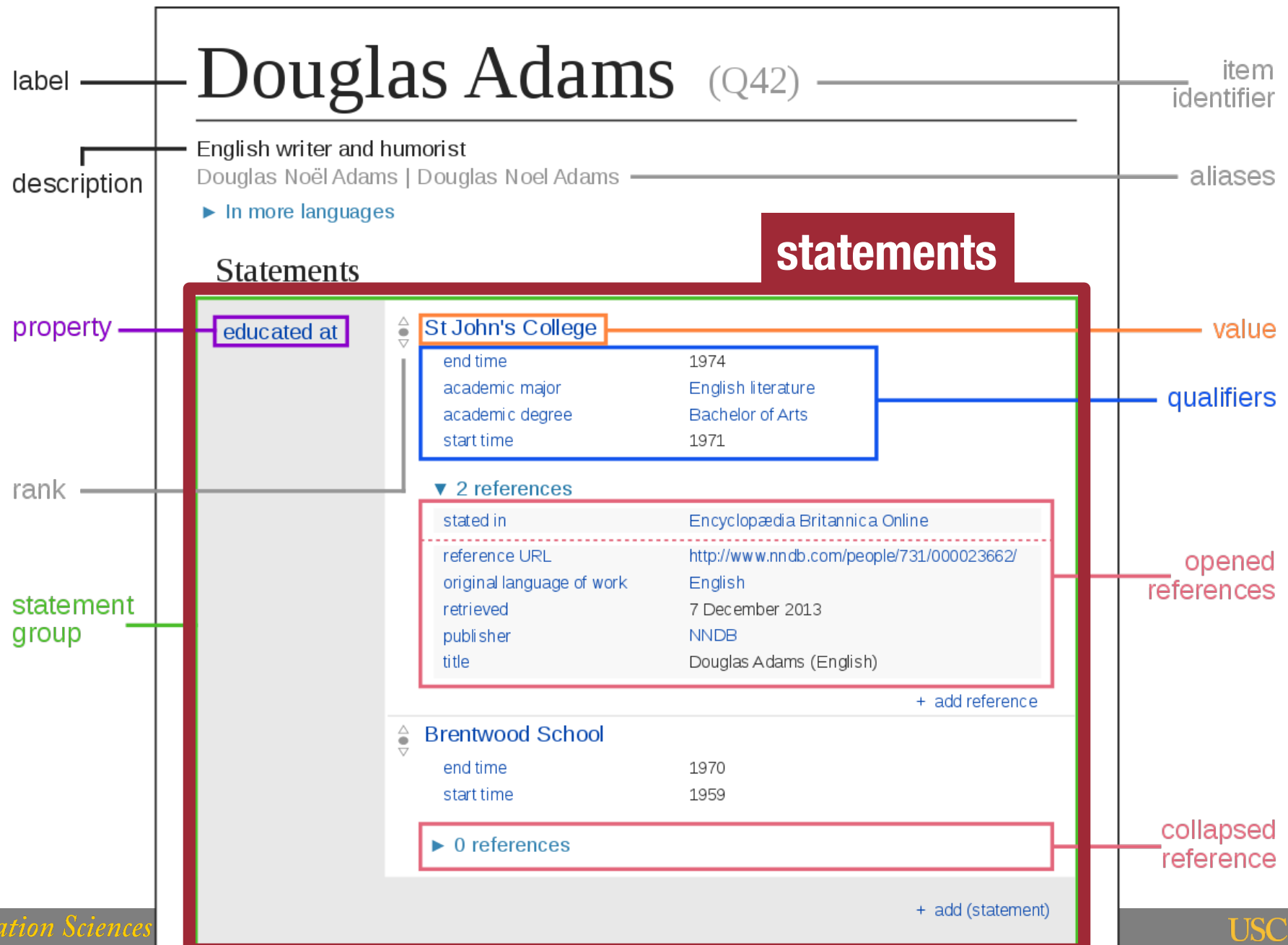


# **Large Knowledge Graphs (cont.)**

**Craig Knoblock**

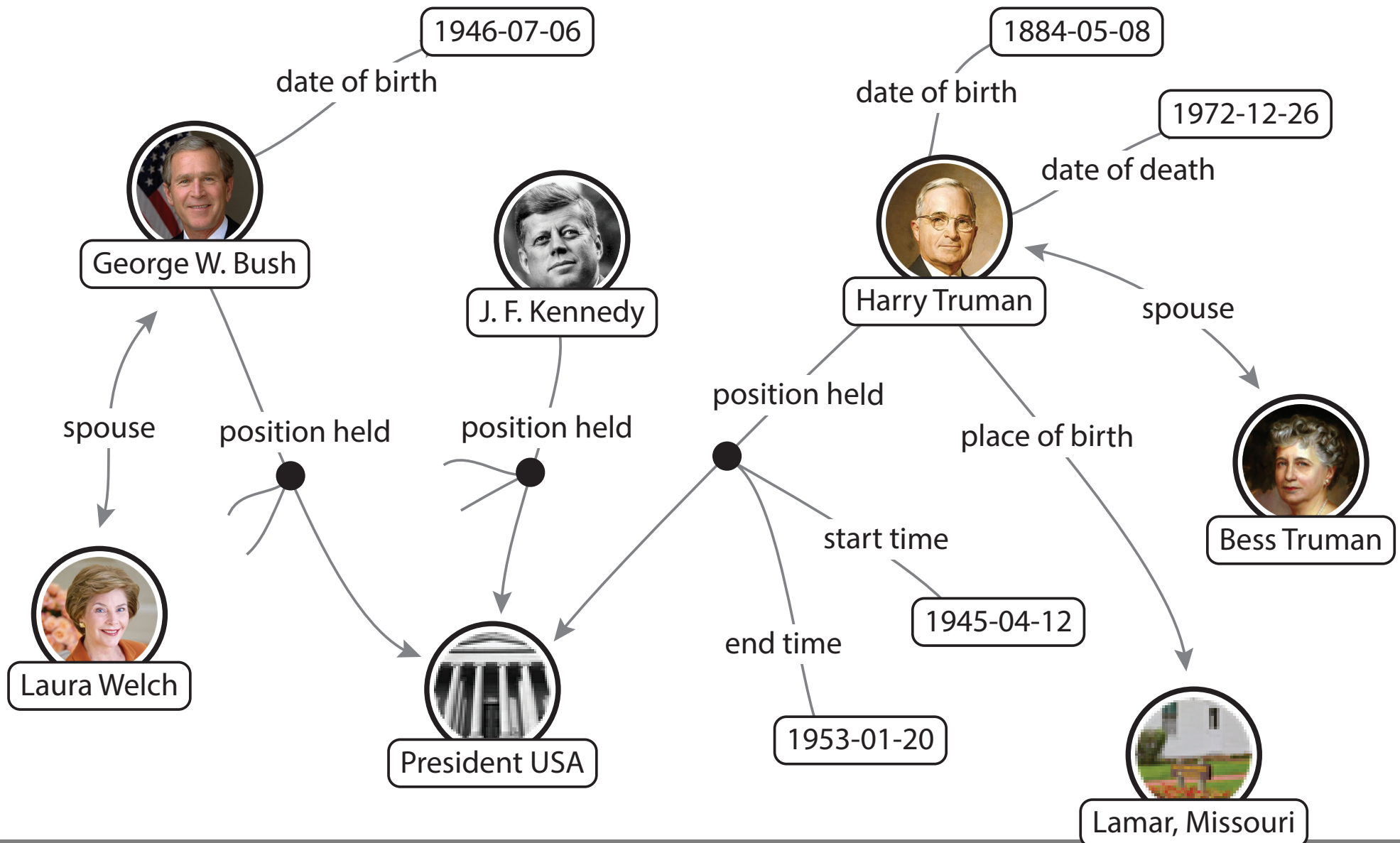
**DSCI-558/CSCI-563**

# Data Model: Statements



# RDF Data Model

# Edges Are Statements (Reification)



# Truthy Statements

**All statements with rank = preferred**

if there are any such statements

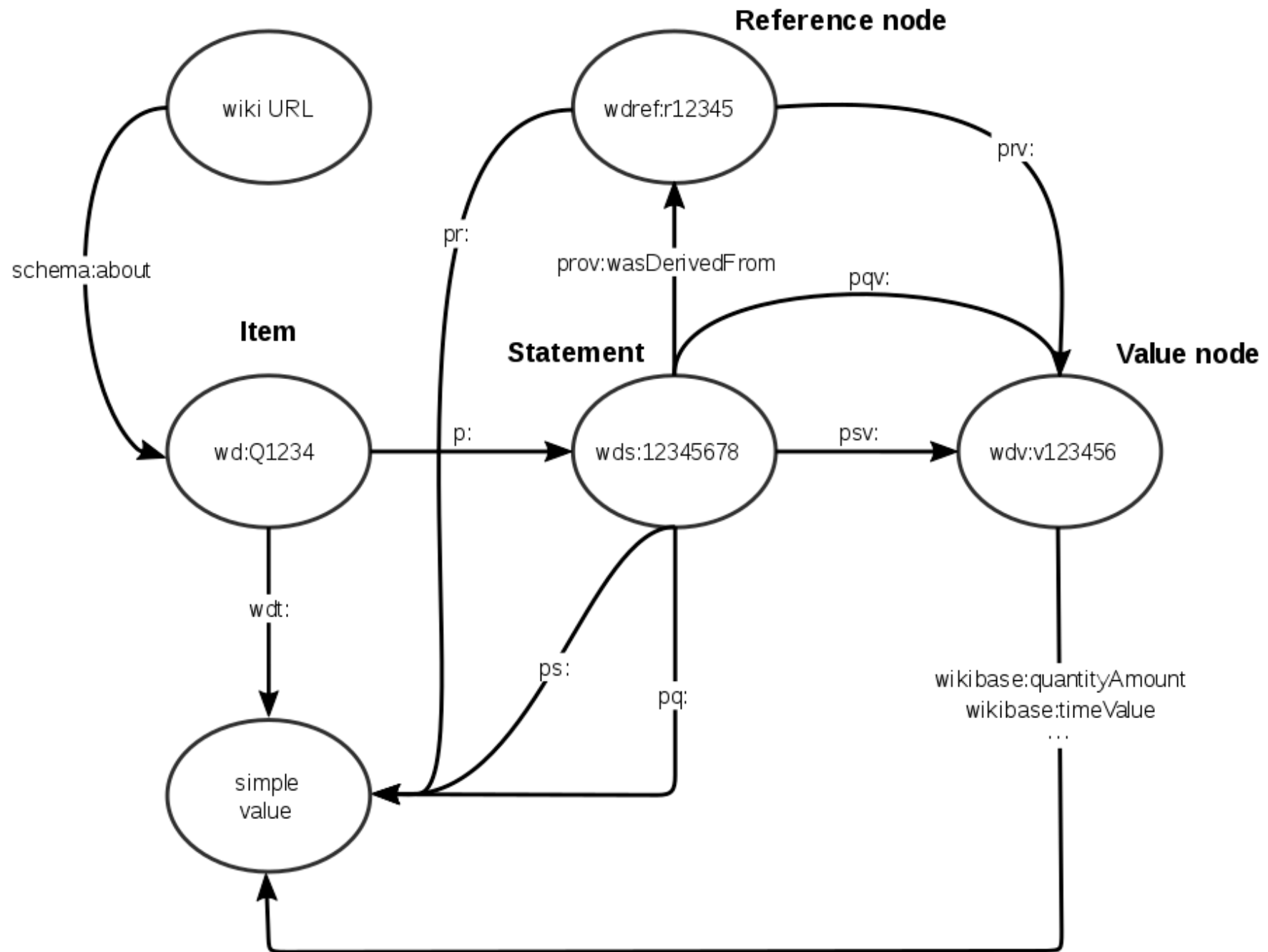
**All statements with rank = normal**

if there are no preferred statements

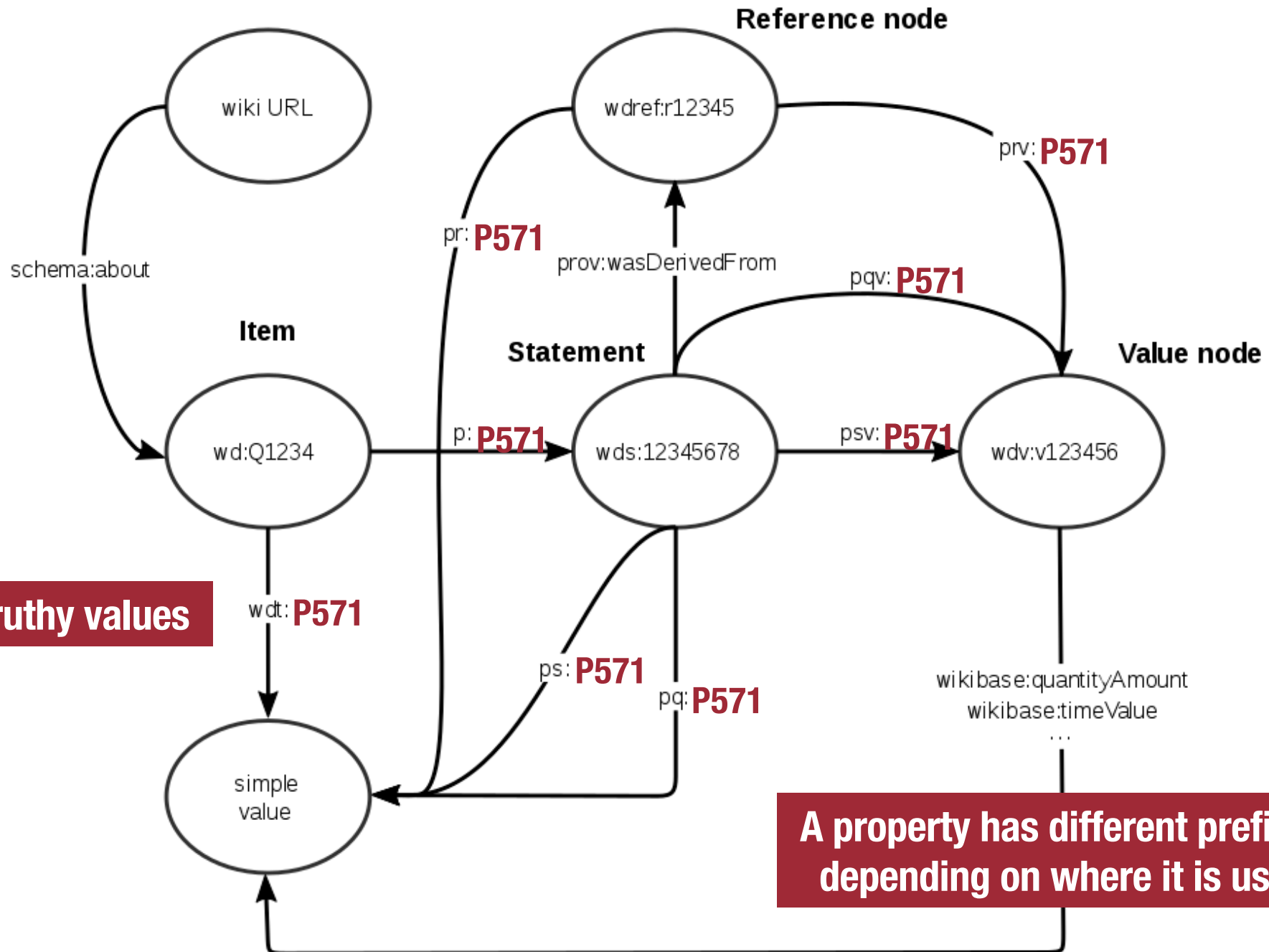
and there are statements with rank = normal

**Statements with rank = deprecated are not truthy**

# RDF Model



# RDF Model



# Wikidata RDF Format

[https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF\\_Dump\\_Format](https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format)

Item

```
wd:Q3 a wikibase:Item ;  
    rdfs:label "The Universe"@en ;  
    skos:prefLabel "The Universe"@en ;  
    schema:name "The Universe"@en ;  
    schema:description "The Universe is big"@en ;  
    skos:altLabel "everything"@en ;  
    wdt:P2 wd:Q3 ;  
    wdt:P7 "value1", "value2" ;  
    p:P2 wds:Q3-4cc1f2d1-490e-c9c7-4560-46c3cce05bb7 ;  
    p:P7 wds:Q3-24bf3704-4c5d-083a-9b59-1881f82b6b37,  
        wds:Q3-45abf5ca-4ebf-eb52-ca26-811152eb067c .
```



# RDF Representation

```
wds:Q3-24bf3704-4c5d-083a-9b59-1881f82b6b37 a
wikibase:Statement, wikibase:BestRank ;
    ps:P2 wd:Q3 ;
    wikibase:rank wikibase:PreferredRank ;
    pq:P8 "-13000000000-01-01T00:00:00Z"^^xsd:dateTime ;
    pqv:P8 wdv:382603eaa501e15688076291fc47ae54 ;
    prov:wasDerivedFrom
    wdref:87d0dc1c7847f19ac0f19be978015dfb202cf59a,
        wdref:d95dde070543a0e0115c8d5061fce6754bb82280 .
```

Statement

# RDF Representation

```
wds:Q3-24bf3704-4c5d-083a-9b59-1881f82b6b37 a
wikibase:Statement, wikibase:BestRank ;
    ps:P2 wd:Q3 ;
    wikibase:rank wikibase:PreferredRank ;
    pq:P8 "-13000000000-01-01T00:00:00Z"^^xsd:dateTime ;
    pqv:P8 wdv:382603eaa501e15688076291fc47ae54 ;
    prov:wasDerivedFrom
    wdref:87d0dc1c7847f19ac0f19be978015dfb202cf59a,
    wdref:d95dde070543a0e0115c8d5061fce6754bb82280 .
```

Statement

```
wdv:b74072c03a5ced412a336ff213d69ef1 a wikibase:TimeValue ;
    wikibase:timeValue "+1976-01-12T00:00:00Z"^^xsd:dateTime ;
    wikibase:timePrecision "11"^^xsd:integer ;
    wikibase:timeTimezone "0"^^xsd:integer ;
    wikibase:timeCalendarModel
    <http://www.wikidata.org/entity/Q1985727> .
```

Time Value

# RDF Representation

```
wds:Q3-24bf3704-4c5d-083a-9b59-1881f82b6b37 a
wikibase:Statement, wikibase:BestRank ;
    ps:P2 wd:Q3 ;
    wikibase:rank wikibase:PreferredRank ;
    pq:P8 "-13000000000-01-01T00:00:00Z"^^xsd:dateTime ;
    pqv:P8 wdv:382603eaa501e15688076291fc47ae54 ;
    prov:wasDerivedFrom
    wdref:87d0dc1c7847f19ac0f19be978015dfb202cf59a,
    wdref:d95dde070543a0e0115c8d5061fce6754bb82280 .
```

Statement

```
wdv:b74072c03a5ced412a336ff213d69ef1 a wikibase:TimeValue ;
    wikibase:timeValue "+1976-01-12T00:00:00Z"^^xsd:dateTime ;
    wikibase:timePrecision "11"^^xsd:integer ;
    wikibase:timeTimezone "0"^^xsd:integer ;
    wikibase:timeCalendarModel
    <http://www.wikidata.org/entity/Q1985727> .
```

Time Value

```
v:cb213eea7a0b90d1d7f65c6eabfab9da a wikibase:QuantityValue ;
    wikibase:quantityAmount "+123"^^xsd:decimal ;
    wikibase:quantityUpperBound "+124"^^xsd:decimal ;
    wikibase:quantityLowerBound "+122"^^xsd:decimal ;
    wikibase:quantityUnit <http://www.wikimedia.org/entity/Q199> .
```

Quantity  
Value

# Wikidata Statistics

# <https://grafana.wikimedia.org/d/0000000175/>

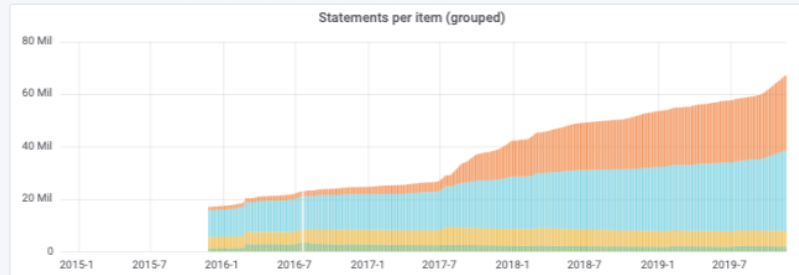
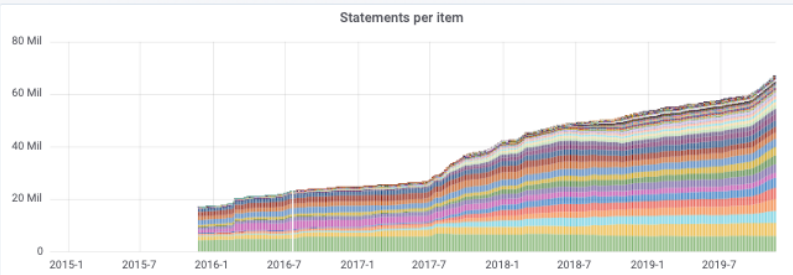
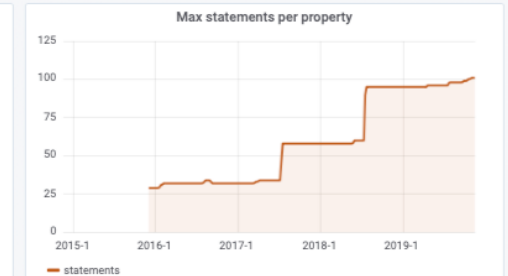
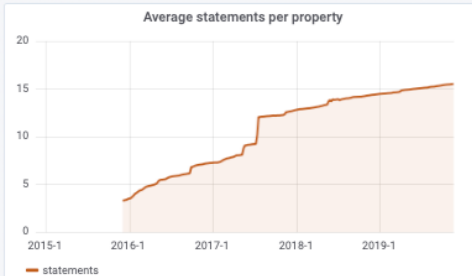
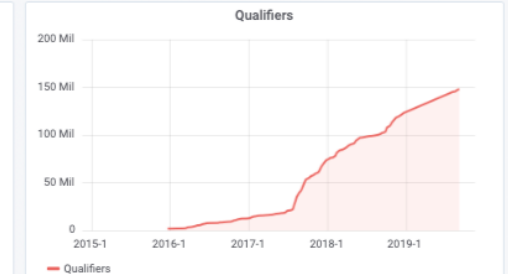
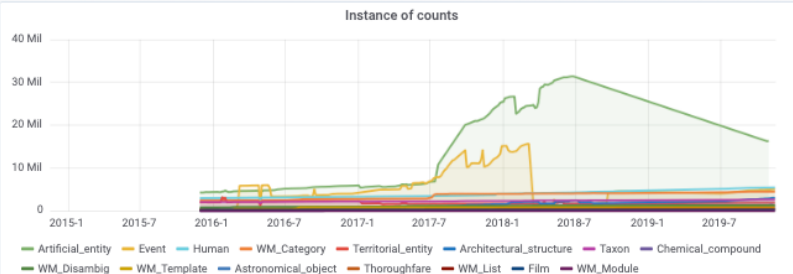
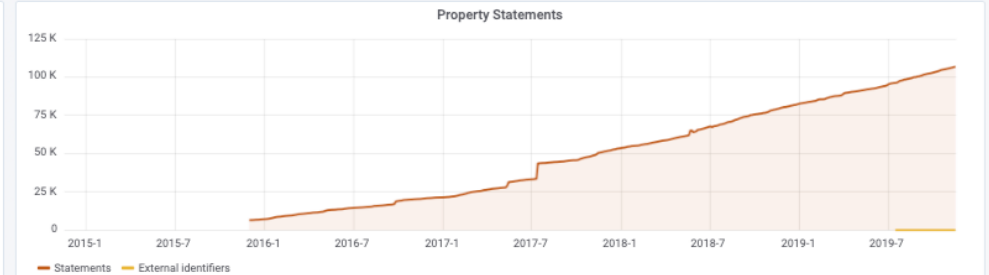
Wikidata dashboards / Wikidata Datamodel Statements

Last 5 years Refresh every 30m

Data about statements. Data is generated via a combination of the WDQS, JSON dumps and SQL queries.

NOTE: Most / All of this dashboard will NOT include LEXEMES. Only use the numbers if you know what they actually are? :D

Description



### Statements per item

Metric	Current
1	5.90 Mil
6	4.99 Mil
10	4.62 Mil
9	4.59 Mil
11	4.26 Mil
12	3.79 Mil

# Google Trends



Google Trends

Explore



● **wikidata**  
Search term

+ Compare

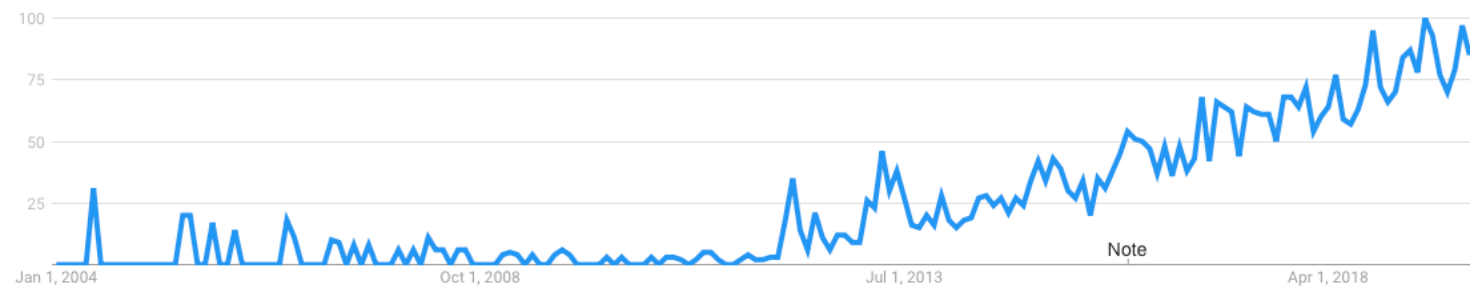
Worldwide ▾

2004 - present ▾

All categories ▾

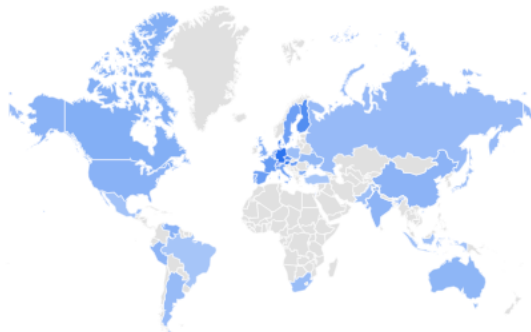
Web Search ▾

Interest over time ?



Interest by region ?

Region ▾



1	Switzerland	100	<div></div>
2	Germany	76	<div></div>
3	Austria	68	<div></div>
4	Denmark	63	<div></div>
5	Israel	61	<div></div>

### Contents

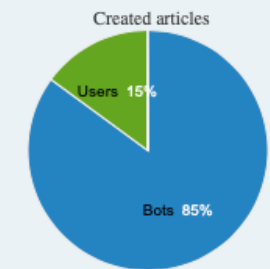
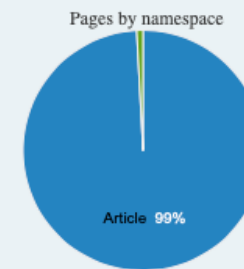
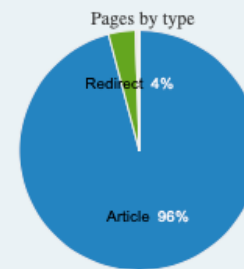
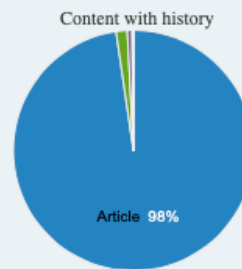
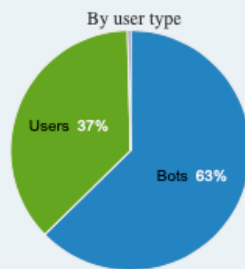
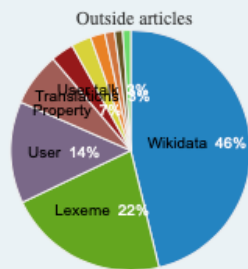
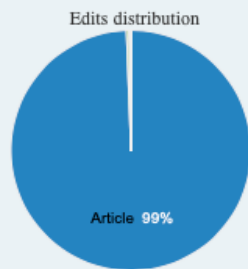
	Articles	Total
<b>Pages</b>	61,020,398	64,050,029
<b>Edits</b>	1,037,514,922	1,045,552,191
<b>Actual content</b>	1219 G	1221 G
<b>Content with history</b>	17133 G	17538 G

### Users

	Actives/month	Total
<b>Users</b>	20,242	302,022
<b>5+ edits</b>	9,322	109,938
<b>100+ edits</b>	1,937	23,298
<b>1000+ edits</b>	432	7,223

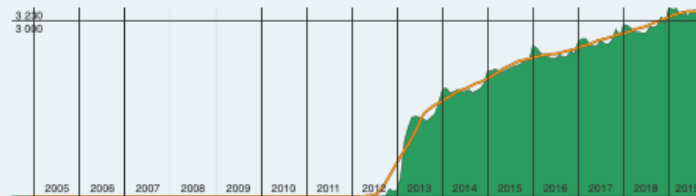
### Average by day

	Articles	All pages
<b>Edits</b>	709,087	712,682
<b>Creations</b>	45,339	46,311
<b>Content</b>	847 M	848 M



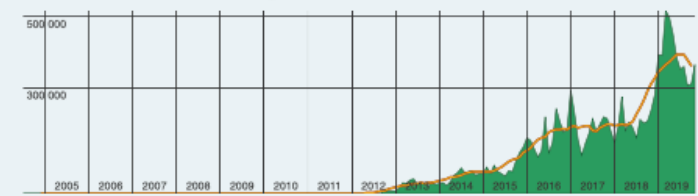
### Users / day

2019	3,171	+9.8 %
2018	2,889	+8.5 %
2017	2,662	+9.3 %
2016	2,436	+10 %
2015	2,209	+22 %
2014	1,818	+60 %
2013	1,136	>500 %



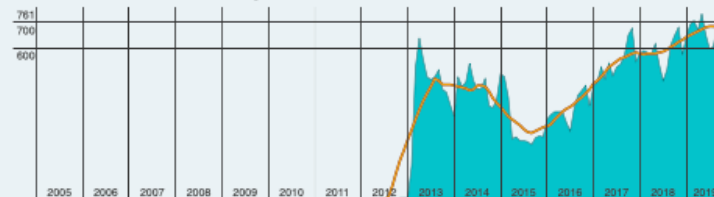
2019	392,461	+92 %
2018	204,708	+7.6 %
2017	190,323	+18 %
2016	161,304	+114 %
2015	75,398	+52 %
2014	49,651	+83 %
2013	27,119	>500 %

### User edits / day



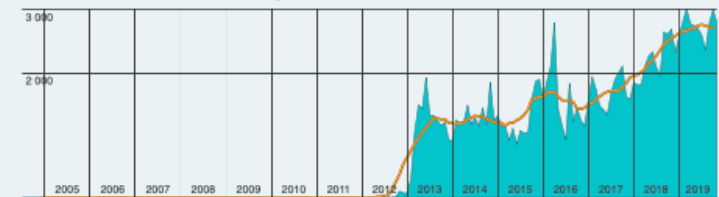
### IP / day

2019	681	+17 %
2018	582	+8.2 %
2017	538	+49 %
2016	360	+21 %
2015	297	-32 %
2014	437	+3.4 %
2013	423	>500 %



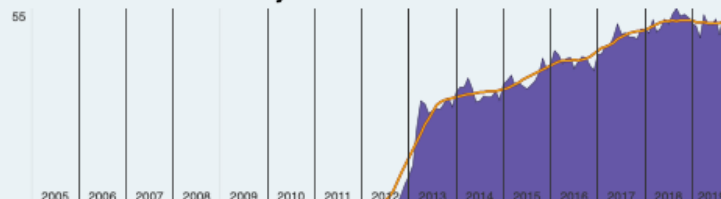
2019	2,744	+25 %
2018	2,205	+30 %
2017	1,691	+8.4 %
2016	1,559	+25 %
2015	1,246	-2.3 %
2014	1,275	+14 %
2013	1,119	>500 %

### IP edits / day



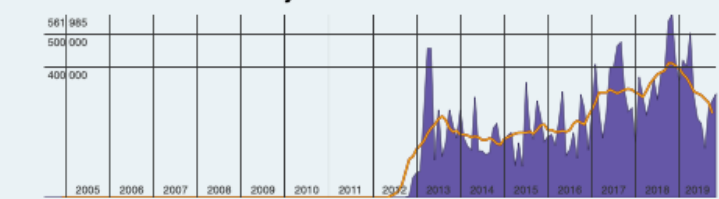
### Bots / day

2019	51	-1.1 %
2018	51	+11 %
2017	46	+15 %
2016	40	+15 %
2015	35	+12 %
2014	31	+29 %
2013	24	>500 %



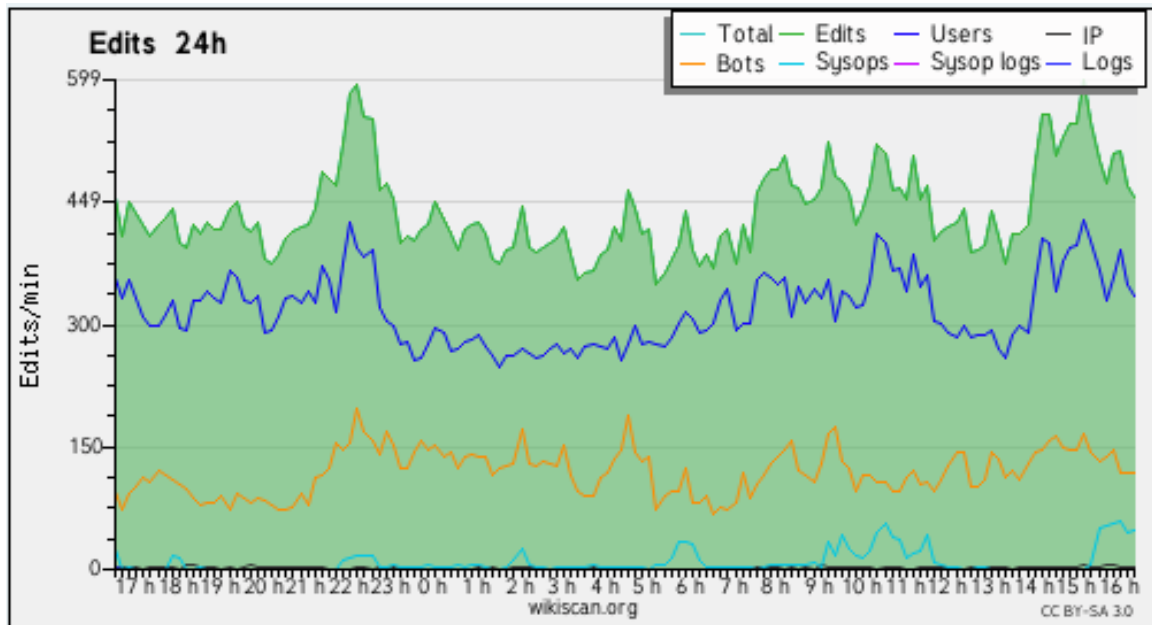
2019	319,103	-12 %
2018	363,948	+9.5 %
2017	332,476	+64 %
2016	203,112	+1.7 %
2015	199,775	+7.9 %
2014	185,117	-18 %
2013	224,259	>500 %

### Bot edits / day

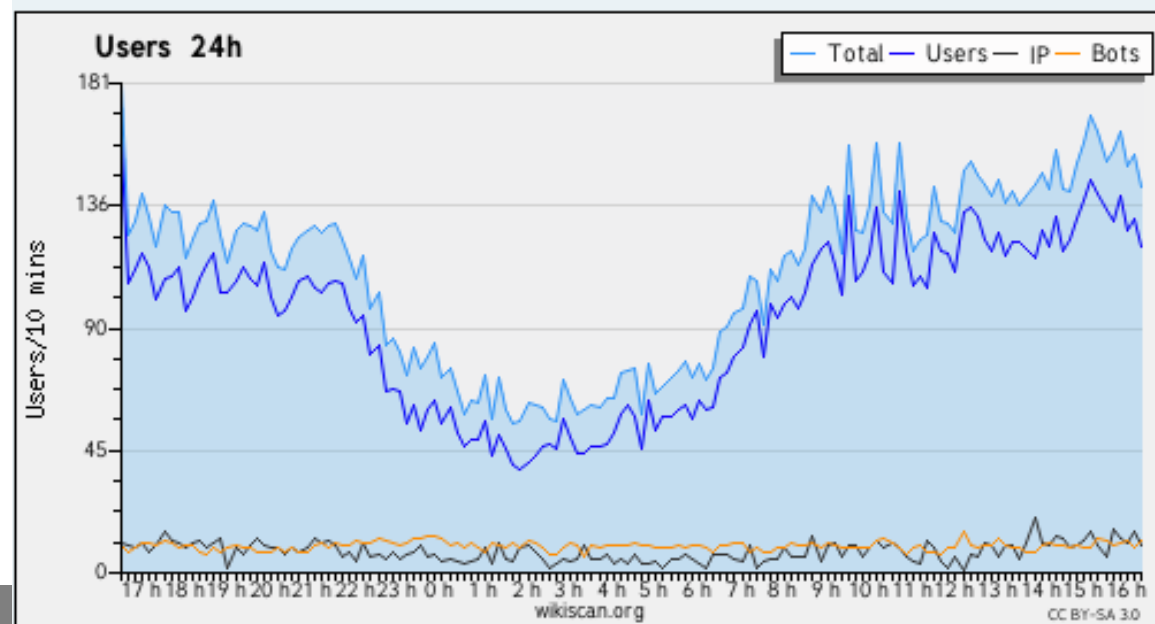




# <https://wikidata.wikiscan.org/hours>



**~10 edits/second**





# Wikidata:Status updates/2019 11 11

[< Wikidata:Status updates](#)

This is the Wikidata summary of the week before 2019-11-11.

## Contents [\[hide\]](#)

- [1 Discussions](#)
- [2 Events](#)
- [3 Tool of the week](#)
- [4 Other Noteworthy Stuff](#)
- [5 Did you know?](#)
- [6 Development](#)
- [7 Monthly Tasks](#)

[https://www.wikidata.org/wiki/Wikidata:Status\\_updates](https://www.wikidata.org/wiki/Wikidata:Status_updates)

## Discussions [\[ edit | edit source \]](#)

- New request for comments: [Wikidata to use data schemas to standardise data structure on a subject](#)

## Events [\[ edit | edit source \]](#)

- Past: Wikidata office hour ([notes](#))
- Past: "Using Wikidata to describe the structure of a book" Webinar by Martin Poulter for LD4P2 project ([Video](#)), ([Slides](#)), ([Notes](#))
- Upcoming: [Wikidata Workshop OpenRefine](#) in Prague on 12 November
- Upcoming: [Online editathon in Swedish](#) 17 November

## Tool of the week [\[ edit | edit source \]](#)

- [Speedpatrolling](#) is a tool to easily patrol Wikidata recent changes, typically on mobile devices.

## Other Noteworthy Stuff [\[ edit | edit source \]](#)

- Wikipedia Personality Infobox in [Catalan](#) and [French](#) now displays the Heritage Institution in charge of their archives via [P485 \(archives at\)](#).
- The [Discord server Wikimedia Community](#) has now a [Wikidata](#) channel.
- [@Wikimedia talks about @Wikidata](#) by GerardM

## Did you know? [\[ edit | edit source \]](#)

- Newest [properties](#):
  - General datatypes: [landscape architect](#), [charge conjugation quantum number](#), [statement is regarded as spoiler for](#), [room number](#), [scope and content](#)
  - External identifiers: [startrek.com Database ID](#), [Alfred P. Sloan Foundation Grant ID](#), [Chinese Professional Baseball League player ID](#), [Software Preservation Society ID](#), [Sega Retro ID](#), [classicamiga ID](#), [Filmový přehled film ID](#), [Filmový přehled person ID](#), [Gaming-History company ID](#), [SK cinema authority ID](#), [SK cinema film ID](#), [CA PROP 65 ID](#), [ePSD identifier](#), [Czech cadastral area ID](#), [dp.ru company ID](#), [Fortuna liga player ID](#), [MEA Indian Mission ID](#), [AaRC person ID](#), [MIC market code](#)
- New [property proposals](#) to review:

# Keeping Up With New Developments

[https://www.wikidata.org/wiki/Wikidata:Development\\_plan](https://www.wikidata.org/wiki/Wikidata:Development_plan)



Main page  
Community portal  
Project chat  
Create a new Item  
Create a new Lexeme  
Recent changes  
Random Item  
Query Service  
Nearby  
Help  
Donate

Print/export

Create a book  
Download as PDF  
Printable version

Tools

What links here  
Related changes  
Special pages  
Permanent link  
Page information

In Wikipedia

Add links

English Szeke Talk Preferences Beta Watchlist Contributions Log out

Project page [Discussion](#)

Read

[Edit](#)

[Edit source](#)

[View history](#)



Search Wikidata



Shortcuts: [WD:DP](#), [WD:DEVPLAN](#)

## Wikidata:Development plan

[Development plan](#)

[Usability and usefulness](#)

[Status updates](#)

[Development input](#)

[Contact the development team](#)

### Wikidata and Wikibase roadmap for 2019 [\[ edit \]](#) [\[ edit source \]](#)

You will find below the roadmap of the Wikidata development team (Wikimedia Germany) for Wikidata and [Wikibase](#) for 2019. If you have any question or comment, feel free to write on [Wikidata:Contact the development team](#).

Please note that the roadmap presents the main projects that the development team will work on during 2019. Things like maintenance of the software and fixing pressing bugs are not mentioned in the roadmap, but will be included in the workflow over the year.

### Overview of the roadmap [\[ edit \]](#) [\[ edit source \]](#)

The most up-to-date version of the roadmap can be found on an online project management tool: [view in blocks](#), [view in timeline](#). You can click on the different items to see further information, like a description or planned start date.

You can also find screenshots below:

### WIKIDATA ROADMAP - ROUGH OVERVIEW 2019

INCREASE DATA QUALITY & TRUST	BUILD OUT THE ECOSYSTEM	ENCOURAGE MORE DATA USE	ENABLE MORE DIVERSE DATA & USERS	OTHER
<a href="#">basic ShEx support</a>	<a href="#">federation (ontology reuse)</a>	<a href="#">editing Wikidata from Wikipedia and co (client editing)</a>	<a href="#">better mobile web support (viewing terms)</a>	<a href="#">WikidataCon</a>
<a href="#">signed statements (signing statements not linking to other entities)</a>	<a href="#">Wikibase website</a>	<a href="#">query builder</a>	<a href="#">better mobile web support (editing terms)</a>	<a href="#">storage infrastructure improvements</a>
<a href="#">improved reference support (Citoid)</a>	<a href="#">develop a partnership model around Wikibase</a>			
<a href="#">automated adding of references based on semantic markup</a>	<a href="#">define the landscape and segments of Wikibase users</a>			

WIKIDATA ROADMAP - ROADMAP OF EVERYONE

Jan 1, 2019 - Dec 31, 2019

# Wikidata Summary

**Free**

**Collaborative**

**Multilingual**

**A secondary database**

**Collecting structured data**

**Support for Wikimedia wikis**

**Anyone in the world**

**> 60M items**

**> 6,000 properties**

**> 2,000,000 classes**

**> 4,000 external id types**

**Model using statements**

**Subject, property, value**

**Qualifiers, references**