

Mining the knowledge of the world with Wikidata

Marius Hoch
hoo

2016-05-05

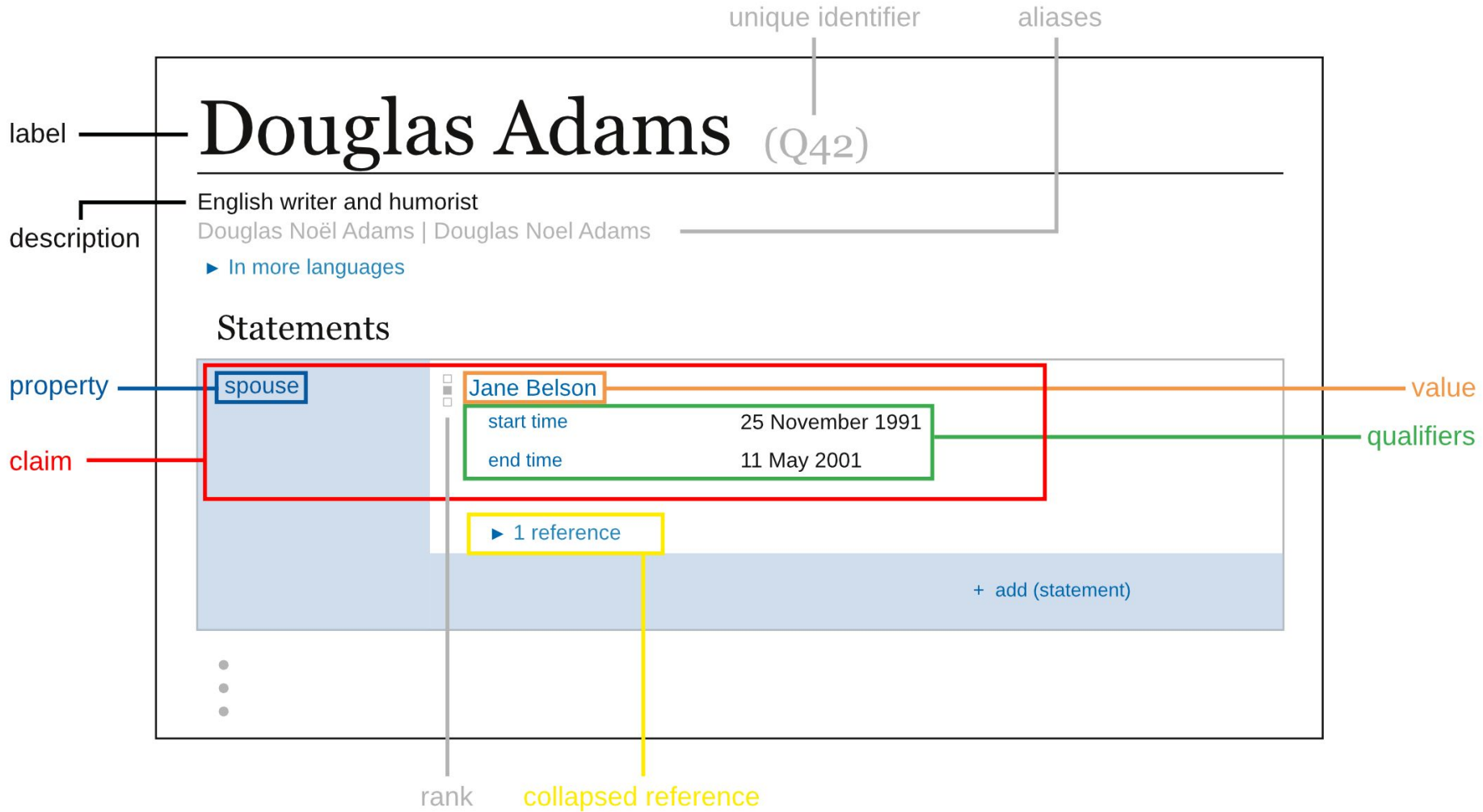
Slides adapted from:
Wikidata's SPARQL introduction for beginners
Lucie-Aimée Kaffee (CC BY-SA)



WIKIDATA

What is Wikidata?

- Free knowledge base by the Wikimedia movement
- Started ~ 3 years ago
- Structured data
- Linked data
- Persistent identifiers
- Multiple languages
- People, places, events, and more
- Powered by Wikibase (open source Mediawiki extension)
- Licenced under CC-0



What is RDF?

RDF triple language

Subject Predicate Object

Alan Turing (Q7251) instance of (P31) human (Q5)

Using URIs

<http://www.wikidata.org/entity/Q7251> <http://www.wikidata.org/prop/direct/P31> <http://www.wikidata.org/entity/Q5>

Turtle (Terse RDF Triple Language)

@prefix wd: <<http://www.wikidata.org/entity/>>

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

wd:Q7251 wdt:P31 wd:Q5 .

What is SPARQL?

W3C standardized RDF query language to form complex queries to SELECT and retrieve data from an endpoint.

Return all items, which are instance of human, limited to 200 items.

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

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

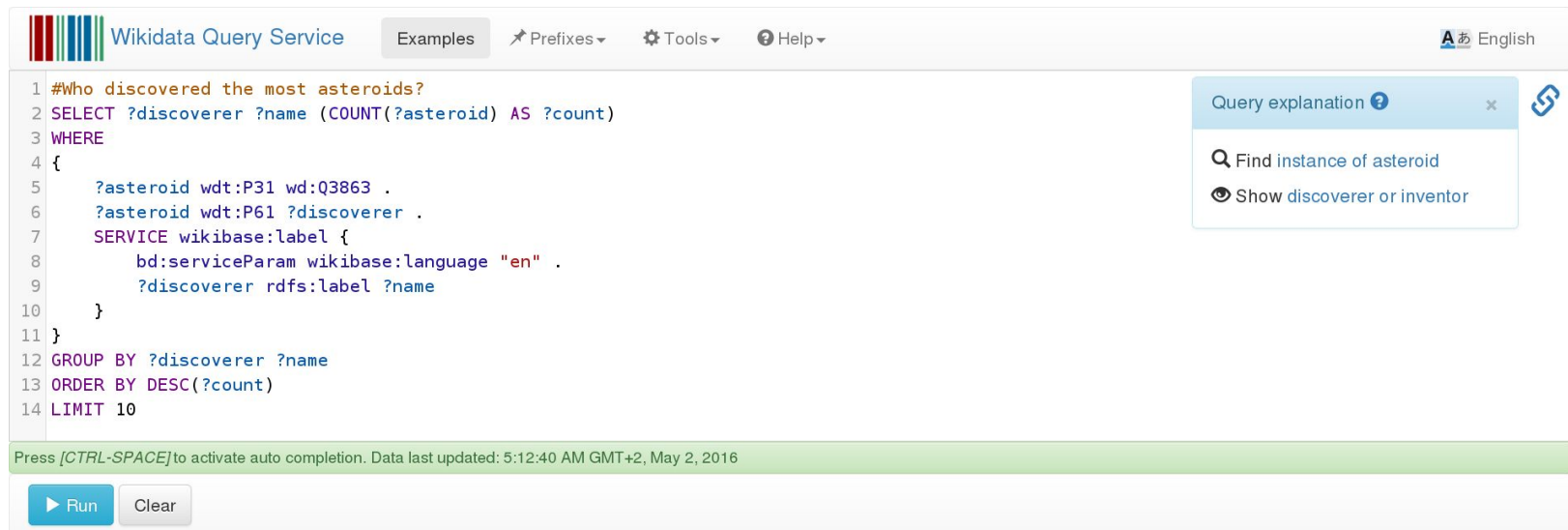
```
SELECT ?subject WHERE { ?subject wdt:P31 wd:Q5 . } LIMIT 200
```

Exploring Wikidata's linked data

Easy to use graphical frontend available at <https://query.wikidata.org/>

Supports some generic visualizations

Many Examples



The screenshot shows the Wikidata Query Service interface. At the top, there is a navigation bar with the Wikidata logo, the text "Wikidata Query Service", and several menu items: "Examples", "Prefixes", "Tools", and "Help". On the right side of the navigation bar, there is a language selector set to "English".

The main area contains a SPARQL query editor with the following code:

```
1 #Who discovered the most asteroids?
2 SELECT ?discoverer ?name (COUNT(?asteroid) AS ?count)
3 WHERE
4 {
5   ?asteroid wdt:P31 wd:Q3863 .
6   ?asteroid wdt:P61 ?discoverer .
7   SERVICE wikibase:label {
8     bd:serviceParam wikibase:language "en" .
9     ?discoverer rdfs:label ?name
10  }
11 }
12 GROUP BY ?discoverer ?name
13 ORDER BY DESC(?count)
14 LIMIT 10
```

Below the query editor, there is a green bar with the text: "Press [CTRL-SPACE] to activate auto completion. Data last updated: 5:12:40 AM GMT+2, May 2, 2016".

At the bottom left, there are two buttons: "Run" and "Clear".

On the right side, there is a "Query explanation" panel with a close button (x) and a link icon. It contains two options:

- Find instance of asteroid
- Show discoverer or inventor

Results can also be accessed in various machine readable formats

Exploring Wikidata's linked data

Show data for Alan Turing (Q7251) limited to 100

```
SELECT ?predicate ?object WHERE {  
  wd:Q7251 ?predicate ?object  
} LIMIT 100
```

[On Wikidata Query Service](#)

Extract of the result

predicate	object
rdfs:label	Alan Turing
rdfs:label	آلان تورنج
skos:altLabel	Алан Цюрынг
wdt:P101	wd:Q21198

Turing's field of work (P101) is computer science (Q21198)

Exploring Wikidata's linked data

Show all fields of work (P101) of Alan Turing

```
SELECT ?object ?objectLabel WHERE {  
  wd:Q7251 wdt:P101 ?object .  
  SERVICE wikibase:label {  
    bd:serviceParam wikibase:language "en" .  
  }  
} LIMIT 100
```

[On Wikidata Query Service](#)

object	objectLabel
wdt:Q21198	computer science
wdt:Q897511	cryptanalysis

Truthy and full statements

represents statements in two formats

truthy

contain only the statements of best rank

best rank are all statements of preferred rank, if those exist or all statements of normal rank

qualifiers are ignored

Prefix *wdt:* (<http://www.wikidata.org/prop/direct/>) - e.g. *wdt:P101*

full statements

represent all data about a statement

represented in a separate node with the unique id of the statement - e.g. *q92741-7ED21209-4E23-4EB1-886A-772CC6A5350F*

Prefix *p:* (<http://www.wikidata.org/prop/>) - e.g. *p:P101*

Example Query

Academy award winning directors

award received		Academy Award for Best Director
	point in time	1998
	for work	Titanic
	subject of	70th Academy Awards

(excerpt from the James Cameron item)

```

SELECT DISTINCT ?item ?itemLabel ?awardLabel ?time
WHERE {
  ?item wdt:P106/wdt:P279* wd:Q3455803 .
  ?item p:P166 ?awardStat .
  ?awardStat ps:P166 wd:Q103360 .
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language "en" .
  }
  ?awardStat pq:P805 ?award .
  ?award rdfs:label ?awardLabel filter (lang(?awardLabel) = "en") .
  ?award wdt:P585 ?time .
}
ORDER BY DESC(?time)

```

award received

Academy Award for Best Director

point in time 1998

for work Titanic

subject of 70th Academy Awards

Items with the Occupation (P106) of Director (Q3455803) or a subclass (P279)

with an awarded (P166) statement

that has the value Academy Award for Best Director (Q103360)

include the labels

Get the award (which is "subject of" XXth Academy Awards)

and its label

the "point of time" of the Academy Award

```

SELECT DISTINCT ?item ?itemLabel ?awardLabel ?time
WHERE {
  ?item wdt:P106/wdt:P279* wd:Q3455803 .
  ?item p:P166 ?awardStat .
  ?awardStat ps:P166 wd:Q103360 .
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language "en" .
  }
  ?awardStat pq:P805 ?award .
  ?award rdfs:label ?awardLabel filter (lang(?awardLabel) = "en") .
  ?award wdt:P585 ?time .
}
ORDER BY DESC(?time)

```

award received	Academy Award for Best Director
point in time	1998
for work	Titanic
subject of	70th Academy Awards

Items with the Occupation (P106) of Director (Q3455803) or a subclass (P279)

with an awarded (P166) statement

that has the value Academy Award for Best Director (Q103360)

include the labels

Get the award (which is "subject of" XXth Academy Awards)

and its label

the "point of time" of the Academy Award

```

SELECT DISTINCT ?item ?itemLabel ?awardLabel ?time
WHERE {
  ?item wdt:P106/wdt:P279* wd:Q3455803 .
  ?item p:P166 ?awardStat .
  ?awardStat ps:P166 wd:Q103360 .
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language "en" .
  }
  ?awardStat pq:P805 ?award .
  ?award rdfs:label ?awardLabel filter (lang(?awardLabel) = "en") .
  ?award wdt:P585 ?time .
}
ORDER BY DESC(?time)

```

award received	Academy Award for Best Director
point in time	1998
for work	Titanic
subject of	70th Academy Awards

Items with the Occupation (P106) of Director (Q3455803) or a subclass (P279) with an awarded (P166) statement that has the value Academy Award for Best Director (Q103360) include the labels

Get the award (which is "subject of" XXth Academy Awards) and its label

the "point of time" of the Academy Award

```

SELECT DISTINCT ?item ?itemLabel ?awardLabel ?time
WHERE {
  ?item wdt:P106/wdt:P279* wd:Q3455803 .
  ?item p:P166 ?awardStat .
  ?awardStat ps:P166 wd:Q103360 .
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language "en" .
  }
  ?awardStat pq:P805 ?award .
  ?award rdfs:label ?awardLabel filter (lang(?awardLabel) = "en") .
  ?award wdt:P585 ?time .
}
ORDER BY DESC(?time)

```

award received	Academy Award for Best Director
point in time	1998
for work	Titanic
subject of	70th Academy Awards

Items with the Occupation (P106) of Director (Q3455803) or a subclass (P279) with an awarded (P166) statement that has the value Academy Award for Best Director (Q103360)

include the labels

Get the award (which is "subject of" XXth Academy Awards)

and its label

the "point of time" of the Academy Award

```

SELECT DISTINCT ?item ?itemLabel ?awardLabel ?time
WHERE {
  ?item wdt:P106/wdt:P279* wd:Q3455803 .
  ?item p:P166 ?awardStat .
  ?awardStat ps:P166 wd:Q103360 .
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language "en" .
  }
  ?awardStat pq:P805 ?award .
  ?award rdfs:label ?awardLabel filter (lang(?awardLabel) = "en") .
  ?award wdt:P585 ?time .
}
ORDER BY DESC(?time)

```

award received	Academy Award for Best Director
point in time	1998
for work	Titanic
subject of	70th Academy Awards

Items with the Occupation (P106) of Director (Q3455803) or a subclass (P279)
 with an awarded (P166) statement
 that has the value Academy Award for Best Director (Q103360)
 include the labels
 Get the award (which is "subject of" XXth Academy Awards)
 and its label
 the "point of time" of the Academy Award


```

SELECT DISTINCT ?item ?itemLabel ?awardLabel ?time
WHERE {
  ?item wdt:P106/wdt:P279* wd:Q3455803 .
  ?item p:P166 ?awardStat .
  ?awardStat ps:P166 wd:Q103360 .
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language "en" .
  }
  ?awardStat pq:P805 ?award .
  ?award rdfs:label ?awardLabel filter (lang(?awardLabel) = "en") .
  ?award wdt:P585 ?time .
}
ORDER BY DESC(?time)

```

award received	Academy Award for Best Director
point in time	1998
for work	Titanic
subject of	70th Academy Awards

Items with the Occupation (P106) of Director (Q3455803) or a subclass (P279)
 with an awarded (P166) statement
 that has the value Academy Award for Best Director (Q103360)
 include the labels
 Get the award (which is "subject of" XXth Academy Awards)
 and its label
 the "point of time" of the Academy Award

```

SELECT DISTINCT ?item ?itemLabel ?awardLabel ?time
WHERE {
  ?item wdt:P106/wdt:P279* wd:Q3455803 .
  ?item p:P166 ?awardStat .
  ?awardStat ps:P166 wd:Q103360 .
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language "en" .
  }
  ?awardStat pq:P805 ?award .
  ?award rdfs:label ?awardLabel filter (lang(?awardLabel) = "en") .
  ?award wdt:P585 ?time .
}
ORDER BY DESC(?time)

```

award received	Academy Award for Best Director
	point in time 1998
	for work Titanic
	subject of 70th Academy Awards

Items with the Occupation (P106) of Director (Q3455803) or a subclass (P279)
 with an awarded (P166) statement
 that has the value Academy Award for Best Director (Q103360)
 include the labels
 Get the award (which is "subject of" XXth Academy Awards)
 and its label
 the "point of time" of the Academy Award

Result

item	itemLabel	awardLabel	time
Q wd:Q55215	Alejandro González Iñárritu	88th Academy Awards	2016-02-28T00:00:00Z
Q wd:Q55215	Alejandro González Iñárritu	87th Academy Awards	2015-02-22T00:00:00Z
Q wd:Q28028	Alfonso Cuarón	86th Academy Awards	2014-03-02T00:00:00Z
Q wd:Q160726	Ang Lee	85th Academy Awards	2013-02-24T00:00:00Z
Q wd:Q28556	Michel Hazanavicius	84th Academy Awards	2012-02-26T00:00:00Z
Q wd:Q295912	Tom Hooper	83rd Academy Awards	2011-02-27T00:00:00Z
Q wd:Q34816	Kathryn Bigelow	82nd Academy Awards	2010-03-07T00:00:00Z
Q wd:Q134867	Danny Boyle	81st Academy Awards	2009-02-22T00:00:00Z
Q wd:Q13595311	Joel Coen	80th Academy Awards	2008-02-24T00:00:00Z
Q wd:Q13595531	Ethan Coen	80th Academy Awards	2008-02-24T00:00:00Z
Q wd:Q41148	Martin Scorsese	79th Academy Awards	2007-02-25T00:00:00Z
Q wd:Q160726	Ang Lee	78th Academy Awards	2006-03-05T00:00:00Z
Q wd:Q43203	Clint Eastwood	77th Academy Awards	2005-02-27T00:00:00Z
Q wd:Q4465	Peter Jackson	76th Academy Awards	2004-02-29T00:00:00Z
Q wd:Q51552	Roman Polanski	75th Academy Awards	2003-03-23T00:00:00Z
Q wd:Q103917	Steven Soderbergh	73rd Academy Awards	2001-03-25T00:00:00Z
Q wd:Q223233	Sam Mendes	72nd Academy Awards	2000-03-26T00:00:00Z
Q wd:Q8877	Steven Spielberg	71st Academy Awards	1999-03-21T00:00:00Z
Q wd:Q42574	James Cameron	70th Academy Awards	1998-03-23T00:00:00Z
Q wd:Q188726	Anthony Minghella	69th Academy Awards	1997-03-24T00:00:00Z
Q wd:Q42229	Mel Gibson	68th Academy Awards	1996-03-25T00:00:00Z

Visualizations



```
1 #Map of U1 subway stations in Berlin
2 #defaultView:Map
3 SELECT ?stationLabel ?coord
4 WHERE {
5   ?station wdt:P81 wd:Q99691 .
6   ?station wdt:P625 ?coord
7   SERVICE wikibase:Label { bd:serviceParam wikibase:language "de" }
8 }
```

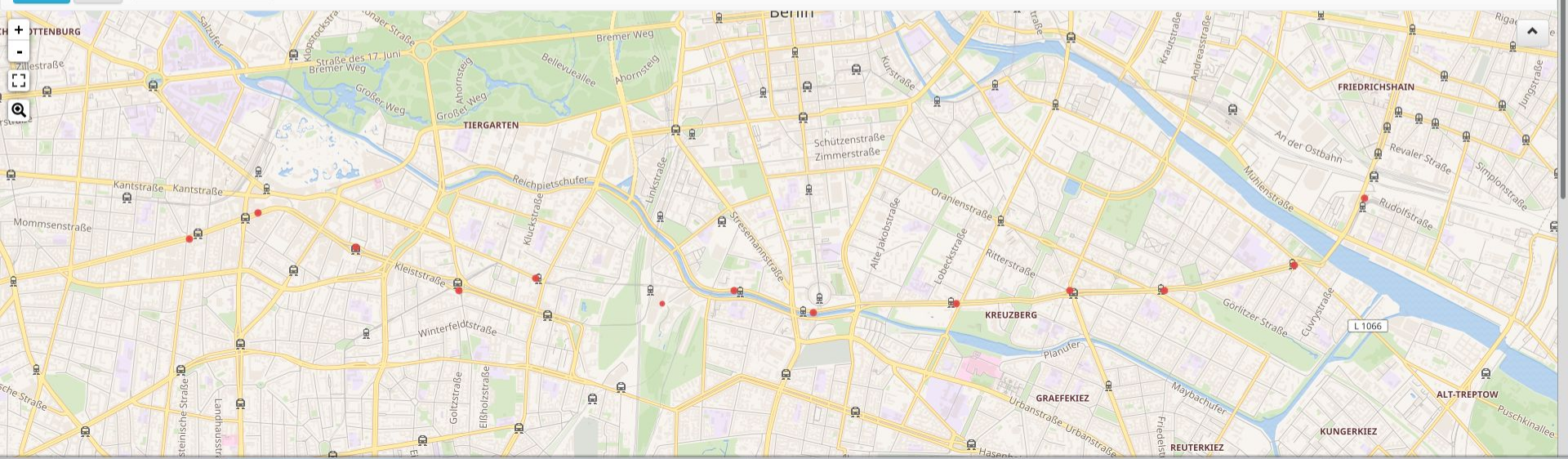
Query explanation ✕ 🔗

- Find
- Show

Press **[CTRL-SPACE]** to activate auto completion. Data last updated: 10:40:57 AM GMT+2, May 5, 2016

▶ Run Clear 13 Results in 3182 ms

📺 Display 📄 Download 🔗 Link



Contact

Wikidata

#wikidata on Freenode

@wikidata (Twitter)

wikidata@lists.wikimedia.org

www.wikidata.org/wiki/Wikidata:Project_chat

Marius

hoo on #wikidata

@mariushoch (Twitter)

Marius.Hoch@wikimedia.de

[\[\[User:Hoo man\]\]](#)

Ressources

[Wikidata Query Service Beta](https://query.wikidata.org/)

<https://query.wikidata.org/>

[Mediawiki: RDF Dump Format](https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format)

https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format

[Prefixes used for Wikidata](https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format#Prefixes_used)

https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format#Prefixes_used

[Mediawiki: Query Examples](https://www.mediawiki.org/wiki/Wikibase/Indexing/SPARQL_Query_Examples)

https://www.mediawiki.org/wiki/Wikibase/Indexing/SPARQL_Query_Examples

[github: wikidata query rdf docs](https://github.com/wikimedia/wikidata-query-rdf/tree/master/docs)

<https://github.com/wikimedia/wikidata-query-rdf/tree/master/docs>

[Wikidata turtle dumps](https://dumps.wikimedia.org/wikidatawiki/entities/)

<https://dumps.wikimedia.org/wikidatawiki/entities/>

[W3C RDF Primer](http://www.w3.org/2007/02/turtle/primer/)

<http://www.w3.org/2007/02/turtle/primer/>

[Slides](https://commons.wikimedia.org/wiki/File:Mining_the_knowledge_of_the_world_with_Wikidata.pdf)

https://commons.wikimedia.org/wiki/File:Mining_the_knowledge_of_the_world_with_Wikidata.pdf