



# Querying the Wikidata Knowledge Graph

Peter Haase

LDBC Meeting, Barcelona

19.3.2015

metaphacts.

# metaphacts GmbH

- Founded 10/2014
- Currently team of four
- Based in Walldorf
- Partnership with Systap

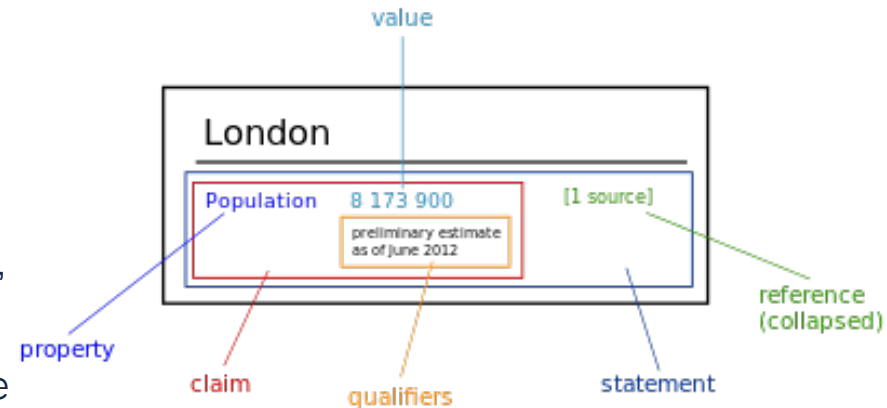


# Portfolio: Software, solutions, services for knowledge graphs

- Storing and querying of knowledge graphs
  - Scalable databases for big graphs, building on bigdata/Blazegraph
  - High-performance graph analytics, based on MapGraph / GAS
  - Light-weight reasoning with large-scale knowledge graphs
- Creation and curation of knowledge graphs
  - Semi-automatic creation of knowledge graphs from existing sources
  - Data integration and ontology-based data access
  - Collaborative management of knowledge graphs
- Application development utilizing knowledge graphs
  - Rapid development of end-user oriented applications
  - Visualization of knowledge graphs, semantic search
  - Mobile applications & augmented and virtual reality

# Wikidata

- **Collecting structured data.** Unlike the Wikipedias, which produce encyclopedic articles, Wikidata collects data, in a structured form.
- **Collaborative.** The data in Wikidata is entered and maintained by Wikidata editors, who decide on the rules of content creation and management in Wikidata
- **Integration with Wikipedia:** serving language labels and data
- **Multilingual.** Editing, consuming, browsing, and reusing the data is fully multilingual. Data entered in any language is immediately available in all other languages.
- **A secondary database.** Wikidata can record not just statements, but also their sources, thus reflecting the diversity of knowledge available and supporting the notion of verifiability.
- **Free.** The data in Wikidata is published under the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)



# Wikidata in RDF

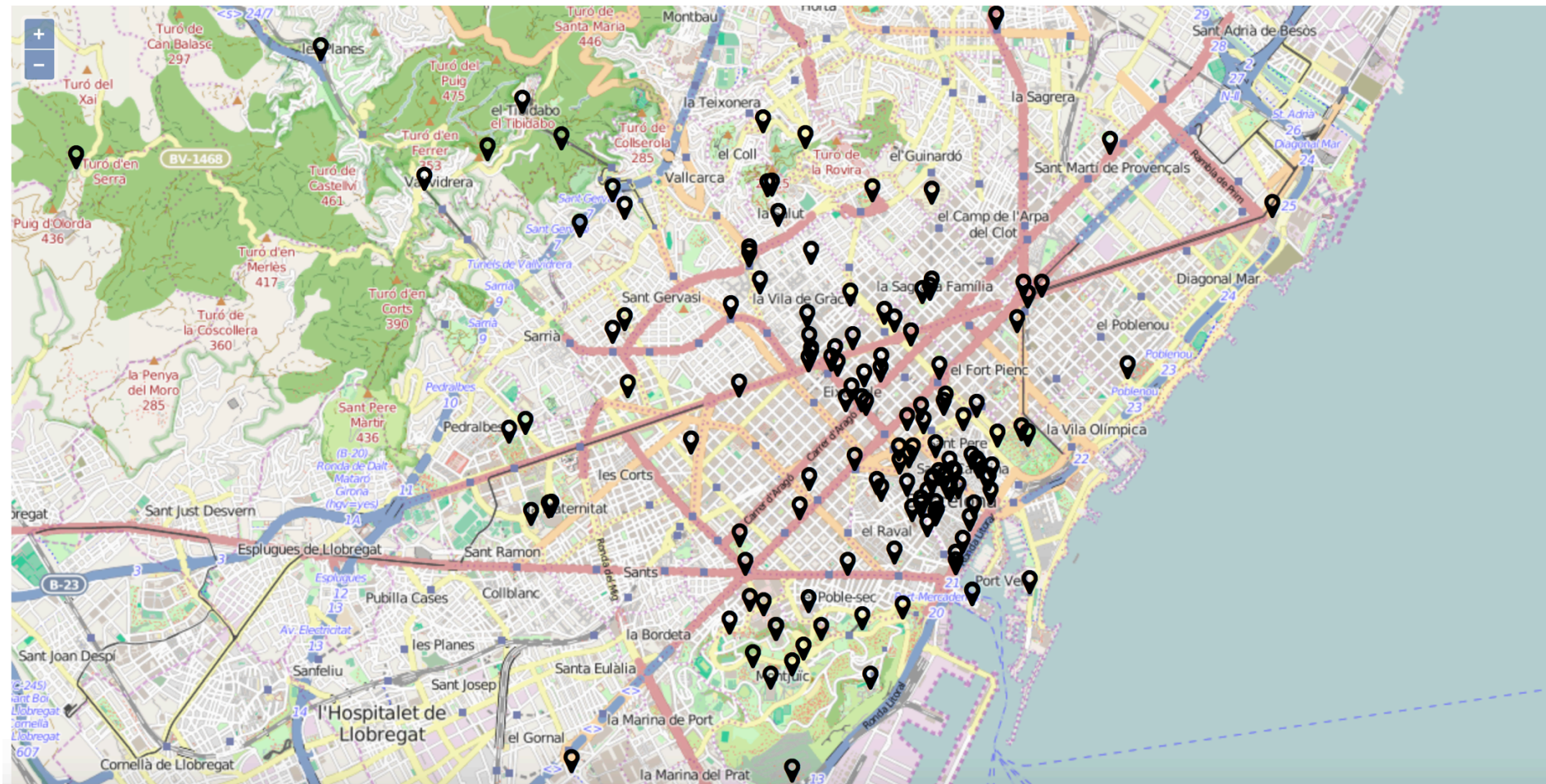
- “Native” Wikidata model data model (not RDF)
  - 16 million entities
  - 34 million statements
  - 80 million labels
  - 350 languages
- RDF exports available
  - Non-trivial mapping to RDF
  - >400 million triples
- Examples:
  - <http://wikidata.metaphacts.com/>
  - <http://wikidata.metaphacts.com/sparql>

# Geo/Spatial

metaphacts sparql

Logout

## Entities in Barcelona



# Geo/Spatial

## = Entities in Barcelona =

```
{{#widget: Map | markers = $
```

```
SELECT ?lat ?lng ?link ?description
```

```
WHERE {
```

```
  ?entity wd:P131c wd:Q1492;
```

```
    wd:P625c ?coordinates;
```

```
    wd:P373c ?description .
```

```
  ?coordinates <http://www.wikidata.org/ontology#longitude> ?lng ;
```

```
                <http://www.wikidata.org/ontology#latitude> ?lat .
```

```
  BIND(?entity as ?link) .
```

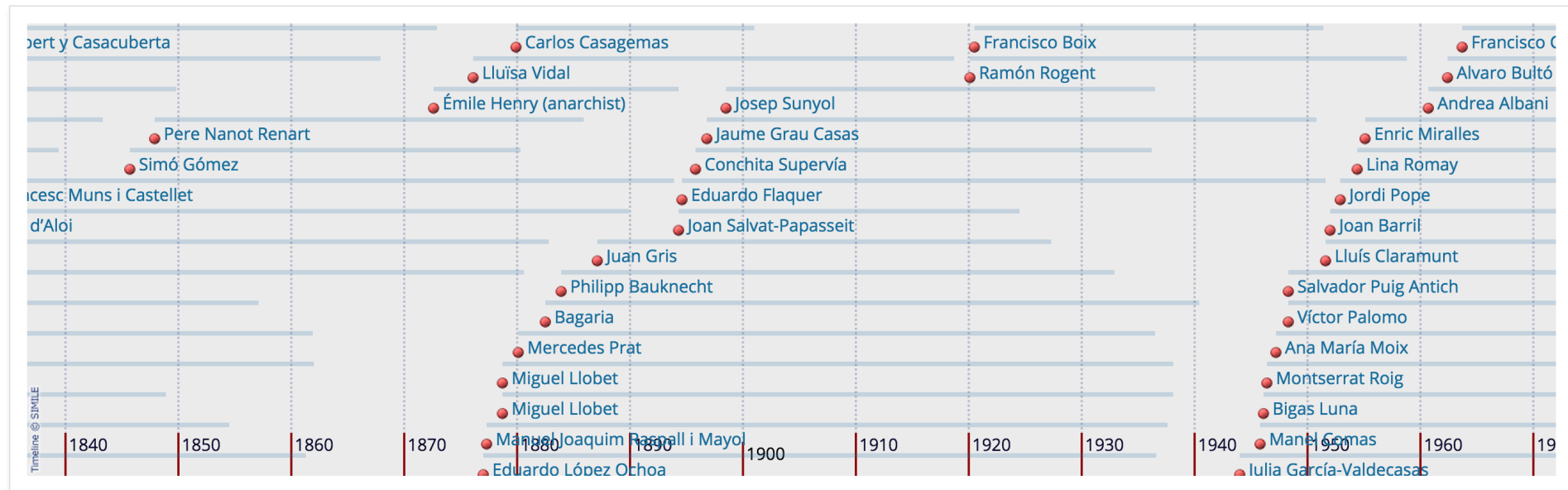
```
}
```

```
$}}
```

# Temporal



## People born in Barcelona on a Timeline





# Temporal

**= People born in Barcelona on a Timeline =**

```
{{#widget:Timeline
|query =

SELECT DISTINCT ?birth ?death ?label ?uri WHERE {
?uri rdfs:label ?label .
?uri wd:P569c ?birth .
?uri wd:P570c ?death .
?uri wd:P19c wd:Q1492 .
FILTER(lang(?label)='en')
}

|
| start='birth'
| end='death'
| label='label'
| link='uri'
| interval='DECADE'
}}
```

# Semantic Search

## Semantic Search

place of birth  ⊕ ⊖

occupation  ⊕ ⊖

Search

María Mercader



Loles León



Jinx Falkenburg



Maria Alba



Lola Dueñas



Astrid Berges-Frisbey



Carmen Amaya



Ariadna Gil



Sílvia Marsó



Juanjo Puigcorbé



Lluís Homar



Vanessa Incontrada



Santi Millán



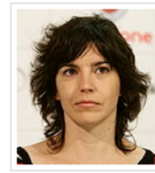
Montserrat Carulla



Leticia Dolera

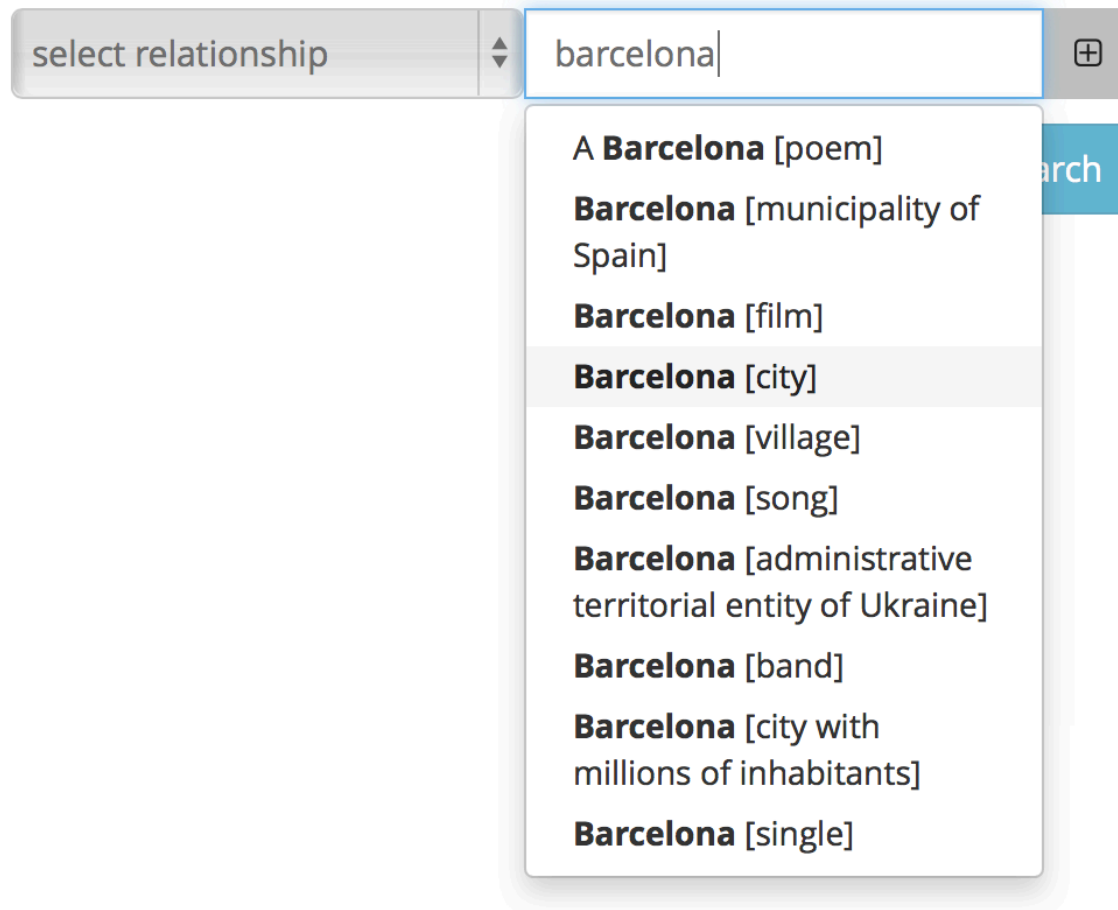


Laia Marull



# Semantic Search

## Semantic Search



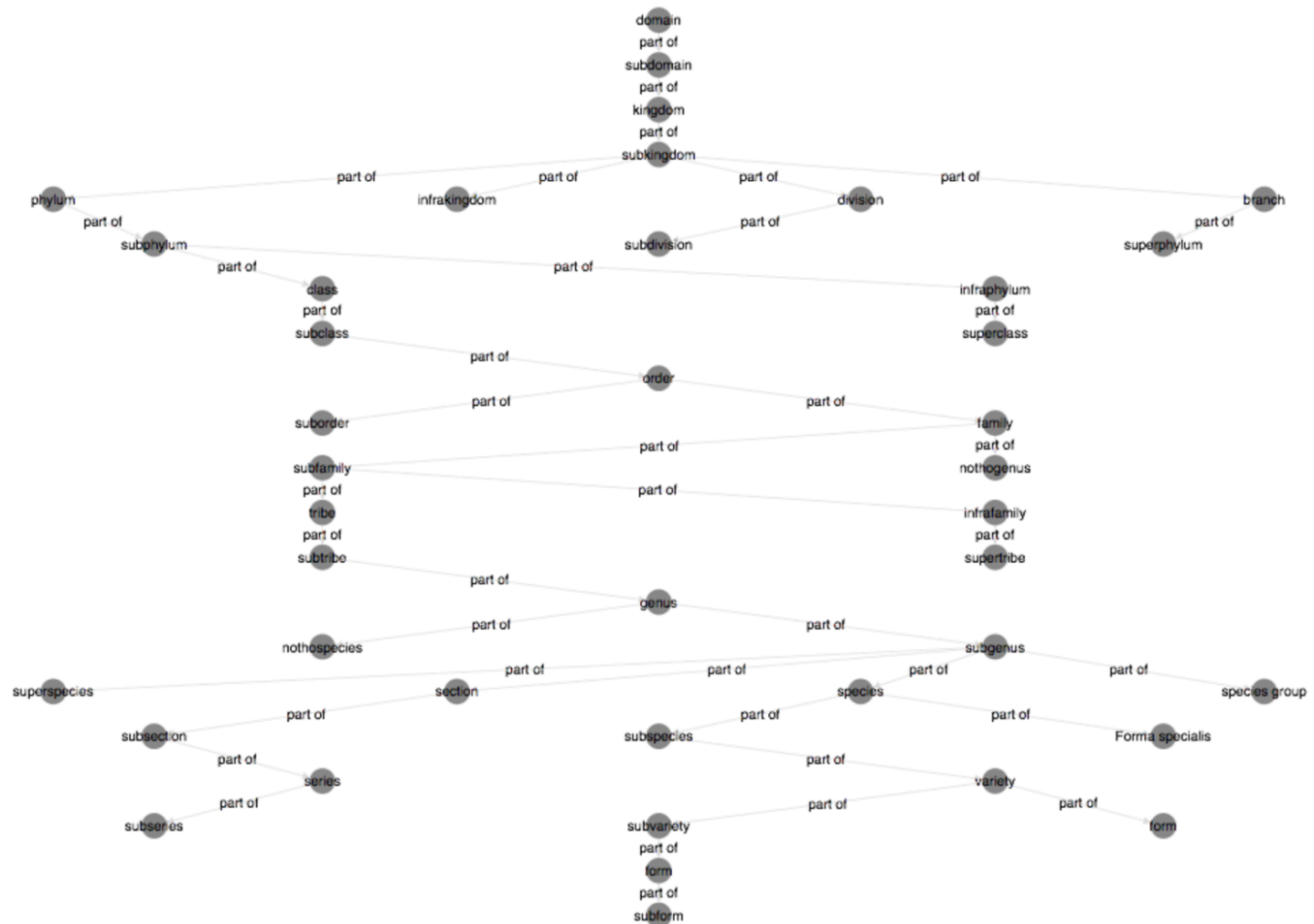
The image shows a search interface with a dropdown menu. The search bar contains the text "barcelona". The dropdown menu lists several results, each with a bolded title and a description in brackets. The results are:

- A **Barcelona** [poem]
- Barcelona** [municipality of Spain]
- Barcelona** [film]
- Barcelona** [city]
- Barcelona** [village]
- Barcelona** [song]
- Barcelona** [administrative territorial entity of Ukraine]
- Barcelona** [band]
- Barcelona** [city with millions of inhabitants]
- Barcelona** [single]

# Graph algorithms

metaphacts sparql

## Taxonomic Ranks (Biological classification)



# Graph algorithms

metaphacts sparql

## Taxonomic Ranks (Biological classification)

domain  
part of  
subdomain

### = Taxonomic Ranks (Biological classification) =

```
<row>
<col width="12" height="1000px">
{{#widget: com.metaphacts.widgets.GraphWidget
| query = 'PREFIX gas: <http://www.bigdata.com/rdf/gas#>
CONSTRUCT { ?pred wd:P361c ?out } WHERE {
SERVICE gas:service {
gas:program gas:gasClass "com.bigdata.rdf.graph.analytics.BFS" .
gas:program gas:in wd:Q146481 . # one or more times, specifies the initial frontier.
gas:program gas:out ?out . # will be bound to the visited vertices.
gas:program gas:out1 ?depth . # will be bound to the depth of the visited vertices.
gas:program gas:out2 ?pred . # | will be bound to the predecessor.
gas:program gas:maxVisited 2000 . # optional limit on the #of visited vertices.
gas:program gas:traversalDirection 'Reverse' .
gas:program gas:linkType wd:P361c
}
}
| layout = {{
name = 'breadthfirst' | roots = {{ 'http://www.wikidata.org/entity/Q146481' }}
}}
}}
</col>
</row>
```

part of  
form  
part of  
subform

# Statistics

## Statistics

### Most popular classes

class	count
<a href="#">Item</a>	16842535
<a href="#">human</a>	2693145
<a href="#">Wikimedia category page</a>	2422875
<a href="#">taxon</a>	1896799
<a href="#">GlobeCoordinatesValue</a>	1847718
<a href="#">Wikimedia disambiguation page</a>	805146
<a href="#">village-level division in China</a>	588520
<a href="#">owl:Class</a>	191385
<a href="#">Wikimedia list article</a>	186733
<a href="#">Wikimedia template</a>	185499

« 1 2 3 4 5 6 7 8 9 10 »

### Most popular properties

property	count
<a href="#">rdf:type</a>	31348748
<a href="#">instance of</a>	12446770
<a href="#">label</a>	8534489
<a href="#">country</a>	2658435
<a href="#">sex or gender</a>	2492158
<a href="#">description</a>	2305311
<a href="#">located in the administrative territorial entity</a>	2036112
<a href="#">taxon rank</a>	1904288
<a href="#">coordinate location</a>	1891469
<a href="#">taxon name</a>	1881588

« 1 2 3 4 5 6 7 8 9 10 »

# Qualified Statements and References

**Douglas Adams** (Q42)

[ edit ]

English writer and humorist

[ edit ]

Also known as:

Douglas Noël Adams

Douglas Noel Adams

DNA

Bop Ad

[ edit ]

⋮

date of birth

11 March 1952

[ edit ]

▶ 1 reference

⋮

Wikipedia pages linked to this item (64 entries)

Language	Code	Linked page	
العربية	arwiki	دوگلاس آدمز	[ edit ]
مصرى	arwiki	دوگلاس ادامز	[ edit ]
Boarisch	barwiki	Douglas Adams	[ edit ]
беларуская	be x oldwiki	Дуглас Адамз	[ edit ]

# Qualified Statements and References

spouse



Jane Belson

start date 25 November 1991

end date 11 May 2001

▼ 1 reference

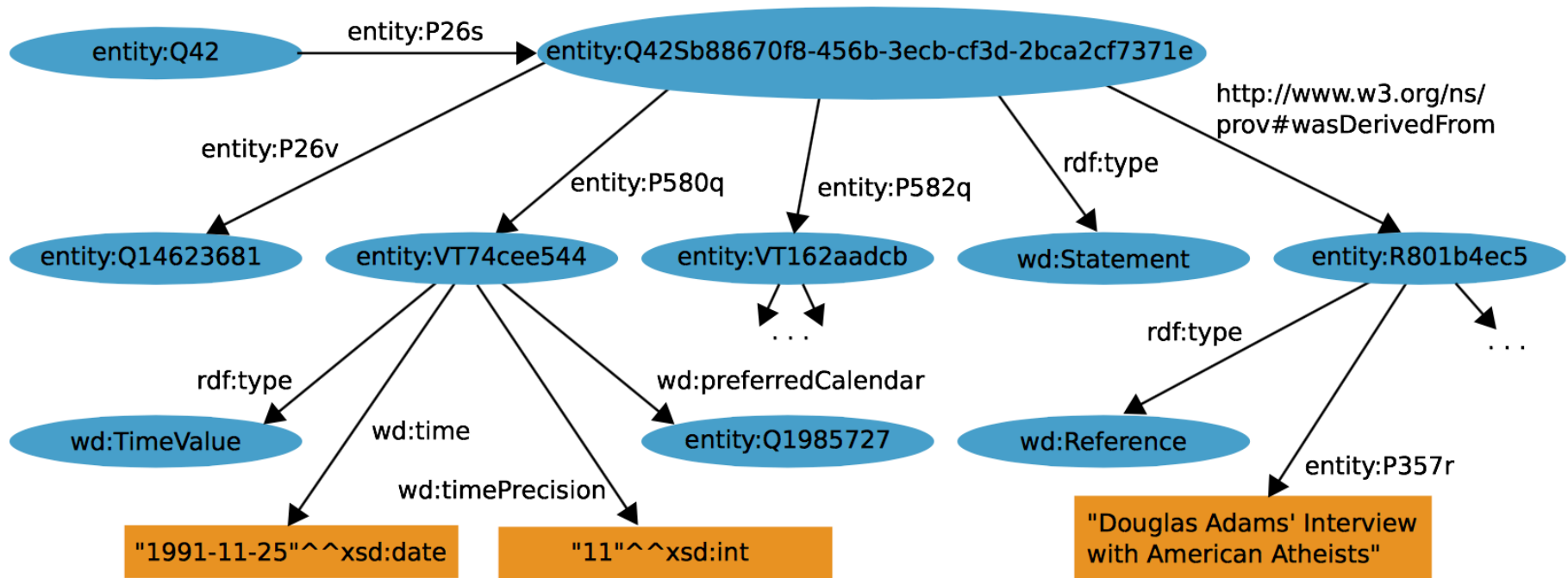
reference URL <http://www.nndb.com/people/731/000023662/>

original language English

title Douglas Adams



# Representing and querying qualified statements and references



Source: <http://korrekt.org/papers/Wikidata-RDF-export-2014.pdf>

```
SELECT ?property ?object ?reference WHERE
{
  entity:Q42 ?s_property ?sid .
  ?sid ?v_property ?object ;
    <http://www.w3.org/ns/prov#wasDerivedFrom> ?reference .
  FILTER ( REPLACE(str(?s_property), ".$", "") =
            REPLACE(str(?v_property), ".$", ""))
}
```

# RDR – Reification Done Right

- Simple extension to RDF and SPARQL
- Allows to use a statement as the subject of another statement
- Formalized by Hartig and Thompson in [Foundations of an Alternative Approach to Reification in RDF](#)
- @prefix : <http://www.wikidata.org/entity/> .  
@prefix prov: <http://www.w3.org/ns/prov#wasDerivedFrom> .  
  
<<:Q42 :p26 :Q14623681>> prov:wasDerivedFrom :R801b4ec5 .
- SELECT ?property ?object ?reference WHERE  
{  
 << :Q42 ?property ?object >> prov:wasDerivedFrom ?reference .  
}

# Conclusions

- Wikidata as useful knowledge graph with real life use cases
- Wikidata query services currently being developed based on RDF / SPARQL
- Expressive Wikidata data model that poses challenges for representation in RDF
- Reification Done Right as possible approach for dealing with qualified statements and references
- Standardization desired
- Interesting from benchmarking perspective

# Contact us!

## metaphacts GmbH

Kautzelweg 13

69190 Walldorf

Germany

p +49 6227 8308660

m +49 157 50152441

e [info@metaphacts.com](mailto:info@metaphacts.com)

 [@metaphacts](https://twitter.com/metaphacts)