

ISWS 2022

Serving and Querying Open Knowledge Graphs on the Web - Basics



Axel Polleres

JULY 2022



What I've planned for today:

- **Basics:**
 - Interlude – some words on syntax...
 - Practical SPARQL on examples querying Open KGs with SPARQL
 - Challenges/limitations of SPARQL over public endpoints
- **Bonus Material (time allowed or upon request):**
 - 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... we need to talk about syntax!



- We already mentioned... 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, common prefixes 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, common prefixes 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) 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 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/knows> <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/knows> <http://www.example.org/bob> .  
<http://www.example.org/alice> <http://xmlns.com/foaf/0.1/knows> <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/knows> <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 like RDF have lead to (really) big open KGs...

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




1,101,215,718 triples/edges

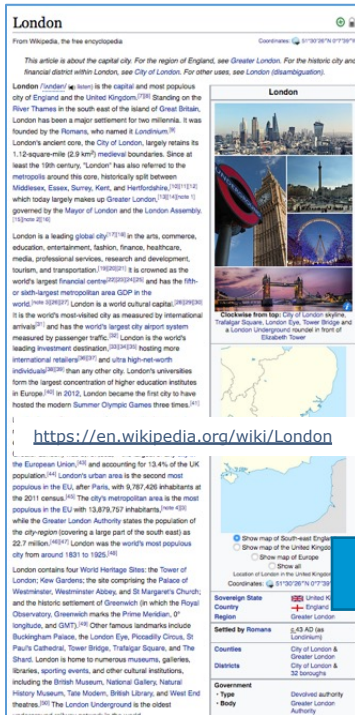


13,602,048,837 triples/edges

- 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.)

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).

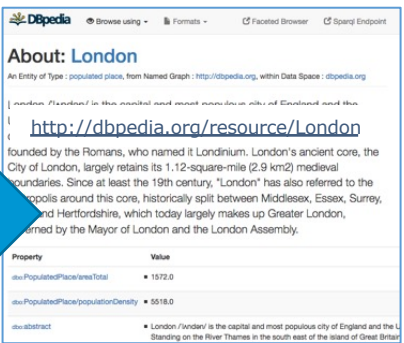
London (/ˈlɒndən/ [ⓘ]) is the capital and most populous city of England and the United Kingdom.^{[1][2]} Standing on the River Thames in the south east of the island of Great Britain, London has been a major settlement for two millennia. It was founded by the Romans, who named it Londinium.^[3] London's ancient core, the City of London, largely retains its 1.12-square-mile (2.9 km²) 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,^{[4][5][6][7]} which today largely makes up Greater London,^{[1][4]} which is governed by the Mayor of London and the London Assembly.^{[13] Nov 2016}

London is a leading global city,^{[17][18]} in the arts, commerce, education, entertainment, fashion, finance, healthcare, media, professional services, research and development, tourism, and transportation.^{[19] [20] [21]} It is crowned as the world's largest financial centre,^{[22] [23] [24] [25]} and has the fifth- or sixth-largest metropolitan area GDP in the world.^{[14][26] [27] [28]} London is a world cultural capital,^{[29] [30] [31]} and is the world's most-visited city as measured by international arrivals^[32] and has the world's largest city airport system measured by passenger traffic.^[33] London is the world's leading investment destination,^{[34] [35]} hosting more international relations^{[36] [37]} and ultra-high-net-worth individuals^{[38] [39]} than any other city. London's universities form the largest concentration of higher education institutes in Europe.^[40] In 2012, London became the first city to have hosted the modern Summer Olympic Games three times.^[41]

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

- One of the central datasets of the Linked Open Data-Cloud
- RDF extracted from Wikipedia-Infoboxes
- You can use a language called SPARQL (roughly: SQL for RDF) to do **structured queries** over RDF via Web accessible **SPARQL endpoints, e.g.** <http://dbpedia.org/sparql>
 - „Cities in the UK with more than 1M population“:

Automatic Extractors



About: London
An Entry of Type : populated place, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

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

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 km²) 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.

Property	Value
Country	Greater London
Region	Greater London
Settled by	Romans
Coordinates	51°30′37″N 0°7′39″W﻿ / ﻿51.51028°N 0.12750°W﻿ / 51.51028; -0.12750
Counties	City of London & Greater London
Districts	City of London & 32 boroughs
Government	
• Type	Devolved authority
• Body	Greater London Authority

Structured queries (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 :United_Kingdom.
    ?city dbo:populationTotal ?pop

    FILTER ( ?pop > 1000000 )
}
```

Try it on yasqui.triply.cc ... short link to the query:
https://api.triplydb.com/s/Of19_c3-e

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>

```
1 SELECT *
2 WHERE {
3     ?X wdt:P31 wd:Q41176
4 }
5 LIMIT 10
```

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>

```
1 SELECT *
2 WHERE {
3   ?X wdt:P31 wd:Q41176 .
4   ?X wdt:P17 wd:Q40 .
5 }
6 LIMIT 10
```

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:

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

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

```
1 SELECT *
2 WHERE {
3     ?X wdt:P31 wd:Q41176 .
4     { {?X wdt:P17 wd:Q40 . }
5       UNION
6       {?X wdt:P17 wd:Q183 . } }
7 }
8 LIMIT 10
```


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>

```
1 SELECT ?L
2 WHERE {
3   ?X wdt:P31 wd:Q41176 ;
4     rdfs:label ?L .
5   { {?X wdt:P17 wd:Q40 . } UNION {?X wdt:P17 wd:Q183 . } }
6   FILTER (lang(?L) = "en")
7 }
8 LIMIT 10
```

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>

```
1 SELECT ?L ?A
2 WHERE {
3   ?X wdt:P31 wd:Q41176 ;
4     rdfs:label ?L ;
5     wdt:P17 wd:Q40 .
6   FILTER (lang(?L) ="en")
7
8   OPTIONAL {?X wdt:P84 ?A }
9 }
10 LIMIT 10
```

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:

- **Basics:**
 - Interlude – some words on syntax...
 - Practical SPARQL on examples querying Open KGs with SPARQL
 - **Challenges/limitations of SPARQL over public endpoints**
- Bonus Material (time allowed or upon request):
 - Serve and query KGs for local processing – HDT
 - Addressing the SPARQL endpoint bottleneck – where are we?
 - Linked Data Fragments
 - Smart-KG
 - Wise-KG

(Some) Challenges:

- 1. Challenge 1:** How to query Contextualized Data (e.g. temporal, provenance,...)
- 2. Challenge 2:** What about real graph queries (paths, paths across distributed data)?
- 3. Challenge 3:** Scalability (and costs of hosting) SPARQL endpoints
- 4. Challenge 4:** Mixing querying and reasoning (how? how to scale?)
- 5. Challenge 5:** Sustainability of RDF and SPARQL resources

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 km² (496.1 sq mi),^[1] 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.^[2] 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)^[3] is an enclave within Rome.

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

reason Rome has been often defined as capital of two states.^[45]

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.^[4] 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.^[5] 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

Automatic Extractors

Country	Italy
Region	Lazio
Government	
Type	Special Comune ("Roma Capitale")
Body	Roma City Council
Mayor	Virginia Raggi (MSI)
Area	
Total	1,285 km ² (496.3 sq mi)
Elevation	21 m (69 ft)
Population (30 April 2018)	
Rank	1st, Italy (4th in EU)
Density	2,236/km ² (5,790/sq mi)
Comune	2,872,800 ^[1]
Metropolitan City	4,365,725 ^[2]

„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

Rome (/roʊm/ RHM; Italian: *Roma* [ˈroma], Latin: *Rōma*) is a city and special city and of the i), it is also the

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

populous 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

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	Q18	8,416,535±0	edit
	point in time	2012	
	determination method	estimation	
	+ 1 reference		
	reference URL	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	
			+ add reference
	Q18	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
```

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

<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

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 Atemezing (Mondeca)

Pavel Klinov (Stardog Union)

Bruno P. Kinoshita (Apache Software Foundation)

Holger Knublauch (TopQuadrant, Inc.)

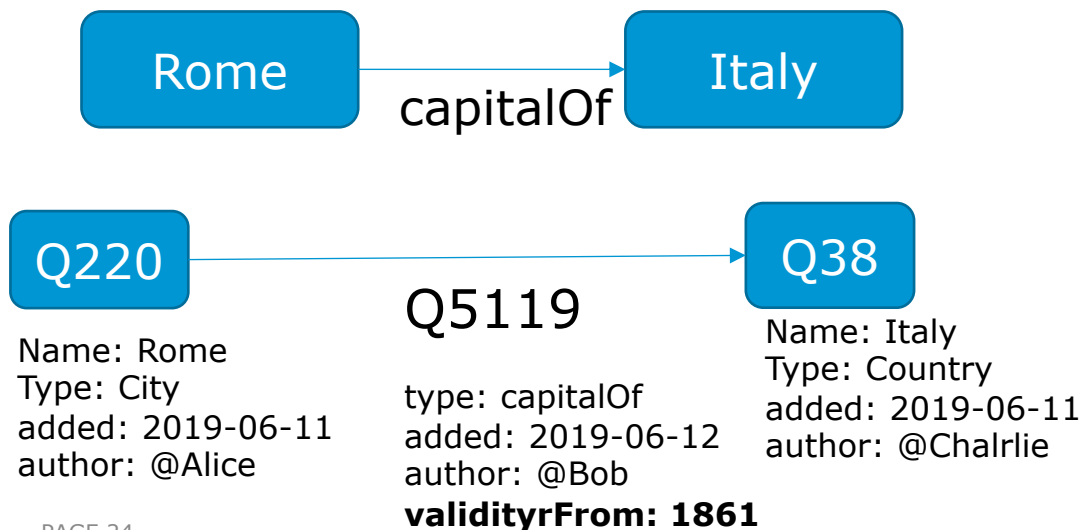
Andreas Kuckartz

Pete Rivett (agnos.ai)

William Van Woensel (Dalhousie University)

Miel Vander Sande (meemoo)

Fabio Vitali (University of Bologna)



Challenge 2: Path queries

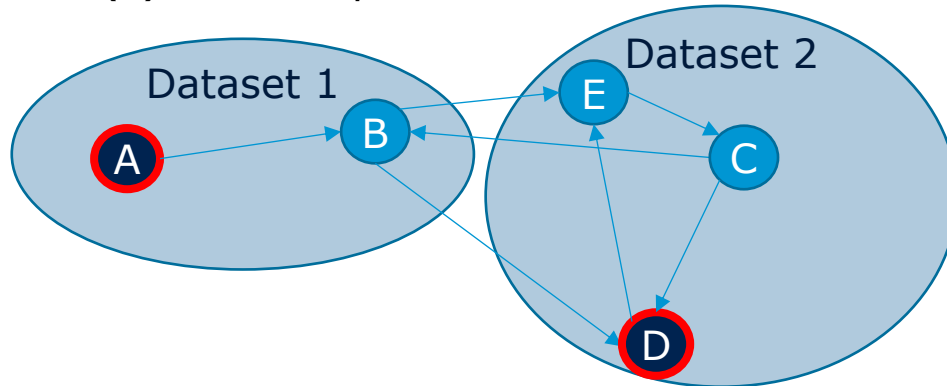
- While it is possible to do path queries in SPARQL via property path expressions, it is still not possible to **return** paths in SPARQL1.1:

```
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/

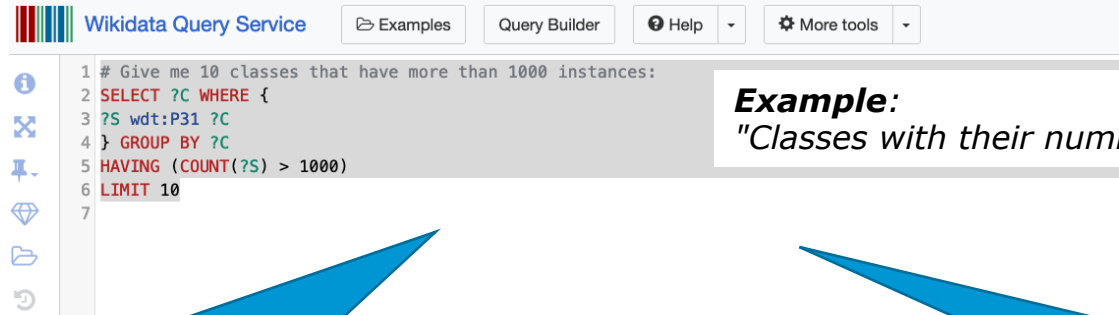
[Savenkov et al, SEMANTICS 2017]

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

k=2

Still interesting question also in (Graph)DB Theory...
regarding entailments, coverage of queries in such settings.

Challenge 3: Scalability of SPARQL endpoints?



The screenshot shows the Wikidata Query Service interface. At the top, there are navigation buttons for 'Examples', 'Query Builder', 'Help', and 'More tools'. Below these is a text area containing a SPARQL query. To the right of the query, there is a text box with an example query result.

```
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
```

Example:
"Classes with their number of instances"

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>

[Fernández et al. 2013, JWS][Vergborough et al. 2016, JWS]


[Fernández et al. 2020, WebConf][Azzam et al. 2022, WebConf]

Challenge 3: Scalability of SPARQL endpoints?

AVAILABILITY




Last update: Sun, 17 Apr 2022 10:51:32 GMT

► Description:

 Operating normally

18.23% (103/565) endpoints are available

Observation 2:

- Serving SPARQL endpoints sustainably is too hard/expensive?
- Linked Data has rather evolved into a :    **few, but huge, popular (Open) Knowledge Graphs:**

<https://sparql.es.ai.wu.ac.at/availability> , operating since 2012 [Käfer et al. 2012, LDOW]

Challenge 3: Scalability of SPARQL endpoints? **WU**

What's the problem?

https://iccl.inf.tu-dresden.de/web/Wikidata_SPARQL_Logs/

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

Wikidata Query Service

```
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
```

Query timeout limit reached

SPARQL-QUERY: queryStr=SELECT ?C WHERE {
?S wdt:P31 ?C
} GROUP BY ?C
HAVING (COUNT(?S) > 1000)
LIMIT 10

Challenge 1: serve complex/long running queries to single users

Interval	First day	Last day	Queries
Interval 1	2017-06-12	2017-07-09	59,547,909

Challenge 2: serve many queries to many users *concurrently*

[Fernández et al. 2013, JWS][Vergborough et al. 2016, JWS]

[Fernández et al. 2020, WebConf][Azzam et al. 2022, WebConf]

Challenge 4: Reasoning and Inconsistencies

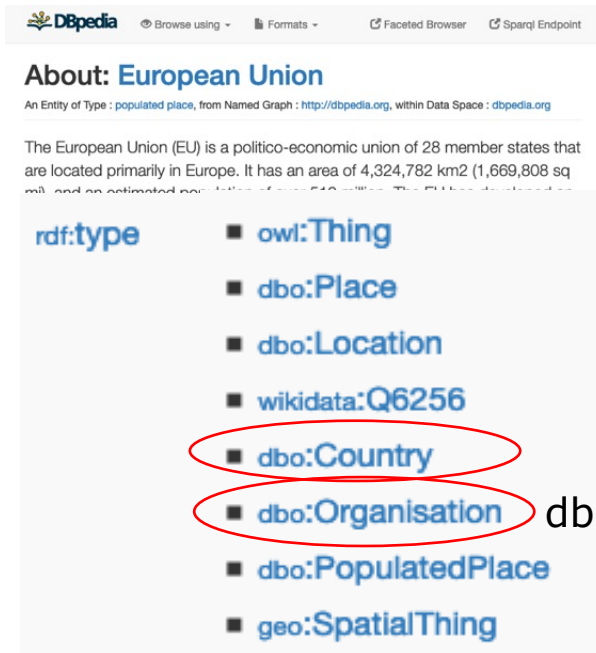
- A lot of work has been done in the past on (deductive reasoning over KGs) in particular to retrieve implicit answers through exploiting the **OWL** and **RDFS** semantics.
- ... e.g. by query rewriting or materialisation.

- 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 510 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.

[Bischof et al. 2014]



Challenge 4: Reasoning and Inconsistencies

important KGs don't use OWL and RDFS

- Wikidata!

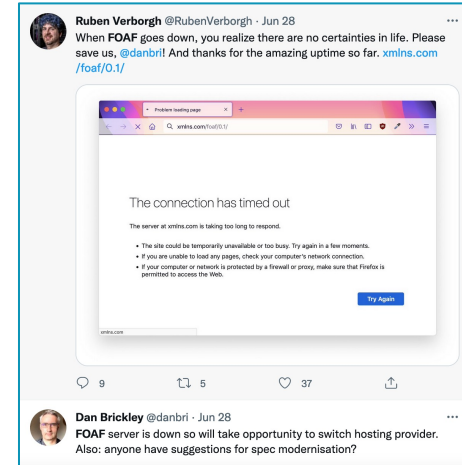
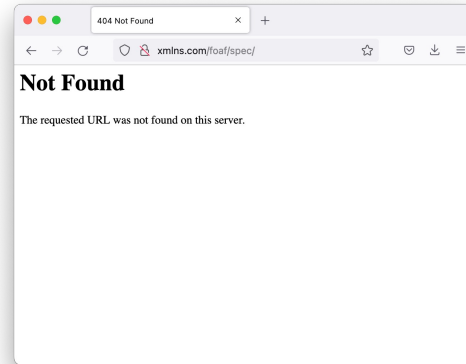
```
SELECT DISTINCT ?city ?Path WHERE {  
  ?city wdt:P31/wdt:P279* wd:Q515.  
  ?city wdt:P1082 ?population .  
  ?city wdt:P17 wd:Q38 .  
  FILTER (?population > 1000000) }
```

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

SAME
SAME
BUT
DIFFERENT

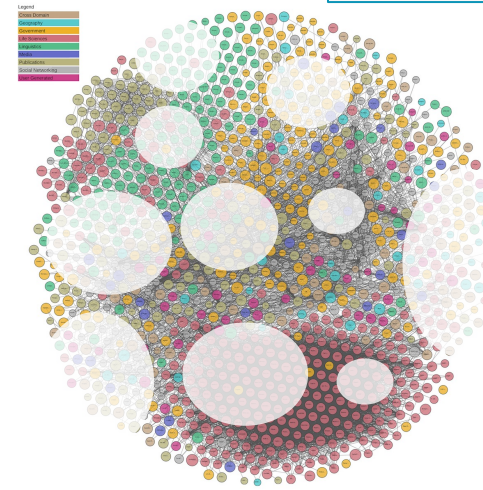
Challenge 5: Sustainability of RDF and OWL resources on the Web...

■ Vocabularies? FOAF:



■ Linked Open Data? [Polleres et al.2020, SWJ]:

"Among the mentioned 5435 resources in the 1281 "LOD"-tagged datasets on datahub.io, there are only 1917 resources URLs that could be dereferenced."



What I've planned for today:

- **Basics:**
 - Interlude – some words on syntax...
 - Practical SPARQL on examples querying Open KGs with SPARQL
 - **Challenges/limitations of SPARQL over public endpoints**
- *Bonus Material (time allowed or upon request):*
 - *Serve and query KGs for local processing – HDT*
 - *Addressing the SPARQL endpoint bottleneck – where are we?*
 - *Linked Data Fragments*
 - *Smart-KG*
 - *Wise-KG*



(subjective, highly incomplete list of) References

1. *Armin Haller, Axel Polleres, Daniil Dobriy, Nicolas Ferranti, Sergio José Rodríguez Méndez: An Analysis of Links in Wikidata. ESWC 2022: 21-38*
2. *Axel Polleres, Maulik Rajendra Kamdar, Javier D. Fernández, Tania Tudorache, and Mark A. Musen. A more decentralized vision for linked data. 11(1):101--113, January 2020. Semantic Web Journal (SWJ) 11(1):101-113, 2020.*
3. *Armin Haller, Javier D. Fernández, Maulik R. Kamdar, and Axel Polleres. What are links in linked open data? a characterization and evaluation of links between knowledge graphs on the web. ACM Journal of Data and Information Quality (JDIQ), 2(2):1–34, 2020.*
4. *Tobias Käfer, Jürgen Umbrich, Aidan Hogan, and Axel Polleres. Towards a dynamic linked data observatory. In WWW2012 Workshop on Linked Data on the Web (LDOW2012), Lyon, France, April 2012.*
5. *Javier D. Fernández, Miguel A. Martínez-Prieto, Claudio Gutiérrez, Axel Polleres, and Mario Arias. Binary RDF Representation for Publication and Exchange (HDT). Journal of Web Semantics (JWS), 19(2), 2013.*
6. *Ruben Verborgh, Miel Vander Sande, Pieter Colpaert, Sam Coppens, Erik Mannens, Rik Van de Walle: Web-Scale Querying through Linked Data Fragments. LDOW 2014*
7. *Ruben Verborgh, Miel Vander Sande, Olaf Hartig, Joachim Van Herwegen, Laurens De Vocht, Ben De Meester, Gerald Haesendonck, Pieter Colpaert: Triple Pattern Fragments: A low-cost knowledge graph interface for the Web. J. Web Semant. 37-38: 184-206 (2016)*
8. *O. Hartig and C. B. Aranda. 2016. Bindings-Restricted Triple Pattern Fragments. In ODBASE 2016. 762–779*
9. *S. Clearly-Strange, Is Happening At ISWS. (2022), check what you can find about <http://dbpedia.org/resource/Bertinoro> https://w3id.org/framester/isws2022_th.owl#hauntedBy some entity, at this SPARQL endpoint: <http://etna.istc.cnr.it/framester2/sparql> and get all the info you are able to find about it!*
10. *Amr Azzam, Javier D. Fernández, Maribel Acosta, Martin Beno, and Axel Polleres. SMART-KG: Hybrid shipping for SPARQL querying on the web. In The Web Conference 2020, Taipei, Taiwan, 2020.*
11. *Amr Azzam, Christian Aebeloe, Gabriela Montoya, Ilkcan Keles, Axel Polleres, and Katja Hose. WiseKG: Balanced Access to Web Knowledge Graphs. In The Web Conference 2021, pages 1422---1434, Ljubljana, Slovenia, 2021. ACM / IW3C2.*
12. *T. Minier, H. Skaf-Molli, and P. Molli. 2019. SaGe: Web Preemption for Public SPARQL Query Services. In WWW 2019. 1268–1278.*