over RDF:
  - „Cities in the UK with more than 1M population“:



**About: London**  
An Entity of Type: populated place, from Named Graph: <http://dbpedia.org>, within Data Space: [dbpedia.org](http://dbpedia.org)

<http://dbpedia.org/resource/London>

Property	Value
<a href="#">db:PopulatedPlace/areaTotal</a>	• 1572.0
<a href="#">db:PopulatedPlace/populationDensity</a>	• 5518.0
<a href="#">db:abstract</a>	• London ( <span><span>/<span><span>ˈ</span><span>l</span><span>ɒ</span><span>n</span><span>d</span><span>ən</span></span>/</span></span> ) is the capital and most populous city of England and the U Kingdom on the River Thames in the south east of the island of Great Britain, founded by the Romans, who named it Londinium. London's ancient core, the City of London, largely retains its 1.12-square-mile (2.9 km <sup>2</sup> ) medieval boundaries. Since at least the 19th century, "London" has also referred to the metropolis around this core, historically split between Middlesex, Essex, Surrey, and Hertfordshire, which today largely makes up Greater London, which is governed by the Mayor of London and the London Assembly.

## Structured queries (SPARQL):

<http://yasgui.org/short/UVOyhX8ft>

```
PREFIX : <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX yago: <http://dbpedia.org/class/yago/>

SELECT DISTINCT ?city ?pop WHERE {
    ?city a yago:City108524735 .
    ?city dbo:country :United_Kingdom.
    ?city dbo:populationTotal ?pop
}

FILTER ( ?pop > 1000000 )
```

# RDF used in practice on the Web: Another Open Knowledge Graph: Wikidata

- slightly different idea than DBpedia:
  - a Wikimedia foundation project itself
  - put simply: "replace factual data within Wikipedia by a (graph) Database"
- Wikidata can also be queried as RDF with SPARQL!



# Let's learn some SPARQL with Wikidata

- “Simple” surface query:

## Which cities in the UK have more than 1M people?

```
SELECT DISTINCT ?city WHERE {  
  ?city wdt:P31/wdt:P279* wd:Q515.  
  ?city wdt:P1082 ?population .  
  ?city wdt:P17 wd:Q38 .  
  FILTER (?population > 1000000) }
```

instance of (P31)  
that class of which this subject is a particular example and member. (Subject typically an individual member with Proper Name label.) Different from P279 (subclass of).

subclass of (P279)  
all instances of these items are instances of those items; this item is a class (subset) of that item. Not to be confused with Property:P31 (instance of).

city (Q515)  
large and permanent human settlement

population (P1082)  
number of people inhabiting the place; number of people of subject

country (P17)  
sovereign state of this item  
United Kingdom (Q145)  
country in Europe

- What's this?



# Let's learn some SPARQL with Wikidata

- You can try out the queries on <http://query.wikidata.org/>

<https://www.wikidata.org/entity/Q41176> (wd:Q41176) ... Building  
<http://www.wikidata.org/prop/direct/P31> (wdt:P31) ... instanceOf

**Triple Patterns (TPs)**: Try this query for

"Give me 10 buildings"

<https://w.wiki/4TAP>

# Let's learn some SPARQL with Wikidata

- You can try out the queries on <http://query.wikidata.org/>

<https://www.wikidata.org/entity/Q41176> (wd:Q41176) ... Building  
<http://www.wikidata.org/prop/direct/P31> (wdt:P31) ... instanceOf

**Basic Graph patterns (BGPs)**: "Join" between edges/triples:

"Give me 10 buildings **in Austria**"

<https://w.wiki/4TAY>

# Let's learn some SPARQL with wikidata

- You can try out the queries on <http://query.wikidata.org/>

<https://www.wikidata.org/entity/Q41176> (wd:Q41176) ... Building  
<http://www.wikidata.org/prop/direct/P31> (wdt:P31) ... instanceOf

**UNION** between patterns:

Try this query for "Give me 10 buildings in **Austria or Germany**"

<https://w.wiki/4TAf>

# Let's learn some SPARQL with wikidata

- You can try out the queries on <http://query.wikidata.org/>

<https://www.wikidata.org/entity/Q41176> (wd:Q41176) ... Building  
<http://www.wikidata.org/prop/direct/P31> (wdt:P31) ... instanceOf

**FILTERS** (similar to WHERE conditions in SQL):

"Give me **the German labels of** 10 buildings in Austria or Germany"

<https://w.wiki/4TAK>

# Let's learn some SPARQL with wikidata

- You can try out the queries on <http://query.wikidata.org/>

<https://www.wikidata.org/entity/Q41176> (wd:Q41176) ... Building  
<http://www.wikidata.org/prop/direct/P31> (wdt:P31) ... instanceOf

**OPTIONAL** (similar to OUTER JOIN in SQL):

"Give me the German labels of 10 buildings in Austria **and their architect (if available)**"

<https://w.wiki/4TAn>

# Full details of SPARQL and many more examples:

- <https://www.w3.org/TR/sparql11-query/>
- Supported by various modern graph databases.

# What I've planned for today:

- **Part 1:**
  - Interlude
  - Practical tutorial on querying Open KGs with SPARQL
  - ***(some) Challenges/limitations of SPARQL over public endpoints***
- **Part 2:**
  - Serve and query KGs for local processing – HDT
  - Addressing the SPARQL endpoint bottleneck – where are we?
    - Linked Data Fragments
    - Smart-KG
    - Wise-KG

# Challenge 1: Often, you also need to deal with *contextualized* information

■ E.g. from



**Rome**  
From Wikipedia, the free encyclopedia

For other uses, see **Rome (disambiguation)**.

**Rome** (Latin and Italian: *Roma* [ˈroma]) is the capital city and a special comune of Italy (named *Comune di Roma Capitale*). Rome also serves as the capital of the Lazio region. With 2,872,800 residents in 1,285 km<sup>2</sup> (496.1 sq mi),<sup>[1]</sup> it is also the country's most populated comune. It is the fourth most populous city in the European Union by population within city limits. It is the centre of the Metropolitan City of Rome, which has a population of 4,365,725 residents, thus making it the most populous metropolitan city in Italy.<sup>[2]</sup> Rome is located in the central-western portion of the Italian Peninsula, within Lazio (Latium), along the shores of the Tiber. The Vatican City (the smallest country in the world)<sup>[3]</sup> is an enclave within Rome.

https://en.wikipedia.org/wiki/Rome

reason Rome has been often defined as capital of two states.<sup>[45]</sup>

Rome's history spans 28 centuries. While Roman mythology dates the founding of Rome at around 753 BC, the site has been inhabited for much longer, making it one of the oldest continuously occupied sites in Europe.<sup>[4]</sup> The city's early population originated from a mix of Latins, Etruscans, and Sabines. Eventually, the city successively became the capital of the Roman Kingdom, the Roman Republic and the Roman Empire, and is regarded by some as the first ever metropolis.<sup>[5]</sup> It was first called The Eternal City (Latin: *Urbs Aeterna*; Italian: *La Città Eterna*) by the Roman poet Tibullus in the 1st century BC, and the

**Rome**  
Roma  
Capital city and comune  
Roma Capitale

Clockwise from top: Basilica, Castel Sant'Angelo, Trevi Fountain

**Country**  
Italy

**Region**  
Lazio

**Government**  
- **Type** Special Comune ("Roma Capitale")  
- **Body** Roma City Council  
- **Mayor** Virginia Raggi (MSI)

**Area**  
- **Total** 1,285 km<sup>2</sup> (496.3 sq mi)

**Elevation** 21 m (69 ft)

**Population** (30 April 2018)  
- **Density** 2,236/km<sup>2</sup> (5,790/sq mi)  
- **Rank** 1st, Italy (4th in EU)

**Comune** 2,872,800<sup>[1]</sup>

**Metropolitan City** 4,365,725<sup>[2]</sup>

Automatic  
Extractor

„Cities in the **Italy** with more than 1M population“

**About: Rome**  
An Entry of Type *Comune*, from Named Graph: <http://dbpedia.org>, within Data Space: [dbpedia.org](http://dbpedia.org)

Rome (/roʊm/ RHM; Italian: *Roma* [ˈroma], Latin: *Rōma*) is a city and special ly and of the i), it is also the opulous city in the European Union by population within city limits. The Metropolitan City of Rome has a population of 4.3 million residents. The city is located in the central-western portion of the Italian Peninsula, within Lazio (Latium), along the shores of Tiber river. The Vatican City is an independent country geographically located within the city boundaries of Rome, the only existing example of a country within a city, for this reason Rome has been often defined as capit

<http://dbpedia.org/resource/Rome>

**dbp:populationAsOf** ■ 2014 (xsd:integer)

**dbp:populationBlank** ■ 2869461 (xsd:integer)

■ 4321244 (xsd:integer)

Structured query (SPARQL):

```
PREFIX : <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX yago: <http://dbpedia.org/class/yago/>

SELECT DISTINCT ?city ?pop WHERE {
  ?city a yago:City108524735 .
  ?city dbo:country :Italy .
  ?city dbo:populationTotal ?pop

  FILTER ( ?pop > 1000000 )
}
```

Doesn't work!



# Challenge 1: Wikidata as RDF ... In Wikidata even context information can be queried by SPARQL

- However, Wikidata has more complex info: (**temporal** context, **provenance**,...)
  - Rome:
  - <https://www.wikidata.org/wiki/Q220>

... Can I query that with SPARQL? Yes!

population	8,416,535±0	edit
point in time	2012	
determination method	estimation	
1 reference		
reference URL	<a href="http://www.ons.gov.uk/ons/rel/pop-estimate/population-estimates-for-england-and-wales/mid-2012/mid-2012-population-estimates-for-england-and-wales.html">http://www.ons.gov.uk/ons/rel/pop-estimate/population-estimates-for-england-and-wales/mid-2012/mid-2012-population-estimates-for-england-and-wales.html</a>	
		+ add reference
1,011,157±0		edit
point in time	1801	
determination method	census	

Wikidata Query Service

Examples Help More tools

```
1
2 SELECT ?city (min(?time) as ?year) WHERE {
3   ?city wdt:P31/wdt:P279* wd:Q515.
4   ?city wdt:P17 wd:Q38 .
5   ?city p:P1082 ?statement .
6   ?statement <http://www.wikidata.org/prop/statement/value/P1082> ?value .
7   ?statement <http://www.wikidata.org/prop/qualifier/P585> ?time .
8   ?value <http://wikiba.se/ontology#quantityAmount> ?population .
9   FILTER (?population > 1000000 )
10  } GROUP BY ?city
```

<a href="http://www.visionofbritain.org.uk/data_cube_page.jsp?data_theme=T_POP&amp;data_cube=N_TOT_POP&amp;u_id=10097836&amp;c_id=10001043&amp;add=N">http://www.visionofbritain.org.uk/data_cube_page.jsp?data_theme=T_POP&amp;data_cube=N_TOT_POP&amp;u_id=10097836&amp;c_id=10001043&amp;add=N</a>	+ add reference
1811	edit
method	census
<a href="http://www.visionofbritain.org.uk/data_cube_page.jsp?data_theme=T_POP&amp;data_cube=N_TOT_POP&amp;u_id=10097836&amp;...">http://www.visionofbritain.org.uk/data_cube_page.jsp?data_theme=T_POP&amp;data_cube=N_TOT_POP&amp;u_id=10097836&amp;...</a>	

<https://w.wiki/4rs>

# Challenge 1: Contextualized information in RDF

- no standard as of yet. State of affairs:
  - Wikidata has its own proprietary extension (cf. last slide)
  - Alternative representations/engines involve **Property Graphs**
  - ongoing work: RDF\*/SPARQL\* community group



## RDF-star and SPARQL-star

Final Community Group Report 17 December 2021

This version:

<https://www.w3.org/2021/12/rdf-star.html>

Latest published version:

<https://w3c.github.io/rdf-star/cg-spec>

Latest editor's draft:

[https://w3c.github.io/rdf-star/cg-spec/editors\\_draft.html](https://w3c.github.io/rdf-star/cg-spec/editors_draft.html)

Test suite:

<https://w3c.github.io/rdf-star/tests/>

Implementation report:

<https://w3c.github.io/rdf-star/reports/>

Previous version:

<https://w3c.github.io/rdf-star/cg-spec/2021-07-01.html>

Editors:

Olaf Hartig (Linköping University)

Pierre-Antoine Champin (ERCIM)

Gregg Kellogg (no affiliation)

Andy Seaborne (Apache Software Foundation)

Authors:

Dörthe Arndt (TU Dresden)

Jeen Broekstra (metaphacts)

Bob DuCharme (CCRI)

Ora Lassila (Amazon)

Peter F. Patel-Schneider (PARC)

Eric Prud'hommeaux (Janeiro Digital, W3C/MIT)

Ted Thibodeau, Jr. (OpenLink Software, Inc.)

Bryan Thompson (Amazon)

Acknowledgements:

James Anderson (datagraph gmbh)

Ghislain Atermezang (Mondeca)

Pavel Klinov (Stardog Union)

Bruno P. Kinoshita (Apache Software Foundation)

Holger Knublauch (TopQuadrant, Inc.)

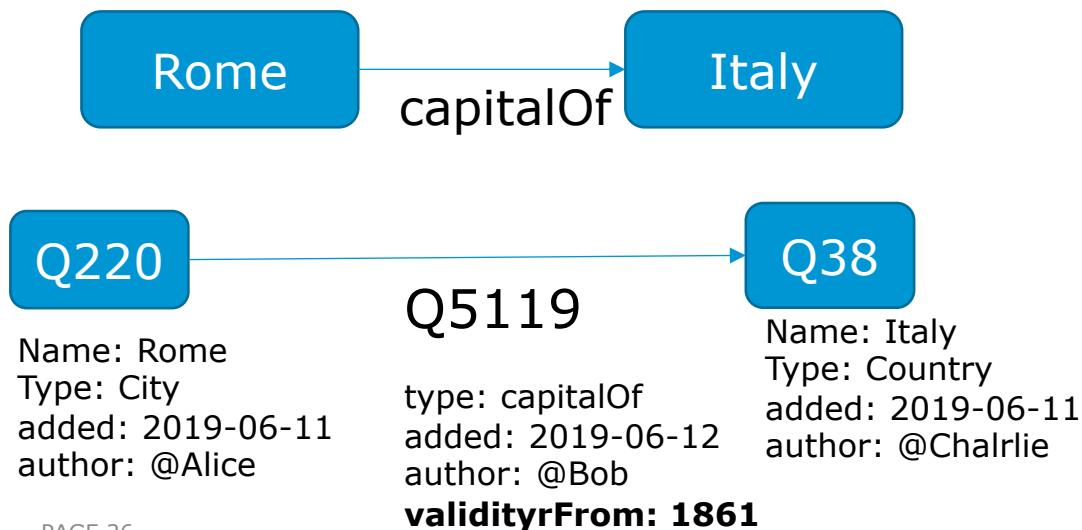
Andreas Kuckartz

Pete Rivett (agnos.ai)

William Van Woensel (Dalhousie University)

Miel Vander Sande (meemoo)

Fabio Vitali (University of Bologna)



# Challenge 2: Path queries

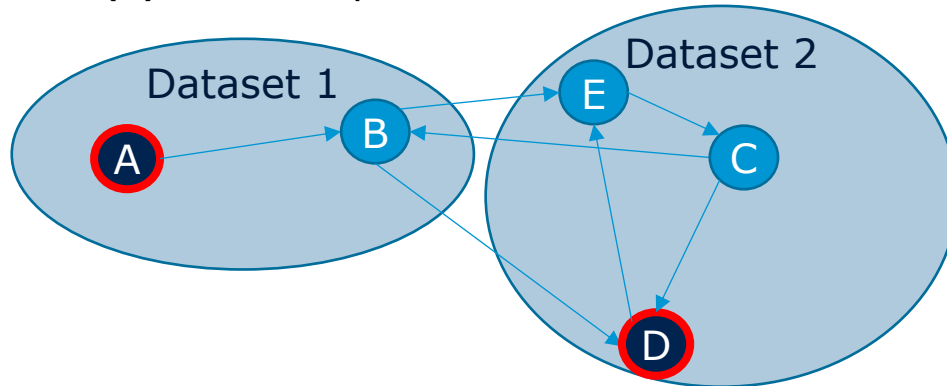
- While it is possible to do path queries in SPARQL via property path expressions, it is still not possible to **return** paths in SPARQL1.1:

```
SELECT DISTINCT ?city ?Path WHERE {  
  ?city wdt:P31/wdt:P279* wd:Q515.  
  ?city wdt:P1082 ?population .  
  ?city wdt:P17 wd:Q38 .  
  FILTER (?population > 1000000) }  
}
```

i.e.: what is the  
(shortest) path ?Path  
connecting ?city and  
wd:Q515?

# Challenge 2: Path queries – prototype solution

Common problem in graphs, not doable with SPARQL, but with extensions [1]:  
"Give me the (k) shortest paths between two nodes?"



:a :p :b.  
:b :p :d, :e.  
:c :p :b, :d.  
:d :p :e.  
:e :p :c.

```
rd2hdt.sh -rdftype turtle testgraph.ttl testgraph.hdt
```

```
hdtsparql.sh testgraph.hdt "PREFIX ppf: <java:at.ac.wu.argext.path.>  
SELECT * WHERE{ ?path ppf:topk (:a :d 2) }"
```

We solved this by extending SPARQL [1] with  
bidirectional BFS over HDT

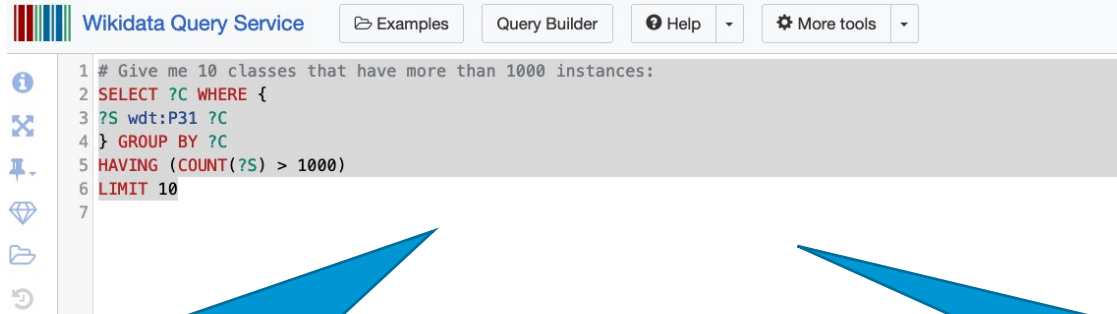
[https://bitbucket.org/vadim\\_savenkov/topk-pfn/](https://bitbucket.org/vadim_savenkov/topk-pfn/)

Open research question(s): e.g.  
But how to do this effectively in a  
Federated setting?

k=2

1. Vadim Savenkov, Qaiser Mehmood, Jürgen Umbrich, and Axel Polleres. Counting to k, or how SPARQL 1.1 could be efficiently enhanced with top k shortest path queries. In 13th International Conference on Semantic Systems (SEMANTiCS), pages 97--103, Amsterdam, the Netherlands, September 2017. ACM. [ .pdf ]

# Challenge 3: Scalability of SPARQL endpoints?



```
Wikidata Query Service [Examples] [Query Builder] [Help] [More tools]
1 # Give me 10 classes that have more than 1000 instances:
2 SELECT ?C WHERE {
3   ?S wdt:P31 ?C
4 } GROUP BY ?C
5 HAVING (COUNT(?S) > 1000)
6 LIMIT 10
7
```

Challenge 3.1: serve complex/long running queries to single users

Challenge 3.2: serve many queries to many users **concurrently**

<https://w.wiki/4mTj>

more on that in Part 2

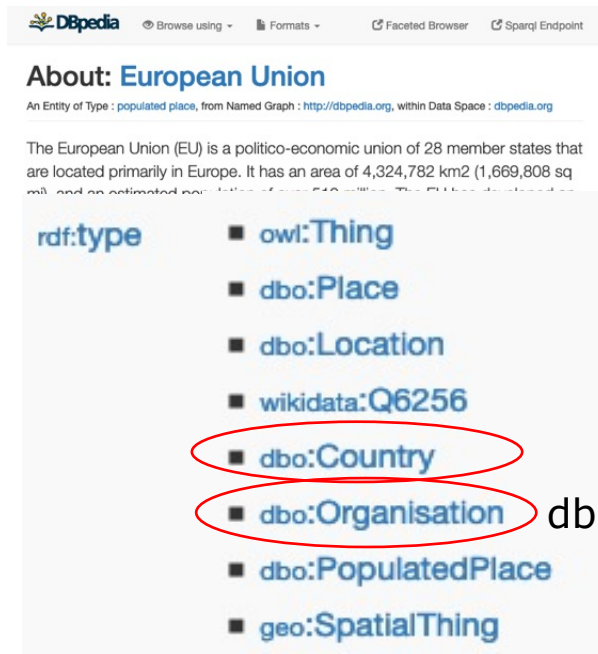
# Challenge 4: Reasoning and Inconsistencies

- A lot of work has been done in the past on (deductive reasoning over KGs) in particular to retrieve implicit answers through exploiting the **OWL** and **RDFS** semantics.
- ... e.g. by query rewriting or materialisation.
  
- However:
  - 1) existing KGs are inconsistent
  - 2) some important KGs don't use OWL and RDFS

# Challenge 4: Reasoning and Inconsistencies

## Existing KGs aren't consistent ☹ [1]

- E.g. 



DBpedia

Browse using - Formats - Faceted Browser Sparql Endpoint

### About: European Union

An Entity of Type : [populated place](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)

The European Union (EU) is a politico-economic union of 28 member states that are located primarily in Europe. It has an area of 4,324,782 km2 (1,669,808 sq mi) and an estimated population of over 500 million. The EU has developed a

**rdf:type**

- owl:Thing
- dbo:Place
- dbo:Location
- wikidata:Q6256
- **dbo:Country**
- **dbo:Organisation**
- dbo:PopulatedPlace
- geo:SpatialThing

Dbpedia Ontology:

dbo:Agent **owl:disjointWith** dbo:Place.

dbo:Country rdfs:subClassOf dbo:Place.

dbo:Organisation rdfs:subClassOf dbo:Agent.



1. Stefan Bischof, Markus Krötzsch, Axel Polleres, and Sebastian Rudolph. Schema-agnostic query rewriting in SPARQL 1.1. In *Proceedings of the 13th International Semantic Web Conference (ISWC 2014)*, Lecture Notes in Computer Science (LNCS). Springer, October 2014. [ [.pdf](#) ]

# Challenge 4: Reasoning and Inconsistencies

## important KGs don't use OWL and RDFS

- Wikidata!

```
SELECT DISTINCT ?city ?Path WHERE {
  ?city wdt:P31/wdt:P279* wd:Q515 .
  ?city wdt:P1082 ?population .
  ?city wdt:P17 wd:Q38 .
  FILTER (?population > 1000000) }
```

use "somewhat similar"  
properties:  
wdt:P31 ~ rdf:type  
wdt:P279 ~ rdfs:subClassOf

SAME  
SAME  
BUT  
DIFFERENT