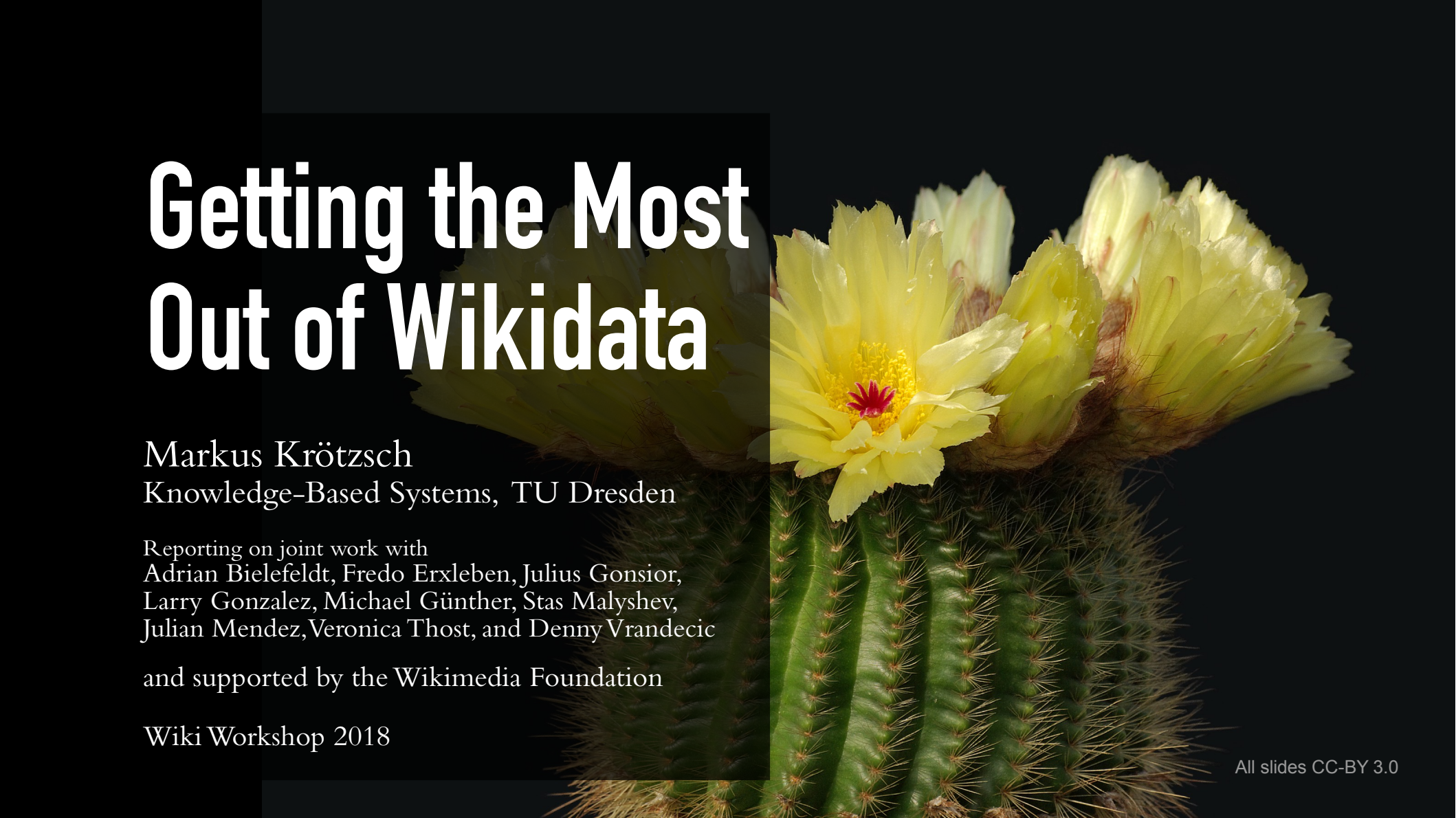
Getting the Most Out of Wikidata

Markus Krötzsch
Knowledge-Based Systems, TU Dresden

Reporting on joint work with
Adrian Bielefeldt, Fredo Erxleben, Julius Gonsior, Larry Gonzalez, Michael Günther, Stas Malyshev, Julian Mendez, Veronica Thost, and Denny Vrandecic

and supported by the Wikimedia Foundation

Wiki Workshop 2018



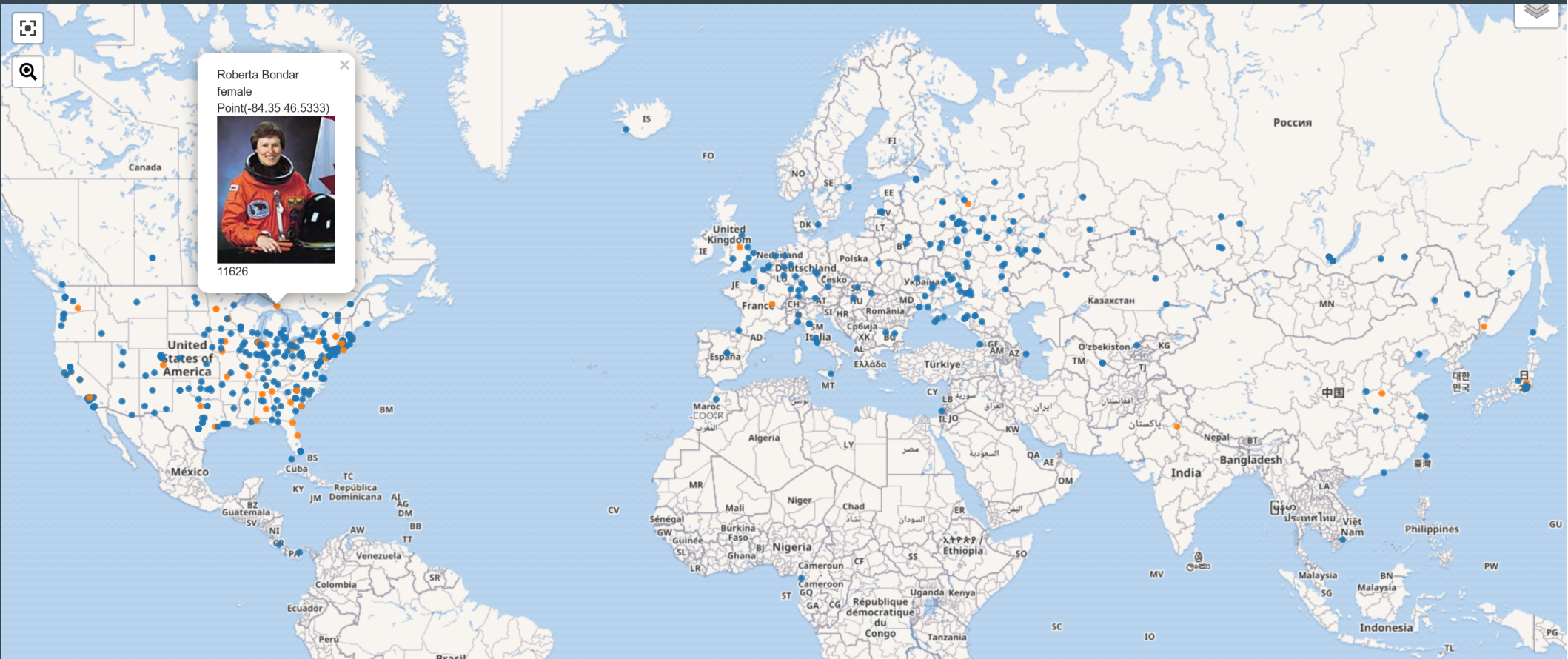
“What are the world’s largest cities with a female mayor?”

“What are the world’s largest cities with a female mayor?”

cityLabel	mayorLabel	population
Tokyo	Yuriko Koike	13742906
Hong Kong	Carrie Lam	7336585
Baghdad	Zekra Alwach	6960000
Surabaya	Tri Rismaharini	4975000
Yokohama	Fumiko Hayashi	3733234
Madrid	Manuela Carmena	3182981
Rome	Virginia Raggi	2873494
Kaohsiung City	Chen Chu	2777384
Antananarivo	Lalao Ravalomanana	2610000
Paris	Anne Hidalgo	2206488

“Where are people born who travel to space?”

(Colour-coded by gender)



“Which days of the week do disasters occur on?”

Date	Mon	Tue	Wed	Thu	Fri	Sat	Sun
1	25	33	22	18	26	28	23
2	24	26	23	23	22	32	12
3	24	27	21	31	23	28	38
4	24	25	33	25	26	26	24
5	37	23	32	18	19	17	29
6	25	28	32	20	24	33	22
7	18	22	25	16	22	18	17
8	32	28	19	25	22	23	19
9	20	25	29	29	27	21	23
10	20	20	19	14	25	25	29
11	30	34	28	23	22	20	20
12	41	33	27	30	20	20	23
13	35	26	29	21	25	24	25
14	24	23	27	23	22	28	17

“Which 19th century paintings show the moon?”


commons:Xi Bannister, Moon P.jpg
Moonlight Marine
Moon


commons:Maufraleverdeluneaucrepusculei1899.jpg
Moonrise at Twilight
Moon


commons:Johan Christian Clausen Dahl - Kysten ved Lauvvig i Norge i måneskin - Thorvalds...
The coast at Lauvvig, Norway
Moon


commons:Johan Christian Claussen Dahl - Bugten ved Napoli i måneskin (1821).jpg
The Bay of Naples by moonlight
Moon


commons:'Dresden in Moonlight' by Johan Christian Dahl, Bergen Kunstmuseum.JPG
Dresden by Moonlight
Moon

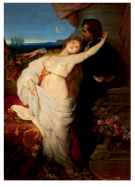

commons:Johan Christian Dahl - Den Esrum sjøen i måneskinn.jpg
Esrom Lake in Moonlight
Moon


commons:Dahl, Der Kopenhagener Hafen im Mondschein, 1831.jpg
Q24704472
Moon


commons:Dupré Landscape by moonlight.jpg
Q24699126
Moon


commons:Szermentowski Morning star.jpg
Morning star
Moon


commons:Lampi Carthusian monastery.jpg
Q23730383
Moon


commons:Max Tannhäuser.jpg
Q23009000
Moon


commons:Louis Rémy Mignot Marsh in Ecuador.jpg
Moonlight over a Marsh in Ecuador
Moon


commons:Johan Christian Dahl - View of Dresden by Moonlight - Google Art Project (NwHK-NsdlnFIMQ).jpg
View of Dresden by Moonlight
Moon


commons:Johan Christian Dahl - Dresden by Moonlight - Google Art Project.jpg
Dresden by Moonlight
Moon


commons:Elin Danielson-Gambogi Kuutamo n 1890.jpg
Moonlight
Moon


commons:Ladislav Medňanszk...
Night Travellers at a Cross
Moon


commons:Night Travellers at a ...
Night Travellers at a Cross
Moon


commons:Martin Johnson Heade - Two Owls at Sunset.jpg
Two Owls at Sunset
Moon


commons:Martin Johnson Heade -Sailing by Moonlight.jpg
Sailing by Moonlight
Moon


commons:'Point Judith, Rhode Island' by Martin Johnson Heade, 1867-68.JPG
Point Judith, Rhode Island
Moon


commons:Midwinter Moonlight by Regis-Francois Gignoux, before 1869, oil on board ...
Mid-Winter Moonlight
Moon


commons:Sanford Robinson Gifford - Crépuscule sur le mont Hunter.jpg
Hunter Mountain, Twilight
Moon


commons:Samuel Colman - The Rock of Salvation - Google Art Pr...
The Rock of Salvation
Moon


commons:John William Casilear-Mo...
Moonrise on the Coast
Moon


commons:Albert Pinkham Ryder - The Lovers' Boat (c.188...
The Lovers' Boat
Moon


commons:Johan Christian Clausen Dahl - Ved den napolitanske golf. Måneskin - Staten,...
The Gulf of Naples. Moonlight
Moon


commons:Johan Christian Clausen Dahl - Ved den napolitanske golf. Måneskin - Staten...
The Gulf of Naples. Moonlight
Moon


commons:Albert Bierstadt - Cloud ...
Cloud Study, Moonlight
Moon


commons:John Martin - The Eve of the Deluge - WGA14146.jpg
The Eve of the Deluge
Moon


commons:Thomas Chambers - Storm-Tossed Frigate.jpg
Storm-Tossed Frigate
Moon


commons:Richards William Trost Moonlight On Mount Lafayette New hampshire.jpg
Moonlight on Mount Lafayette, New Hampshire
Moon


commons:Moonrise by George Inness 1887.jpeg
Moonrise
Moon


commons:'Moonlight' by George Inness, 1893.JPG
Moonlight
Moon

“Which UK bands’ members’ average age is 70 or greater?”

bandLabel	avgAge
The Searchers	80
Rory Storm	79
The Animals	77
Quintessence	77
The Corporation	76.2
The Hollies	76
Cream	75.66666666666666666667
Soft Machine	75.5
The Moody Blues	75.33333333333333333333

“The free knowledge base that anyone can edit”

Wikimania05/Paper-MK2

< Wikimania05

This page is part of the **Proceedings of Wikimania 2005**, Frankfurt, Germany.

0 MISSING	1 Submitted	2 Editing	3 Author review	4 Final edit	5 DONE

Wikipedia and the Semantic Web - The Missing Links [edit]

- **Author(s):** Markus Krötzsch & Denny Vrandečić & Max Völkel
- **License:** CC-NC-SA 2.0 (for further license models, please contact the authors)
- **Slides**: collected but not uploaded yet
- **Video**: 16:44 (talk given by Denny Vrandečić)
- **Note:** Presentation, paper also at Wikipedia and the Semantic Web (PDF, 164K)

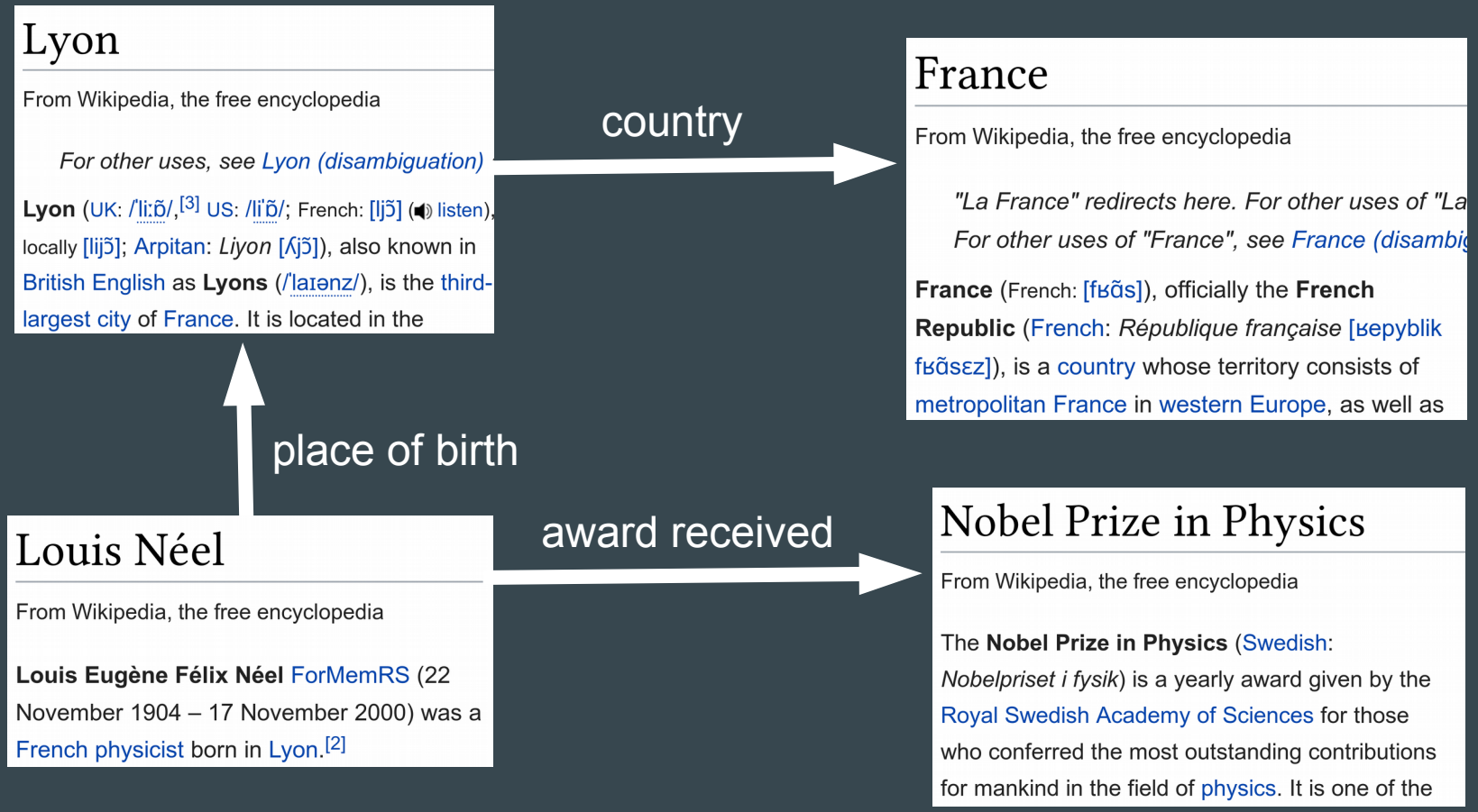
About the author: The authors are research associates at the *Institute of Applied Informatics and Formal Description Methods* (AIFB), University of Karlruhe, Germany, where they are members of the AIFB Research Group Knowledge Management, an interdisciplinary team of computer scientists, mathematicians, and industrial engineers that is one of the world-wide leading institutions in the Semantic Web research community. Other relevant research topics include Semantic Web, ontologies, data and text mining, logic-based knowledge representation, peer-to-peer, and Web services.

Being enthusiastic users and contributors of various Wikis, the authors also have special interest in making emerging semantic technologies available within Wikis, where computer-assisted organization and processing of knowledge plays an important role.

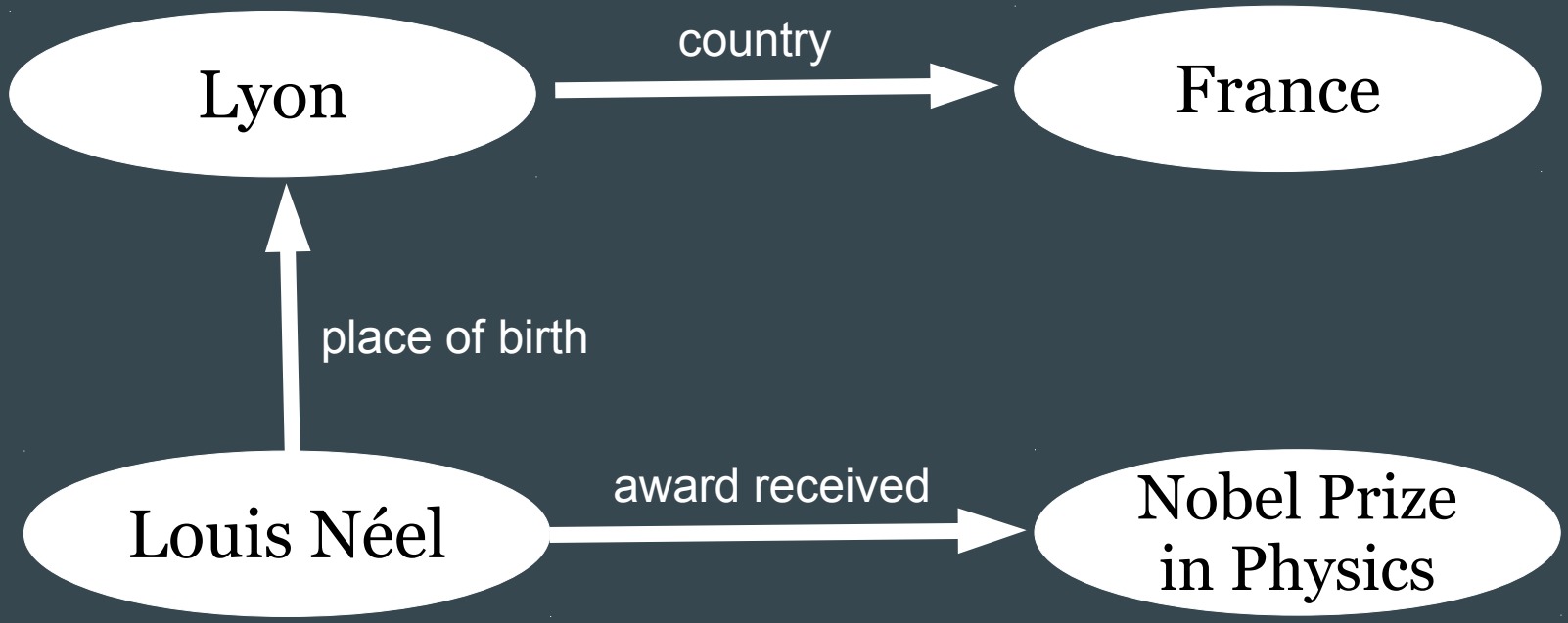
Contents [hide]

1 **Wikipedia and the Semantic Web - The Missing Links**
- 1.1 Introduction
- 1.2 A jump start introduction to semantic technologies
- 1.3 Design
- 1.4 Usability aspects
- 1.5 Implementation, performance and scalability
- 1.6 Additional features
- 1.7 Implementation plan
- 1.8 Applications
- 1.9 Related approaches
- 1.10 Summary and conclusion
- 1.11 Acknowledgements
- 1.12 Bibliography

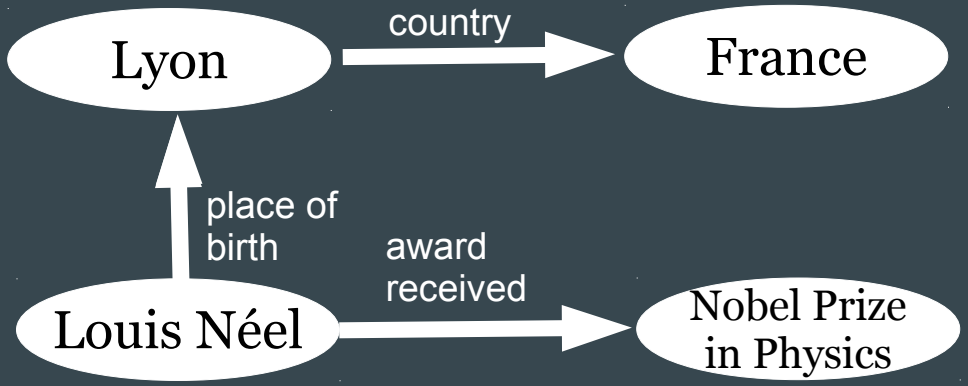
A Simple Idea (2005): “Let’s annotate Wikipedia links!”



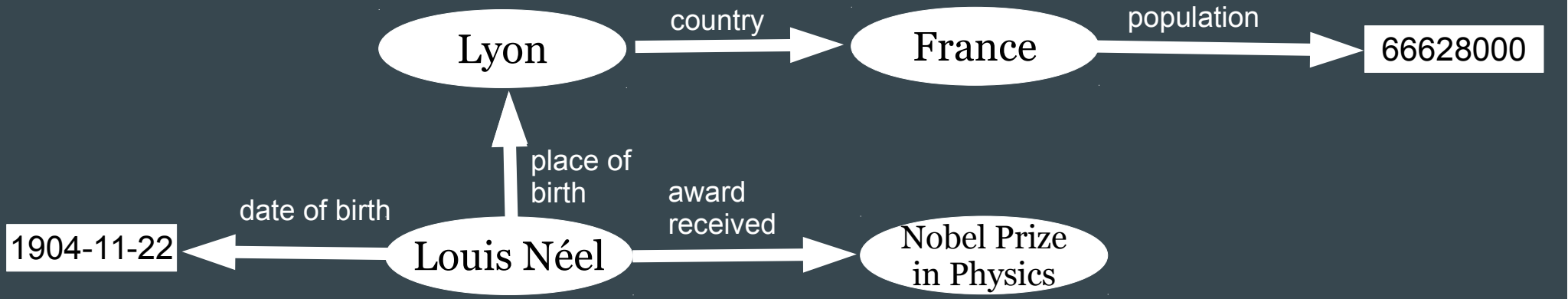
Semantic MediaWiki (2005): From Links to Graphs



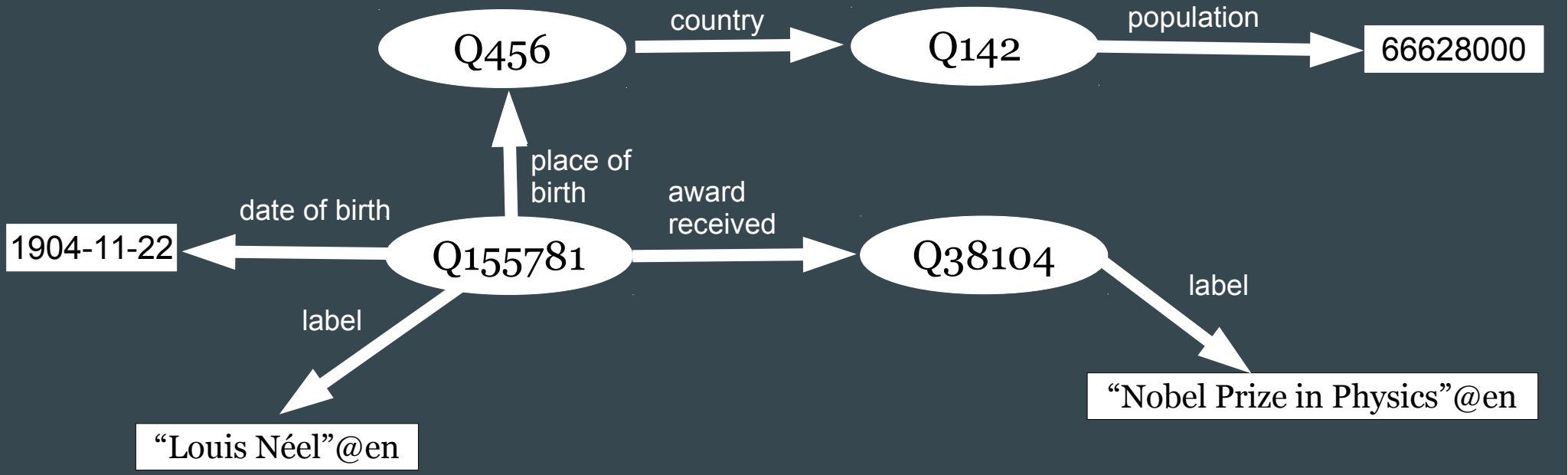
Links are not Enough: Adding Datatypes



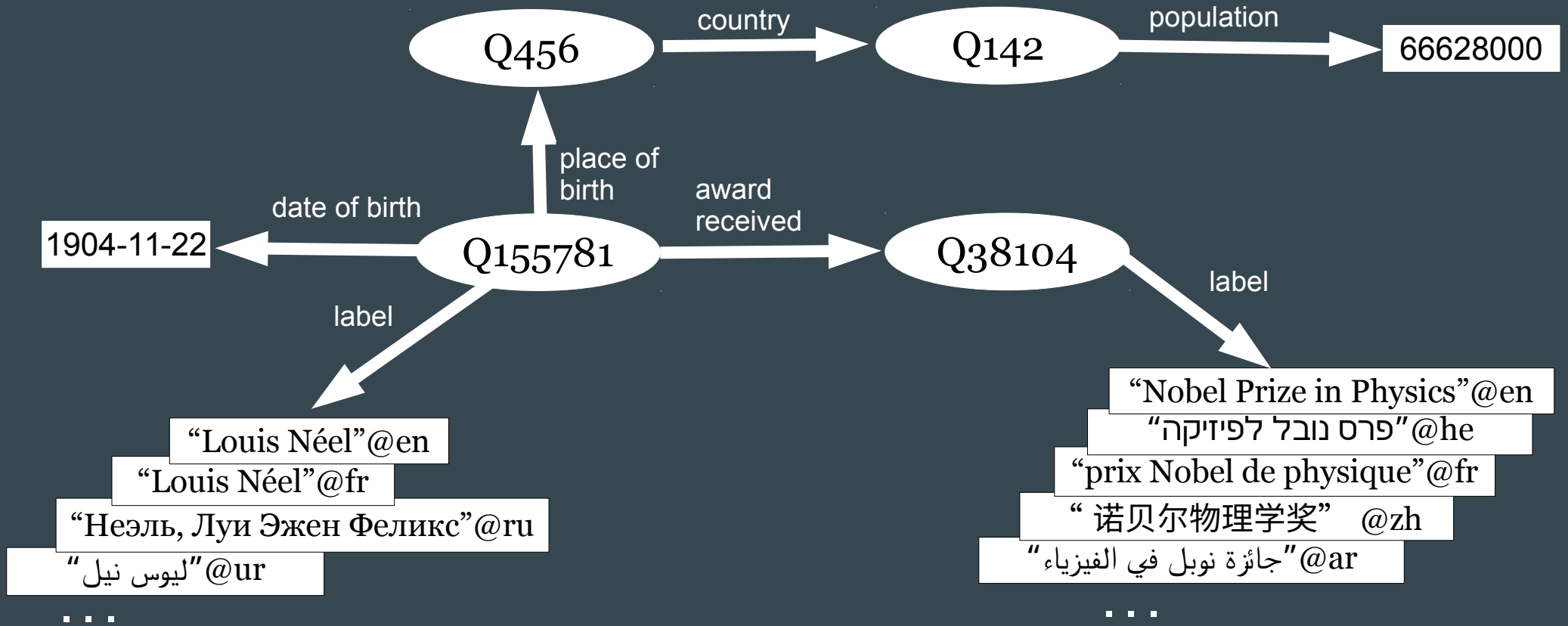
Links are not Enough: Adding Datatypes



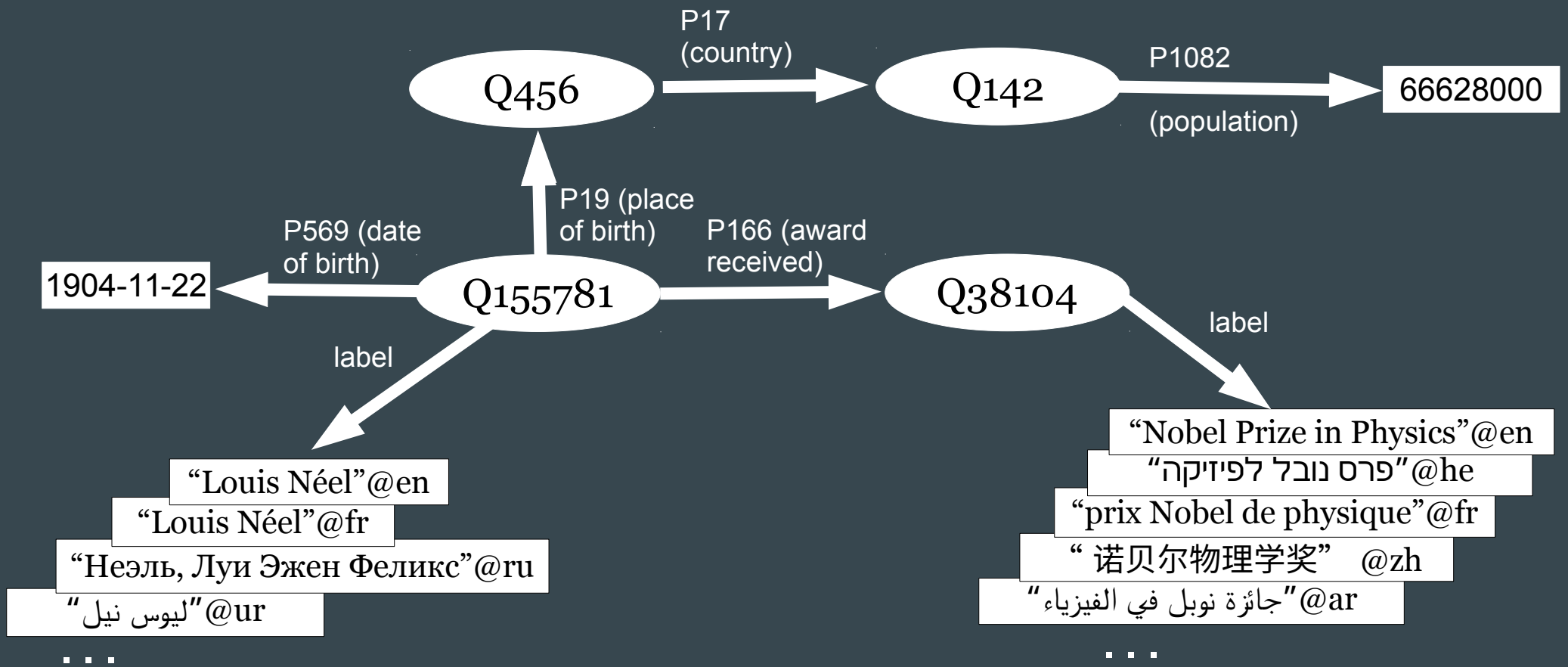
Wikidata: One Graph for Many Languages



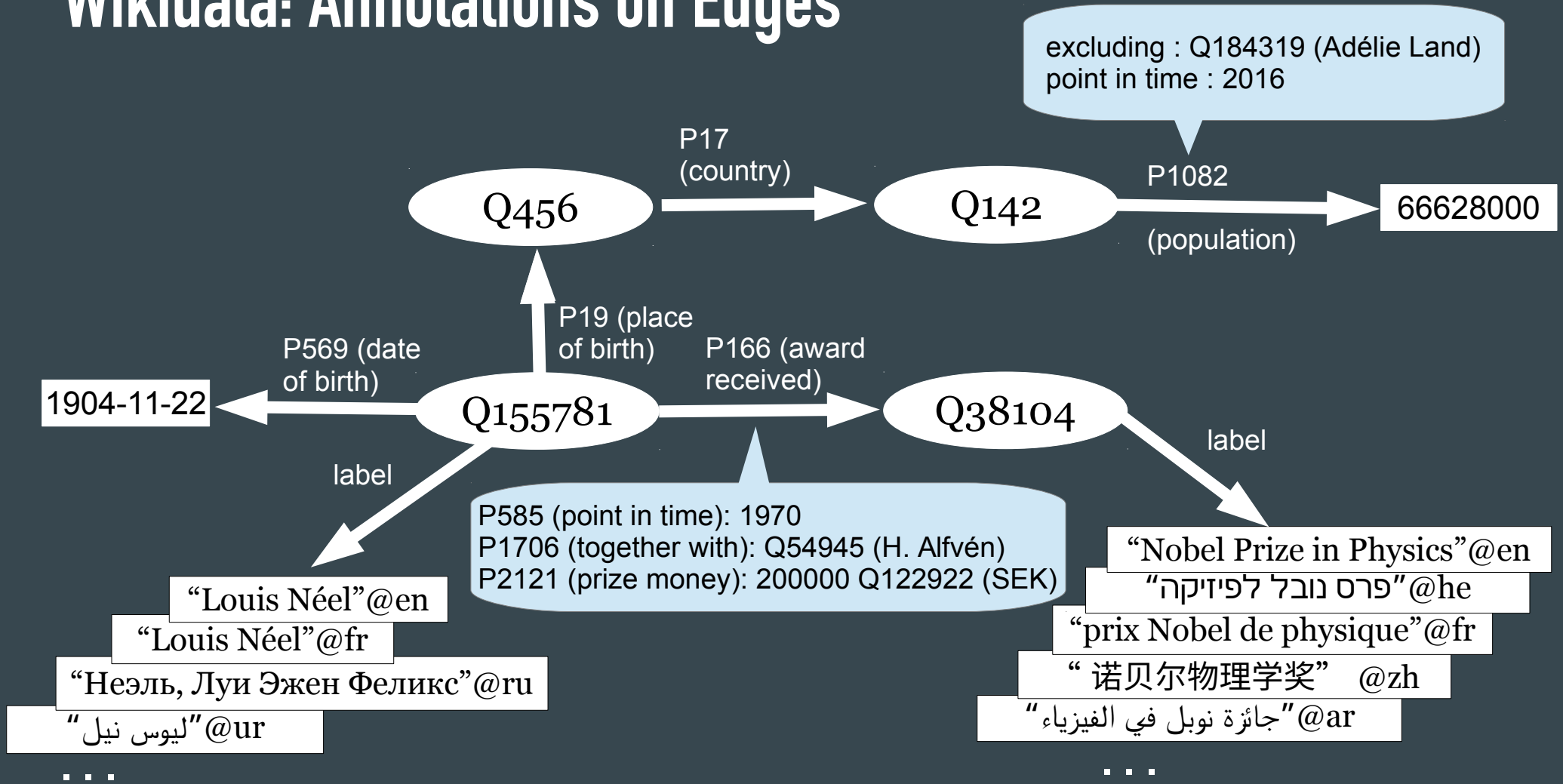
Wikidata: One Graph for Many Languages



Wikidata: One Graph for Many Languages



Wikidata: Annotations on Edges



A Not-So-Simple Idea (2012): Wikidata

Louis Néel (Q155781)

French physicist

Louis Neel | Louis Eugène Felix Néel

award received	Nobel Prize in Physics	edit
	point in time	1970
	together with	Hannes Alfvén
	prize money	200,000 Swedish krona

2 references

copy

reference URL	http://www.nobelprize.org /nobel_prizes/physics/laureates

Wikidata in April 2018

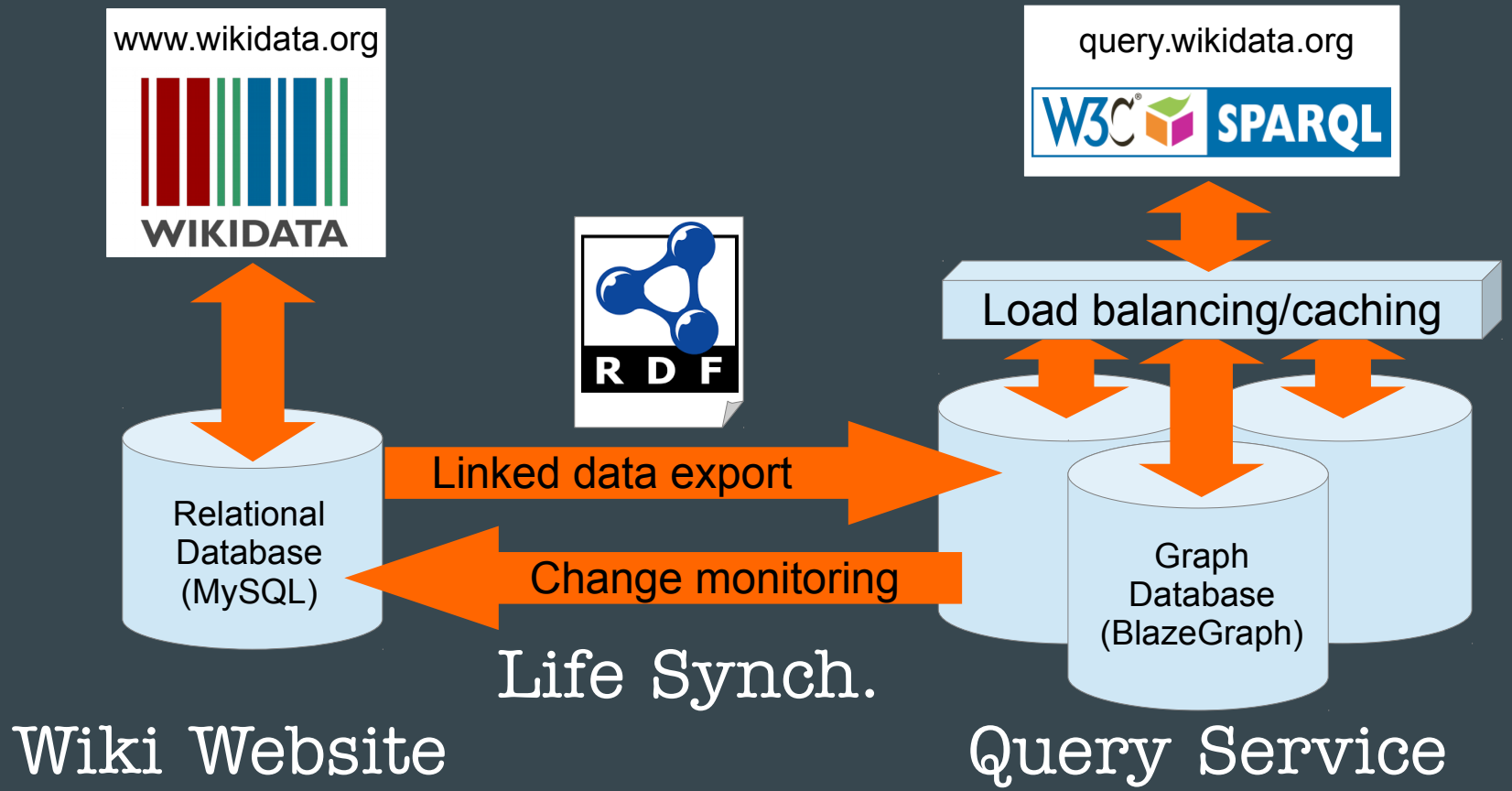
- >400M statements on >45M entities
- >60M links to Wikipedia articles
- >200M labels and aliases
- >1,200M disambiguating descriptions
- >200K registered contributors

“How to query Wikidata?”

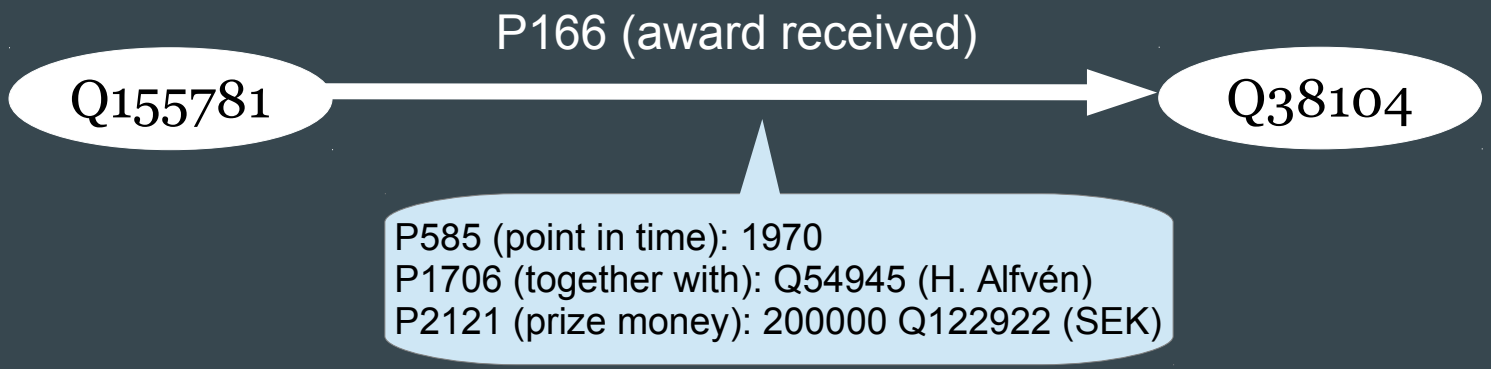
The Wikidata Query Service

Wiki Website

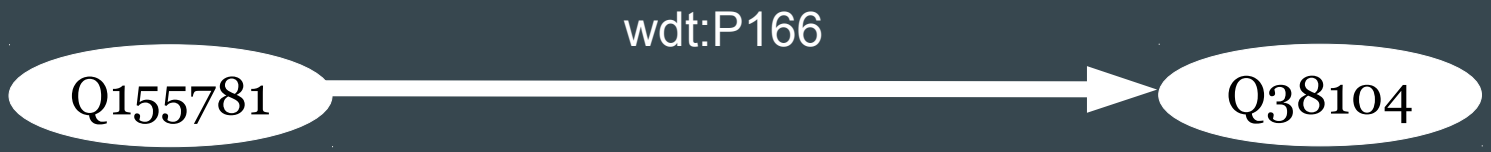
The Wikidata Query Service



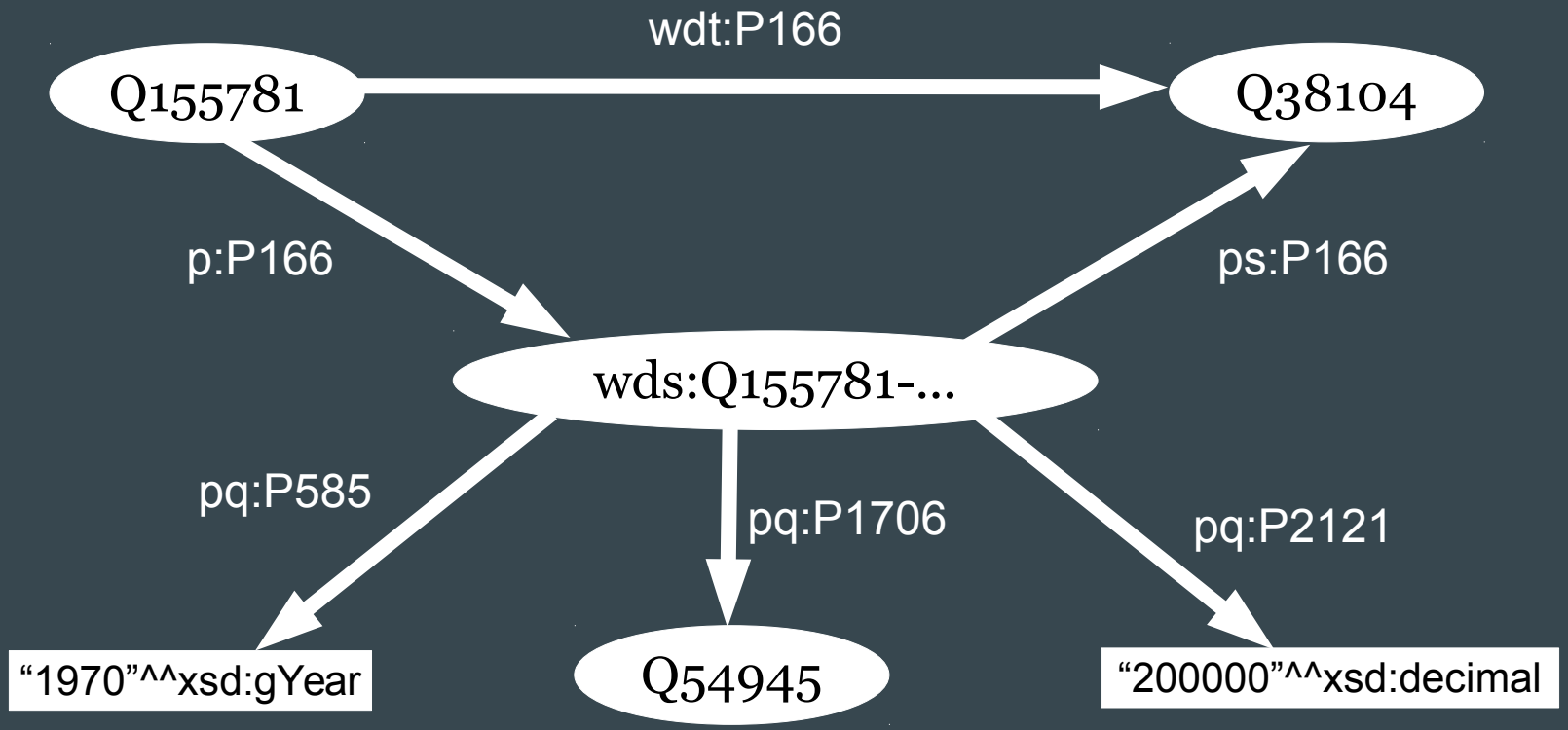
From Wikidata (rich graphs) to RDF (plain graphs)



From Wikidata (rich graphs) to RDF (plain graphs)



From Wikidata (rich graphs) to RDF (plain graphs)



[Erxleben et al., Int. Semantic Web Conf. 2014]

From Wikidata (rich graphs) to RDF (plain graphs)

- Statements get own objects in graph
- Some simple statements also stored directly
- Each Wikidata property becomes many RDF properties
- Complex values get own objects too (not shown)

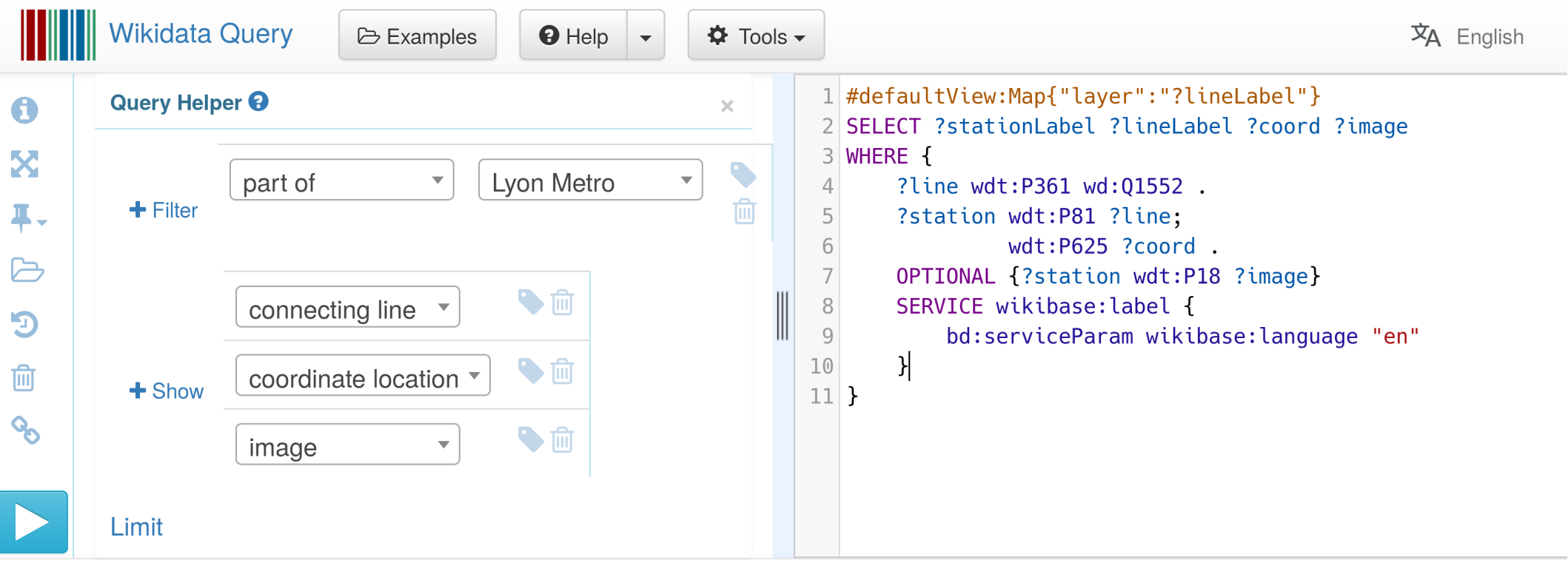
Wikidata RDF Exports

- Weekly full dumps
  - Currently 4.9 billion triples (32 GBit Turtle compressed)
  - At https://dumps.wikimedia.org/wikidatawiki/entities/
- Linked Data Exports
  - Live data in many formats
  - E.g., http://www.wikidata.org/wiki/Special:EntityData/Q42.nt

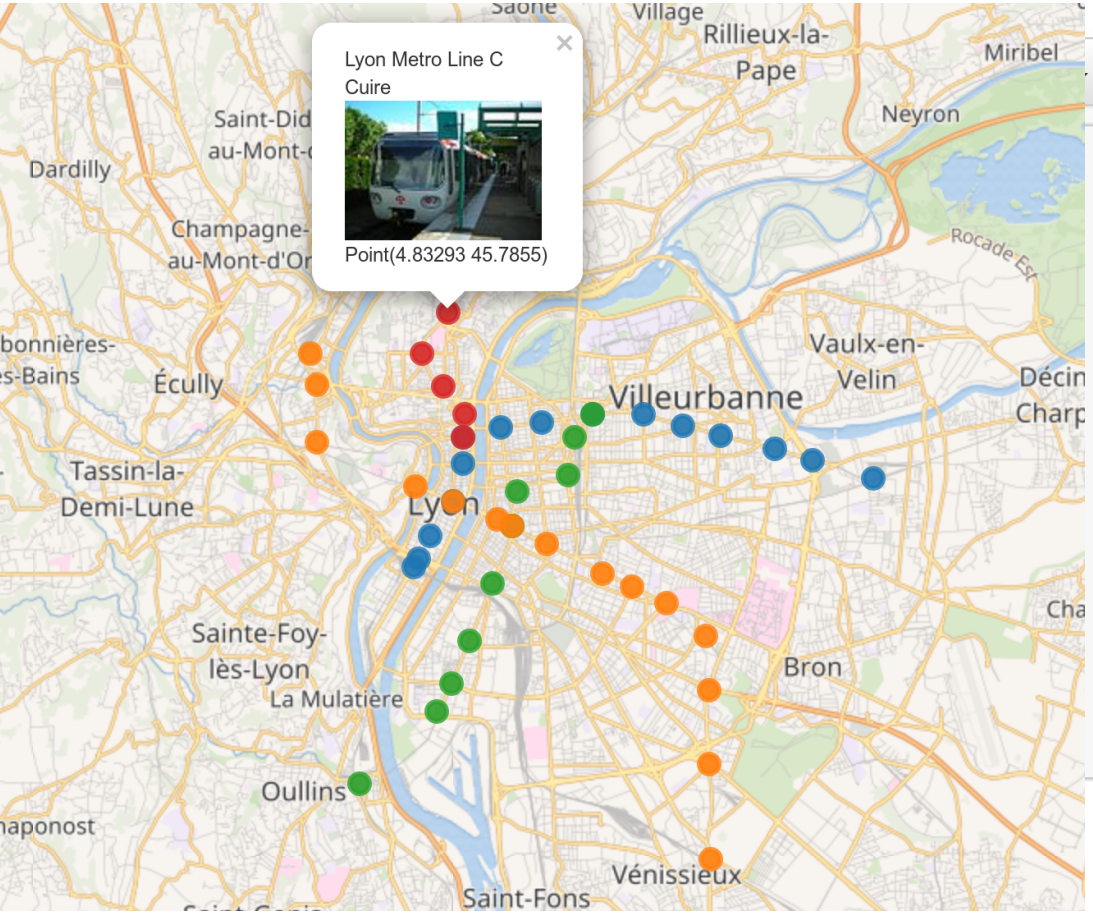
Wikidata SPARQL Query Service

- Official query service since mid 2015
  - User interface at https://query.wikidata.org/
- All the data (4.9B triples), live (latency<60s)
- No limits (well, almost):
  - 60sec timeout
  - No limit on result size (!)
  - No limit on query numbers per IP
  - Clients might be paused after too many parallel requests

A simple SPARQL query



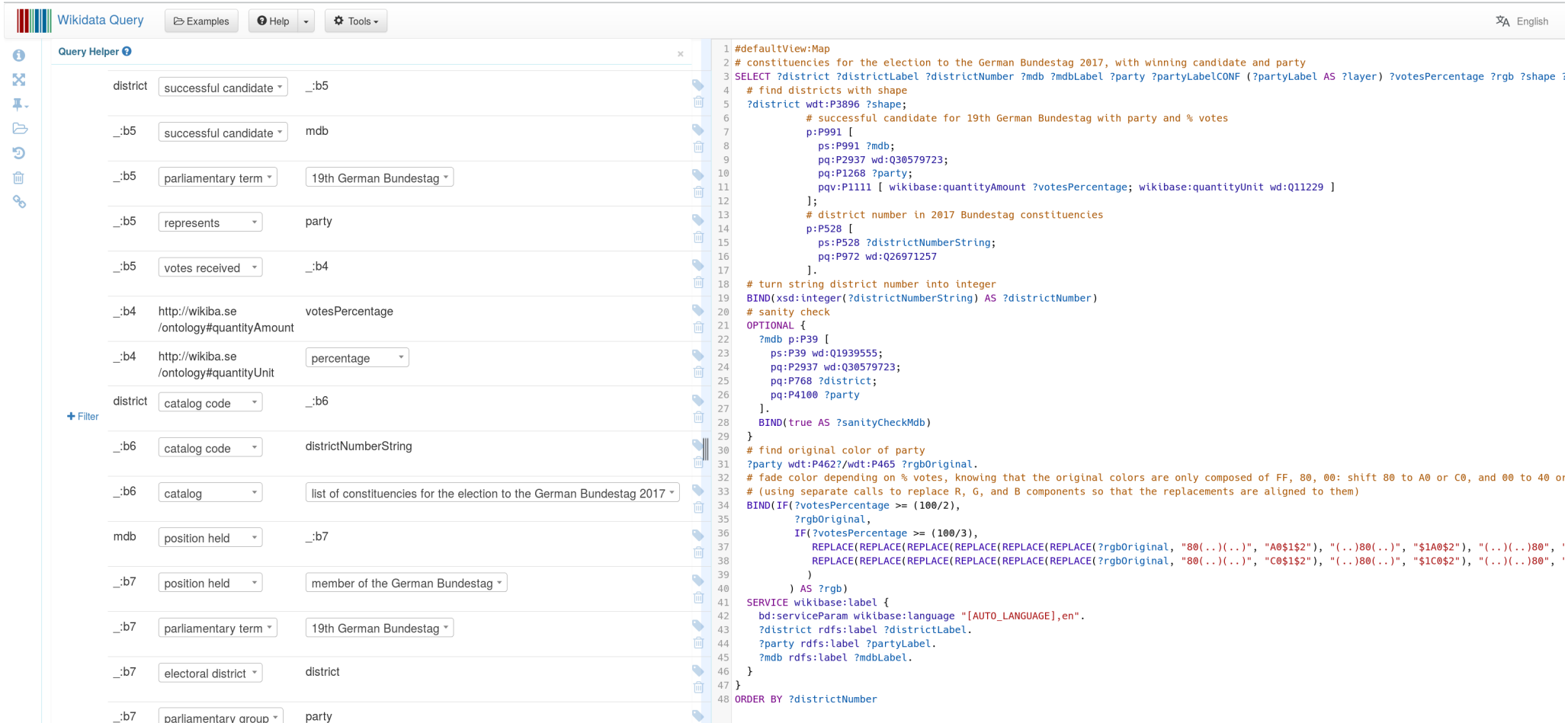
A simple SPARQL query



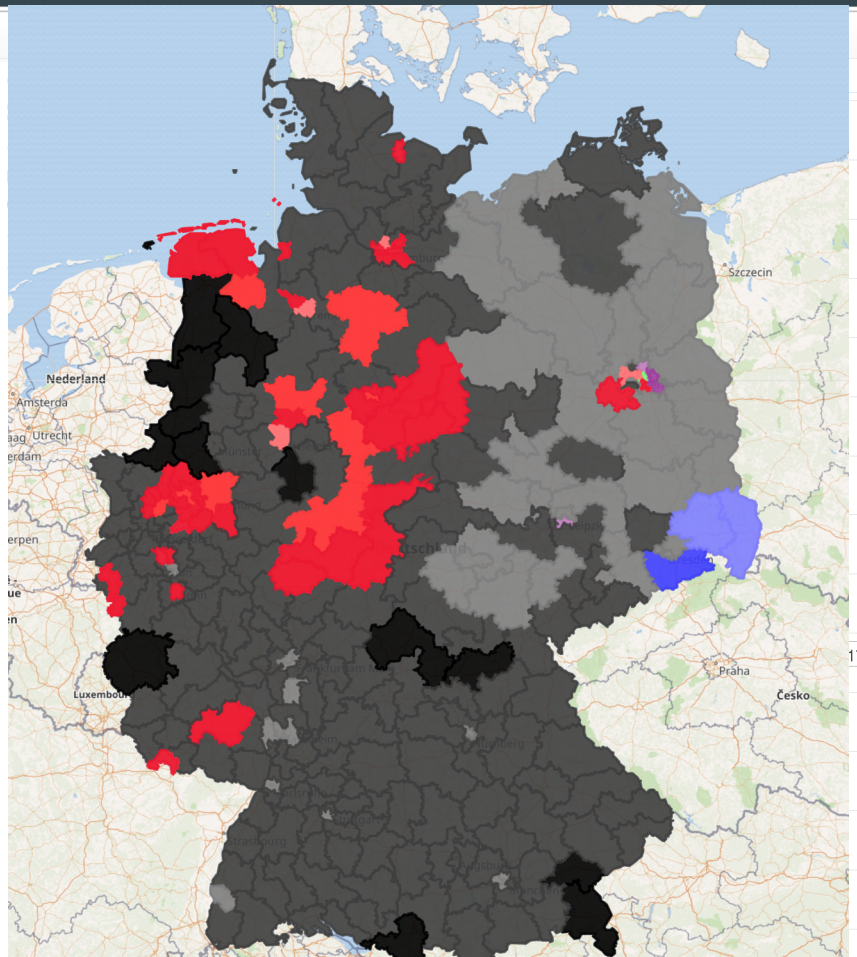
English

```
#defaultView:Map{"layer":"?lineLabel"}
SELECT ?stationLabel ?lineLabel ?coord ?image
WHERE {
    ?line wdt:P361 wd:Q1552 .
    ?station wdt:P81 ?line;
             wdt:P625 ?coord .
    OPTIONAL {?station wdt:P18 ?image}
    SERVICE wikibase:label {
        bd:serviceParam wikibase:language "en"
    }
}
```

A not-so-simple SPARQL query



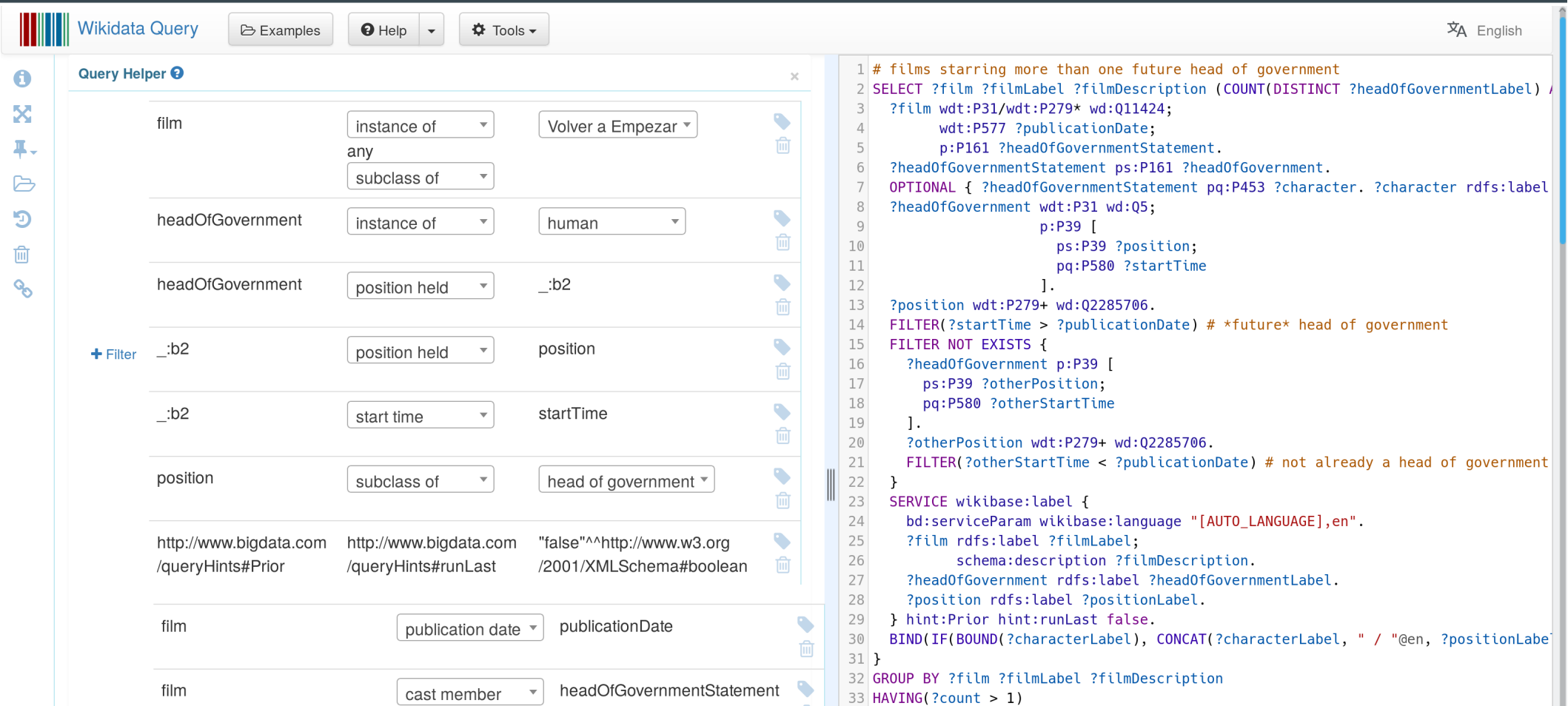
A not-so-simple SPARQL query



English

```
#defaultView:Map
# constituencies for the election to the German Bundestag 2017, with winning candidate and party
SELECT ?district ?districtLabel ?districtNumber ?mdb ?mdbLabel ?party ?partyLabelCONF (?partyLabel AS ?layer) ?votesPercentage ?rgb ?shape ?
  # find districts with shape
  ?district wdt:P3896 ?shape;
            # successful candidate for 19th German Bundestag with party and % votes
            p:P991 [
              ps:P991 ?mdb;
              pq:P2937 wd:Q30579723;
              pq:P1268 ?party;
              pqv:P1111 [ wikibase:quantityAmount ?votesPercentage; wikibase:quantityUnit wd:Q11229 ]
            ];
            # district number in 2017 Bundestag constituencies
            p:P528 [
              ps:P528 ?districtNumberString;
              pq:P972 wd:Q26971257
            ].
  # turn string district number into integer
  BIND(xsd:integer(?districtNumberString) AS ?districtNumber)
  # sanity check
  OPTIONAL {
    ?mdb p:P39 [
      ps:P39 wd:Q1939555;
      pq:P2937 wd:Q30579723;
      pq:P768 ?district;
      pq:P4100 ?party
    ].
    BIND(true AS ?sanityCheckMdb)
  }
  # find original color of party
  ?party wdt:P462?/wdt:P465 ?rgbOriginal.
  # fade color depending on % votes, knowing that the original colors are only composed of FF, 80, 00: shift 80 to A0 or C0, and 00 to 40 or
  # (using separate calls to replace R, G, and B components so that the replacements are aligned to them)
  BIND(IF(?votesPercentage >= (100/2),
          ?rgbOriginal,
          IF(?votesPercentage >= (100/3),
             REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(?rgbOriginal, "80(..)(..)", "A0$1$2"), "(..)80(..)", "$1A0$2"), "(..)(..)80", "
             REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(?rgbOriginal, "80(..)(..)", "C0$1$2"), "(..)80(..)", "$1C0$2"), "(..)(..)80", "
             )
          ) AS ?rgb)
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en".
    ?district rdfs:label ?districtLabel.
    ?party rdfs:label ?partyLabel.
    ?mdb rdfs:label ?mdbLabel.
  }
}
ORDER BY ?districtNumber
```

An advanced SPARQL query



You expect normal people to use SPARQL?!

- If they want … it's really not that difficult
  - Extensive online documentation
  - Over 300 example queries
  - Tutorials and workshops at community events
- But SPARQL is often hidden from users
  - Embedded results on Web pages (incl. Wikipedia)
  - Mobile apps and online apps
  - Crowdsourcing platforms

Wikidata:Request a query

Shortcut: WD:RAQ

This is a page where SPARQL 1.1 Query Language (Q32146616) queries can be requested. Please provide feedback if a query is written for you.

For sample queries, see Examples. Property talk pages include also summary queries for these.

For help writing your own queries, or other questions *about* queries, see Wikidata talk:SPARQL query service/queries.

Help resources about Wikidata Query Service (Q20950365) and SPARQL: Wikidata:SPARQL query service/Wikidata Query Help and Category:SPARQL.

Fishing in the Wikidata river requires both an idea where to look for fish and a suitable fishing method. If you have the former, this page can help you find the latter.

Contents [hide]

1 Slide show with images
2 Retrieve property if available
3 Surname lookup
4 What's in Wikipedia lists?
5 Properties missing a label or description in a given language
6 P: Properties for a set of items
7 About population
8 SPARQL for Q5 externalid statistics
9 Who held what position in the year 420 ?

Current Usage

- SPARQL is widely used
  - >100M requests per month (3.8M per day) in 2018
- Excellent availability and performance
  - 50% of queries answered in <40ms (95% in <440ms; 99% in <40s)
  - Less than 0.05% of queries time out
  - Service has never been down so far
- All software/customisations free & open source
  - See https://github.com/wikimedia/wikidata-query-rdf

“What can we learn from all these SPARQL queries?”

SPARQL Queries Are Interesting

- Which data is actually asked for?
- Which SPARQL features are most important?
- Who is using SPARQL through which tools?

We have analysed complete Wikidata SPARQL query logs (Wikimedia Research Collaboration)

[Bielefeldt et al., "Linked Data on The Web" @ WWW 2018]

Analysing SPARQL logs: The Bot Problem

Analysing SPARQL logs: The Bot Problem

- Query traffic is **ruled** by a few bots

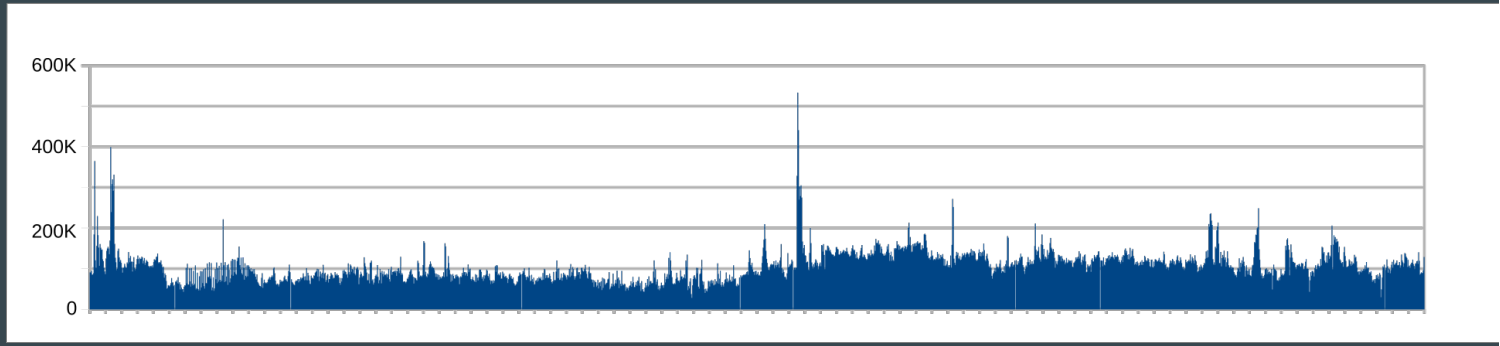


Fig.: Wikidata SPARQL traffic Jun-Sep 2017

Analysing SPARQL logs: The Bot Problem

- Query traffic is **ruled** by a few bots

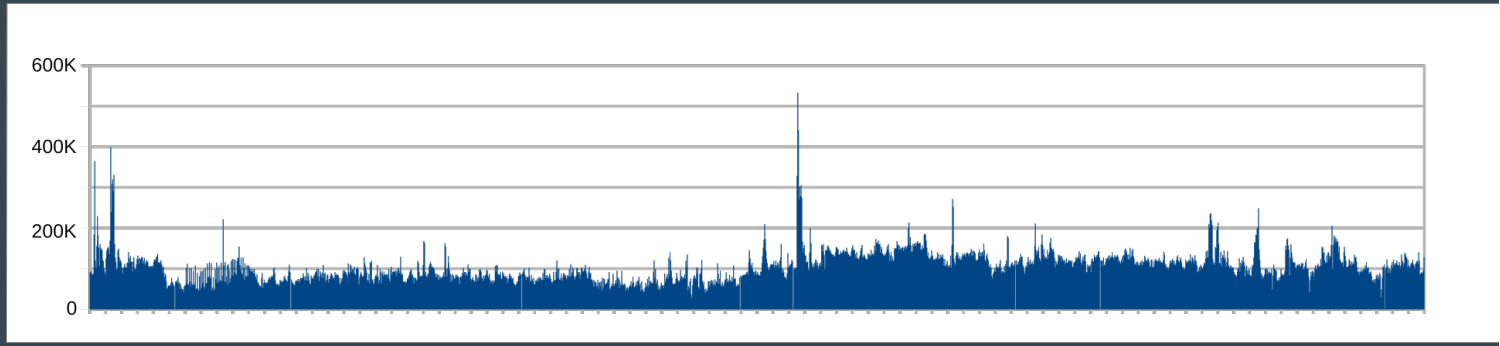


Fig.: Wikidata SPARQL traffic Jun-Sep 2017

- 41% of all Wikidata query traffic from June – September 2017 caused by one super-power user (Magnus Manske)

Analysing SPARQL logs: The Bot Problem

- Query traffic is **ruled** by a few bots

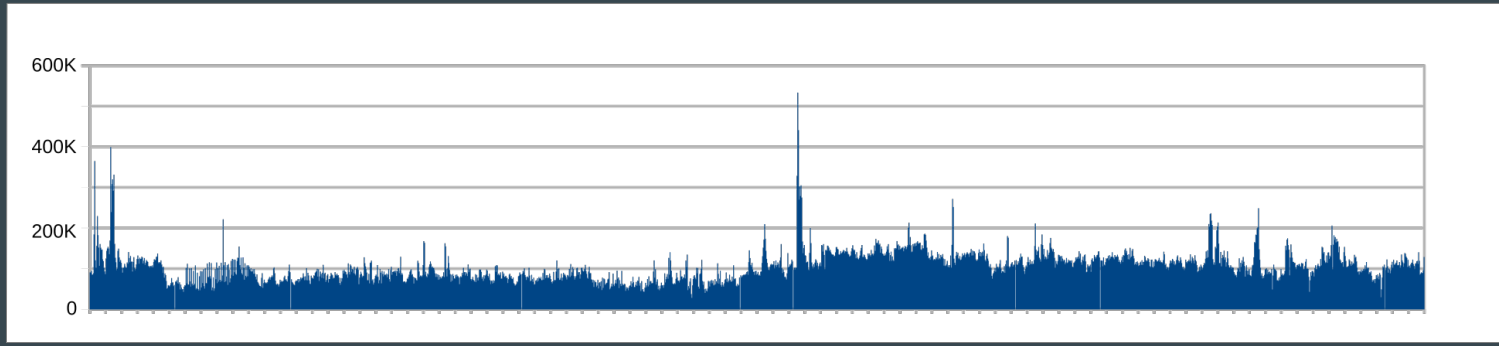


Fig.: Wikidata SPARQL traffic Jun-Sep 2017

- 41% of all Wikidata query traffic from June – September 2017 caused by one super-power user (Magnus Manske)
- The effect does **not** average out, and it affects other sites too

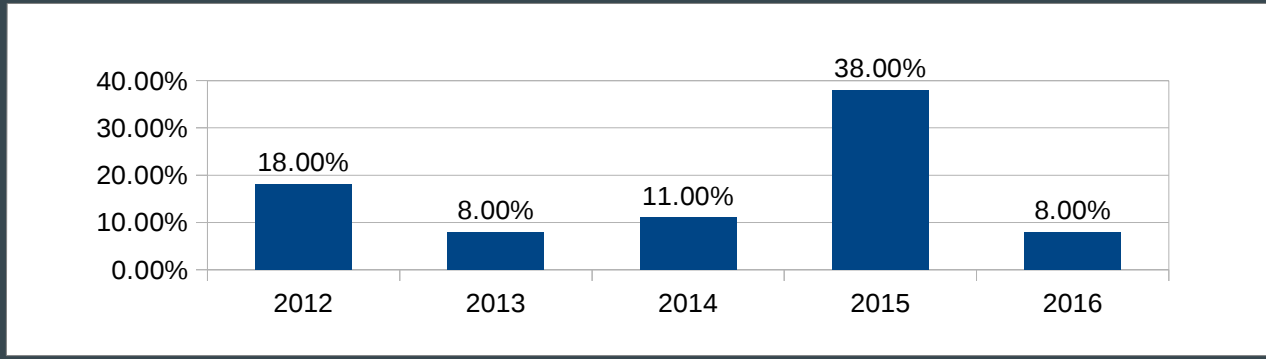


Fig.: Usage of DISTINCT on DBpedia [Bonifati et al. 2017]

Analysing SPARQL logs: The Bot Probl

- Query traffic is **ruled** by a few bots

Fig.: Wikidata SPARQL traffic Jun-Sep 2017

- 41%

une – September 2017
(Magnus Manske)

sites too

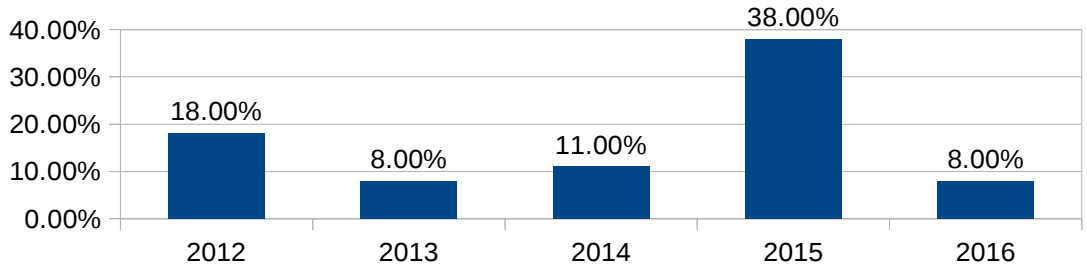


Fig.: Usage of DISTINCT on DBpedia [Bonifati et al. 2017]

No trends!
No predictability!
No insights!

Are SPARQL queries interesting after all?

- Observation: Robotic traffic dominates
  - May not represent any real interest
  - Governed by very few sources
  - Random changes – not uniform on any observed scale

Are SPARQL queries interesting after all?

- Observation: Robotic traffic dominates
  - May not represent any real interest
  - Governed by very few sources
  - Random changes – not uniform on any observed scale
- Hypothesis: Organic traffic also exists
  - Representing human information need during some interaction
  - Composed of many diverse sources
  - Continuous change over months

Note: “Organic” ≠ “hand-written SPARQL” (user apps might use SPARQL to get user-requested data without users actually writing queries)

Extracting organic traffic

- Main signal: User Agents
  - Assumption: organic traffic generally from browser-like agents

Extracting organic traffic

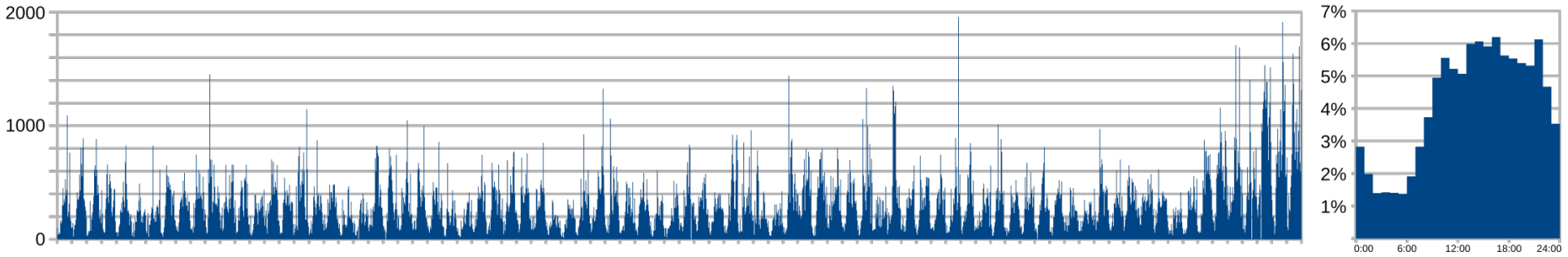
- Main signal: User Agents
  - Assumption: organic traffic generally from browser-like agents
- 2nd signal: query comments
  - Some browser-based tools mark queries using comments
- 3rd signal: activity spikes
  - Group queries by query pattern (following [Raghuveer, USEWOD'12])
  - Find agent-pattern pairs that spike (>2K requests/month)
  - Manually inspect these queries to decide if organic or robotic

→ About 300 further browser-based sources classified "robotic"

Results: Organic component

- Jun–Sep 2017: 658,890 queries (<0.5%)

Temporal distribution of organic queries (12 weeks / time of day)



- More triples
  organic 17%: 1, 97%: ≤11 vs. robotic 57%: 1, 96%: ≤7
- More varied (vocabulary, SPARQL features)

Insights on Wikidata Usage

- Robotic traffic:
  - Mainly information integration bots (comparing database contents)
  - Potentially also selective data download (spider-like)
  - Most queries from a few dominant bots (>60% from top-three bots)
- Organic traffic:
  - Data browsers (often general-purpose)
  - Mobile apps (often topical)
  - Most queries from of unidentified “small” sources
- Reified statements in 4%–10% of queries

What's Next?

More data

- Wikidata: >45M items with >400M statements

More data

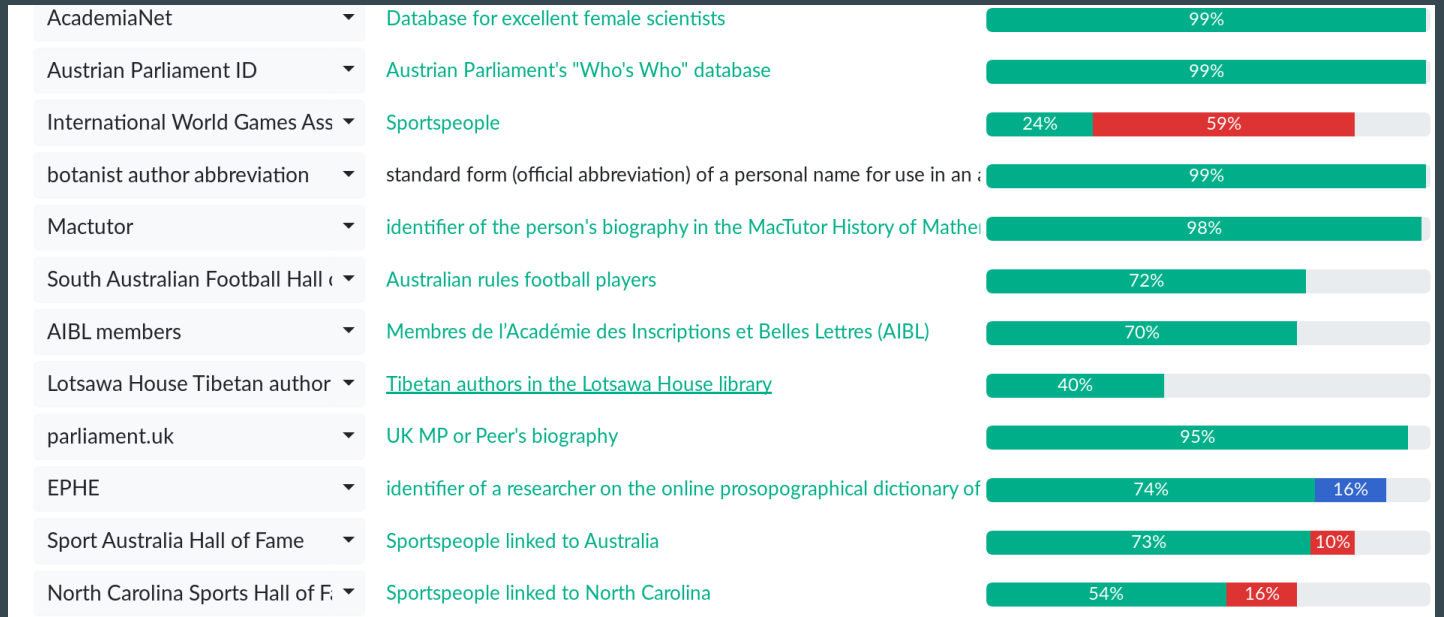
- Wikidata: >45M items with >400M statements
  - OSM: >4B nodes, >230M buildings, >10M trees
  - WDC: >9.5B entities, >38B RDF triples
- Why don’t we just import everything?!

More data

- Wikidata: >45M items with >400M statements
  - OSM: >4B nodes, >230M buildings, >10M trees
  - WDC: >9.5B entities, >38B RDF triples
- Why don’t we just import everything?!
  - Notability? Well, sometimes …
  - Community support! Who will maintain this?

More data: current efforts

- Data donation guidelines
- Wikidata aligns with >2500 databases and catalogues
- Supervised data alignment with crowdsourcing (Mix'n'Match)



More data: current efforts

Soccerdonna

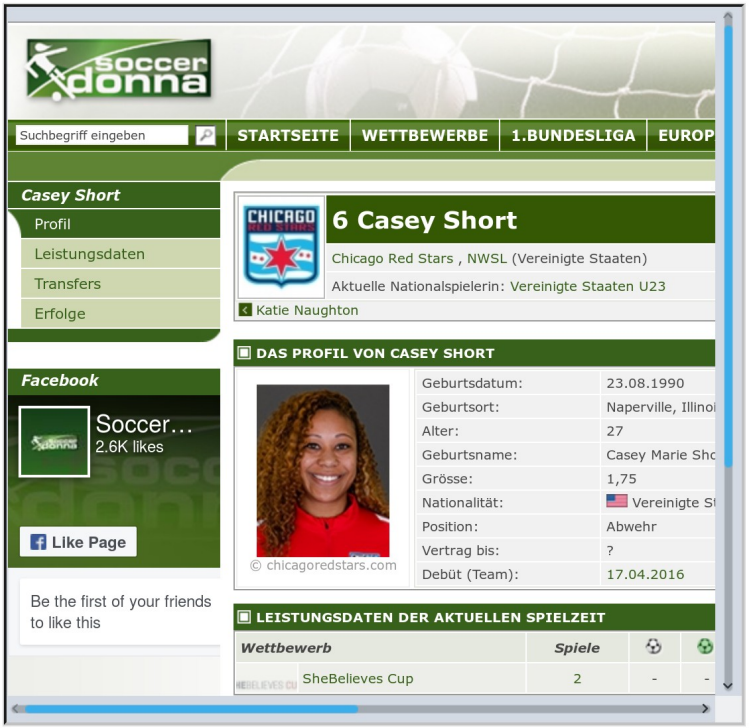
Soccerdonna website female association football player db

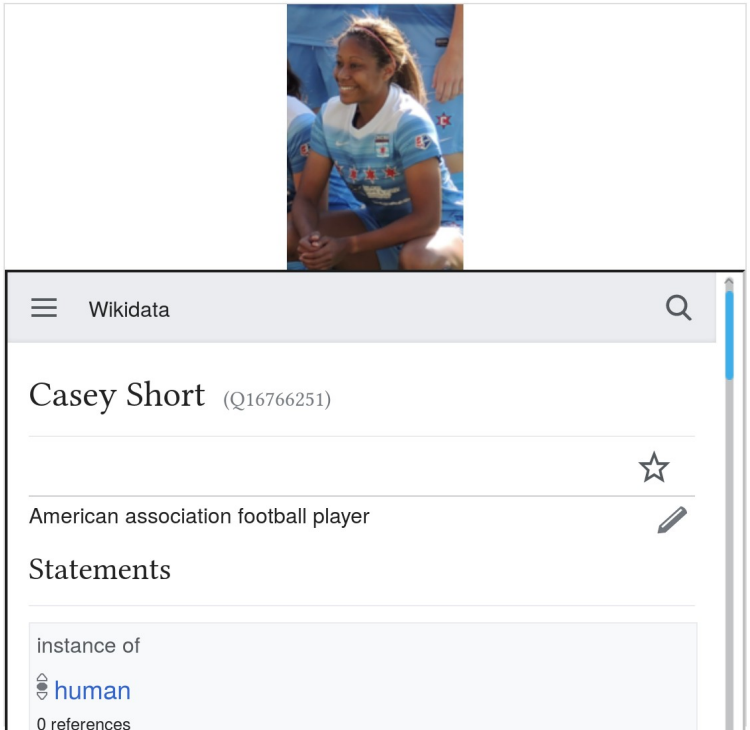
Markus Krötzsch

Load next entry on empty search results

Casey Short

' player, born 23.08.1990 at Naperville, Illinois plays '





New kinds of data

- Coming soon: **lexical data** (dictionary/thesaurus)
  - Exciting & dangerous
- Planned: **media (meta-)data** (Wikimedia Commons)
- Factual knowledge that is not in catalogues?
- Common sense?

In many cases: technical changes/extensions needed

Quality!

- Errors, spam, vandalism
- Global coherency of modelling
- Sources & alignments
- Incompleteness
- Change & data rot

Germany (Q183)

federal parliamentary republic in central-western Europe

FRG | BRD | Bundesrepublik Deutschland | Federal Republic of Germany | de |

basic form of government

federal parliamentary republic

edit

copy

copy

Potential issues

conflicts-with constraint Help Discuss

An entity should not have a statement for basic form of government if it also has a statement for instance of with value republic.

one-of constraint Help Discuss

The value for basic form of government should be one of the following:

- republic
- constitutional monarchy
- federal republic
- representative democracy
- parliamentary system
- soviet republic

edit label

Nauru (Q697)

Republic of Nauru | Pleasant Island | Naoero | nr |

republic in Oceania

head of state

2+28 statements

Baron Waqa (Nauruan politician) ★
start time : 2013-06-11

Sprent Dabwido (president of Nauru)

Frederick Pitcher (President of Nauru) (Proposal)
start time : 2011-11-10
end time : 2011-11-15
Source: MARS

Marcus Stephen (Nauruan sportperson and politician) (Proposal)
start time : 2007-12-19
end time : 2011-11-10
Source: MARS

Ludwig Scotty (Nauruan politician, president) (Proposal)
start time : 2004-06-22
end time : 2007-12-19
Source: MARS

[Marx & MK, International Semantic Web Conf. 2017]

https://tools.wmflabs.org/sqid/

edit label

Frederick Pitcher (Q917601)

position held	President of Nauru (head of state and government in Nauru) start time : 2011-11-10 end time : 2011-11-15 replaces : Marcus Stephen (Nauruan sportperson and politician) replaced by : Sprent Dabwido (president of Nauru)

edit label

Nauru (Q697)

office held by head of government	President of Nauru (head of state and government in Nauru)

A rule of inference:

```
(?headOfState.position held^P39 = ?headOffice)@?X,
(?country.office held by head of state^P1906 = ?headOffice)@?Y
→ (?country.head of state^P35 = ?headOfState)@{start time^P580 = ?X.start time^P580,
end time^P582 = ?X.end time^P582}
```

[Marx et al., International Joint Conf. On Artif. Intellig. 2017]

Conclusion and Outlook

- Wikidata is a fascinating, fast-moving project
  - Large amounts of quality data & much more to come
  - Data export and analysis services for all needs
  - Innovation-friendly community
- Many unsolved questions for research
  - Quality, provenance, social aspects, performance challenges, data integration, internationalisation, …

Literature

- Adrian Bielefeldt, Julius Gonsior, Markus Krötzsch: “Practical Linked Data Access via SPARQL: The Case of Wikidata” Proceedings of the WWW2018 Workshop on Linked Data on the Web (LDOW-18), CEUR Workshop
- Fredo Erxleben, Michael Günther, Markus Krötzsch, Julian Mendez, Denny Vrandečić: “Introducing Wikidata to the Linked Data Web” In Proceedings of the 13th International Semantic Web Conference (ISWC 2014)
- Maximilian Marx, Markus Krötzsch: “SQID: Towards Ontological Reasoning for Wikidata” In Proceedings of the ISWC 2017 Posters & Demonstrations Track, CEUR Workshop Proceedings. CEUR-WS.org
- Maximilian Marx, Markus Krötzsch, Veronika Thost: “Logic on MARS: Ontologies for generalised property graphs” Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI'17), 1188-1194, 2017

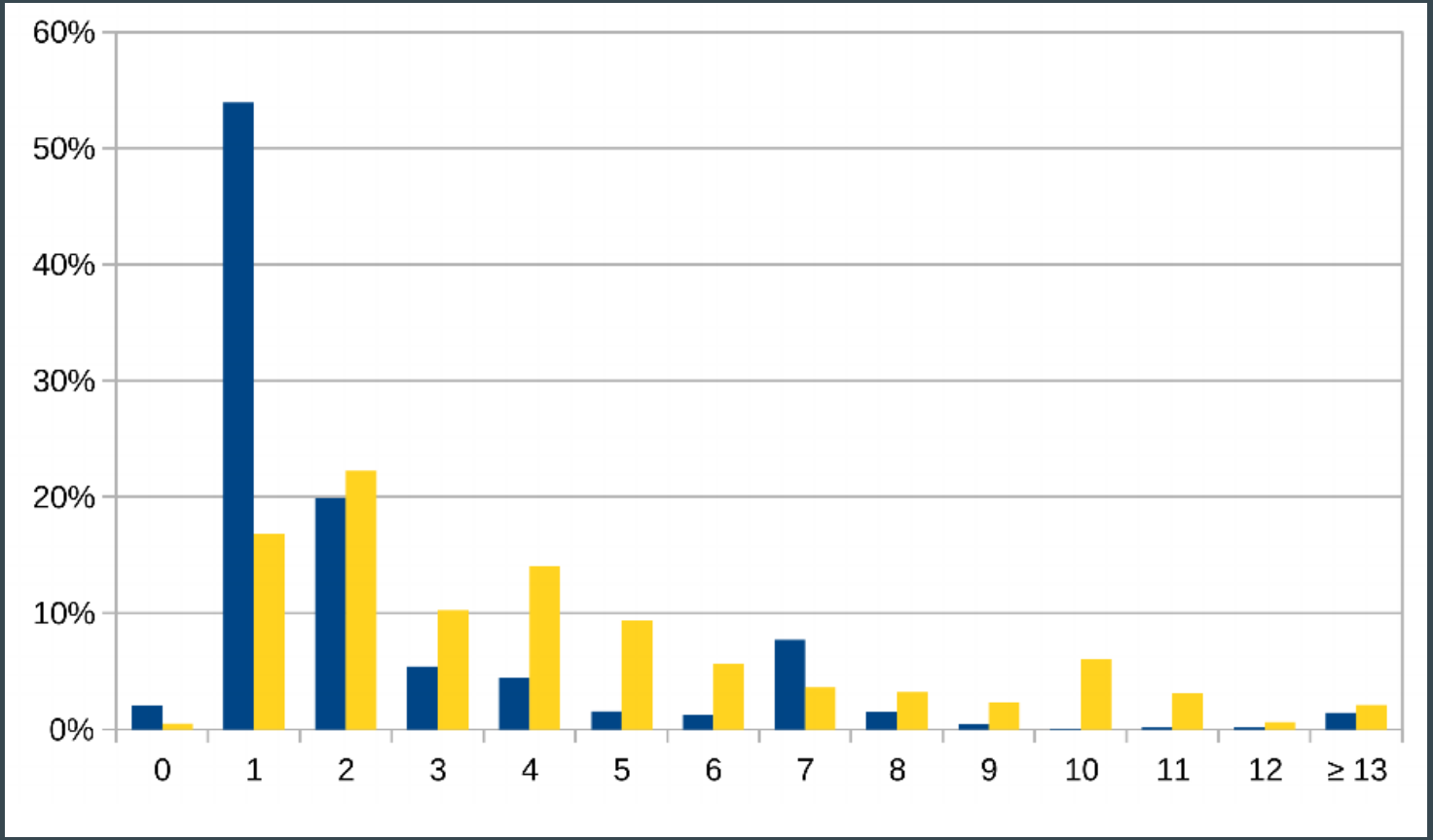
Films with future heads of government

Star in the Dust	1956 film by Charles F. Haas	2	Clint Eastwood, mayor; George Wallace, Governor of Alabama
The Two Who Stole the Moon	1962 Polish film by Jan Batory	2	Jarosław Kaczyński, Prime Minister of Poland; Lech Kaczyński, Mayor of Warsaw
Ragasiya Police 115	1968 film by B. R. Panthulu	2	M. G. Ramachandran, Chief Minister of Tamil Nadu; Jayalalithaa, Chief Minister of Tamil Nadu
Québec : Duplessis et après…	documentary	2	Bernard Landry, Premier of Quebec; René Lévesque, Premier of Quebec
Q3541438	1994 film by Claude Lanzmann	2	Ariel Sharon, Prime Minister of Israel; Ehud Barak, Prime Minister of Israel
Batman & Robin	1997 American superhero film based on the DC Comics character Batman	2	Arnold Schwarzenegger, Mr. Freeze / Governor of California; Jesse Ventura, Governor of Minnesota

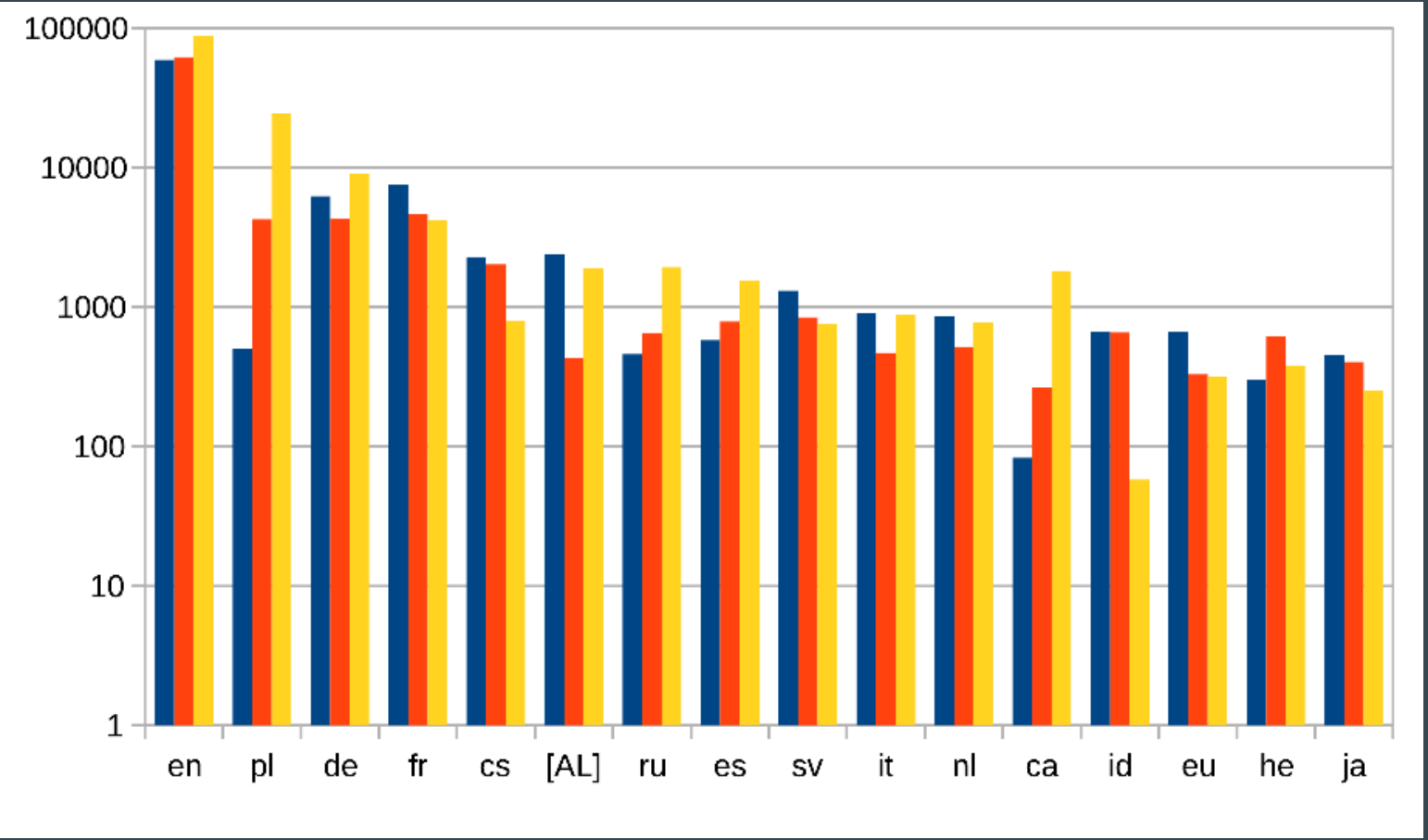
SPARQL Feature Distribution (2017/2018)

	organic						robotic					
	Jun 2017	Jul 2017	Aug 2017	Jan 2018	Feb 2018	Mar 2018	Jun 2017	Jul 2017	Aug 2017	Jan 2018	Feb 2018	Mar 2018
Limit	31.08%	39.55%	46.56%	52.31%	51.23%	36.87%	21.12%	16.86%	17.42%	20.38%	11.47%	15.17%
Distinct	26.50%	31.40%	19.05%	59.30%	60.42%	63.78%	15.84%	5.48%	4.27%	4.32%	7.54%	12.25%
Order By	17.29%	14.75%	13.22%	46.89%	46.99%	34.53%	12.97%	8.01%	6.78%	8.76%	7.68%	17.46%
Offset	0.40%	2.92%	0.37%	0.09%	0.08%	0.06%	7.73%	6.07%	6.29%	0.10%	0.07%	0.10%
Join	87.59%	87.82%	89.76%	82.50%	91.70%	87.02%	88.48%	78.53%	67.41%	73.26%	61.39%	70.19%
Optional	42.36%	46.24%	55.92%	50.90%	41.30%	41.15%	25.08%	11.63%	11.45%	12.73%	15.41%	30.71%
Filter	25.89%	29.12%	22.24%	12.59%	11.76%	11.76%	21.64%	17.92%	13.79%	14.70%	16.83%	29.02%
Path with *	15.02%	15.59%	12.88%	40.92%	32.43%	30.34%	16.43%	19.19%	14.80%	20.56%	17.26%	24.81%
Subquery	13.09%	15.30%	12.79%	6.45%	5.07%	5.39%	0.34%	0.28%	0.33%	0.09%	0.13%	0.11%
Bind	9.85%	9.23%	8.68%	4.72%	3.99%	4.15%	16.29%	12.07%	9.60%	11.94%	13.79%	24.03%
Union	5.10%	5.76%	12.62%	2.56%	2.07%	3.39%	11.26%	8.63%	7.61%	13.96%	13.05%	18.57%
Values	4.44%	3.07%	10.88%	3.29%	3.23%	3.20%	35.72%	30.74%	28.92%	29.82%	23.80%	11.90%
Not Exists	3.31%	3.37%	2.46%	1.24%	0.94%	0.69%	0.19%	0.21%	0.19%	0.27%	0.29%	0.35%
Minus	2.04%	2.91%	1.60%	0.82%	0.57%	0.71%	0.53%	0.92%	1.07%	1.46%	1.26%	1.78%
Service (lang)	44.63%	42.09%	54.78%	50.88%	41.71%	42.95%	10.40%	6.15%	4.27%	7.15%	7.91%	8.90%
Service (other)	11.49%	10.53%	10.32%	7.30%	13.14%	2.31%	4.51%	0.19%	1.16%	0.17%	0.18%	0.51%
Group By	17.12%	19.93%	13.04%	7.00%	5.40%	5.07%	0.41%	0.37%	0.48%	0.22%	0.23%	0.39%
Sample	8.85%	10.93%	4.60%	1.61%	1.63%	1.69%	0.04%	0.04%	0.06%	0.05%	0.04%	0.10%
Count	7.55%	7.60%	8.15%	5.22%	3.88%	3.73%	1.15%	4.30%	0.30%	1.52%	0.65%	0.89%
GroupConcat	1.80%	2.79%	1.17%	0.86%	0.86%	0.74%	0.06%	0.09%	0.02%	0.03%	0.02%	0.28%
Having	1.17%	1.14%	0.72%	0.65%	0.26%	0.33%	0.01%	0.01%	0.00%	0.00%	0.00%	0.04%

Triples per query: organic (blue)/robotic (yellow)



Languages of labels in organic queries



SPARQL feature co-occurrence

							organic		robotic	
J	**F**	**O**	**U**	**P**	**V**	**S**	I1–I3	I4–I6	I1–I3	I4–I6
		(none)					8.04	9.22	19.67	27.67
J							13.29	31.35	11.26	10.09
	F						1.10	0.98	1.92	1.31
J	F						6.68	2.39	2.61	1.68
J				P			2.98	1.62	13.50	13.94
J	F			P			2.48	0.58	0.39	0.07
J					V		0.39	2.01	30.42	17.47
		O					1.26	1.64	0.11	0.63
J		O					22.32	7.04	1.86	1.95
J		O		P			2.07	29.10	0.35	0.05

							organic		robotic	
J	**F**	**O**	**U**	**P**	**V**	**S**	I1–I3	I4–I6	I1–I3	I4–I6
J	F	O					2.66	1.32	2.13	1.18
J		O	U				3.49	0.25	0.02	0.01
J		O			V		3.38	0.41	0.11	0.43
J		O		P	V		1.01	0.06	0.16	0.07
J						S	2.76	1.41	0.06	0.01
J		O				S	4.78	0.62	0.00	0.01
J	F					S	3.19	2.28	0.03	0.01
J	F	O				S	1.02	0.13	0.00	0.00
J	F	O		P			0.79	0.31	0.64	1.58
J			U	P	V		0.01	0.02	0.05	1.92