Wikibase and Wikidata

Wikibase is…

- Wikibase: a structured data repository based on MediaWiki
- Complex/expressive data model has triples, provenance, qualifiers, and alternate values
- Export to standard formats including JSON, RDF/XML, N3, and Turtle
- Access via SPARQL
- Local installs via a Docker image
- Stored in a RDBMS (e.g., mySQL)

Wikidata is…

- Wikidata is “the free knowledge base with 69,192,605 data items that anyone can edit”
- Uses the Wikibase data model and associated software and APIs
- Its data is available to download
  - In bulk as JSON or RDF
  - as individual items in JSON or RDF

https://www.wikidata.org/wiki/Q5

wikidata.org

English | Not logged in | Talk | Contributions | Create account | Log in

WIKIDATA

Item | Discussion | Read | View history | Search Wikidata

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
Concept URI
Cite this page

Douglas Adams (Q42)

English writer and humorist

Douglas Noel Adams | Douglas Noël Adams | Douglas N. Adams

In more languages

Configure

Language	Label	Description	Also known as
English	Douglas Adams	English writer and humorist	Douglas Noel Adams Douglas Noël Adams Douglas N. Adams
Spanish	Douglas Adams	escritor y humorista británico	Douglas Noel Adams Douglas Noël Adams
Traditional Chinese	道格拉斯·亞當斯	英國作家	
Chinese	道格拉斯·亚当斯	英国作家	亞當斯

All entered languages

Statements

instance of	human 2 references
image	

https://wikidata.org/wiki/Special:EntityData/Q42.json

wikidata.org


```
{"entities":{"Q42":{"pageid":138,"ns":0,"title":"Q42","lastrevid":1065557227,"modified":"2019-
11-29T18:46:12Z","type":"item","id":"Q42","labels":{"fr":{"language":"fr","value":"Douglas
Adams"},"ru":{"language":"ru","value":"\u0414\u0443\u0433\u043b\u0430\u0441
\u0410\u0434\u0430\u043c\u0441"},"pl":{"language":"pl","value":"Douglas Adams"},"it":
{"language":"it","value":"Douglas Adams"},"en-gb":{"language":"en-gb","value":"Douglas
Adams"},"nb":{"language":"nb","value":"Douglas Adams"},"es":{"language":"es","value":"Douglas
Adams"},"en-ca":{"language":"en-ca","value":"Douglas Adams"},"hr":
{"language":"hr","value":"Douglas Adams"},"pt":{"language":"pt","value":"Douglas Adams"},"ko":
{"language":"ko","value":"\ub354\uae00\ub7ec\uc2a4 \uc560\ub364\uc2a4"},"nl":
{"language":"nl","value":"Douglas Adams"},"el":
{"language":"el","value":"\u039d\u03c4\u03ac\u03b3\u03ba\u03bb\u03b1\u03c2
\u0386\u03bd\u03c4\u03b1\u03bc\u03c2"},"ar":
{"language":"ar","value":"\u062f\u0648\u063a\u0644\u0627\u0633
\u0622\u062f\u0645\u0632"},"arz":
{"language":"arz","value":"\u062f\u0648\u062c\u0644\u0627\u0633
\u0627\u062f\u0627\u0645\u0632"},"bar":{"language":"bar","value":"Douglas Adams"},"be":
{"language":"be","value":"\u0414\u0443\u0433\u043b\u0430\u0441
\u0410\u0434\u0430\u043c\u0441"},"bg":
{"language":"bg","value":"\u0414\u044a\u0433\u043b\u0430\u0441
\u0410\u0434\u0430\u043c\u0441"},"bs":{"language":"bs","value":"Douglas Adams"},"ca":
{"language":"ca","value":"Douglas Adams"},"cs":{"language":"cs","value":"Douglas Adams"},"cy":
{"language":"cy","value":"Douglas Adams"},"da":{"language":"da","value":"Douglas Adams"},"eo":
{"language":"eo","value":"Douglas Adams"},"et":{"language":"et","value":"Douglas Adams"},"fa":
{"language":"fa","value":"\u062f\u0627\u06af\u0644\u0627\u0633
\u0622\u062f\u0627\u0645\u0632"},"fi":{"language":"fi","value":"Douglas Adams"},"ga":
{"language":"ga","value":"Douglas Adams"},"gl":{"language":"gl","value":"Douglas Adams"},"he":
{"language":"he","value":"\u05d3\u05d0\u05d2\u05dc\u05e1 \u05d0\u05d3\u05de\u05e1"},"hu":
{"language":"hu","value":"Douglas Adams"},"id":{"language":"id","value":"Douglas Adams"},"io":
{"language":"io","value":"Douglas Adams"},"is":{"language":"is","value":"Douglas Adams"},"ja":
{"language":"ja","value":"\u30c0\u30b0\u30e9\u30b9\u30fb\u30a2\u30c0\u30e0\u30ba"},"jv":
{"language":"jv","value":"Douglas Adams"},"ka":
{"language":"ka","value":"\u10d3\u10d0\u10d2\u10da\u10d0\u10e1
\u10d0\u10d3\u10d0\u10db\u10e1\u10d8"},"la":{"language":"la","value":"Duglassius Adams"},"lv":
{"language":"lv","value":"Duglass Adamss"},"mk":
{"language":"mk","value":"\u0414\u0430\u0433\u043b\u0430\u0441
```

The entity in JSON

Q42.json — ~/Downloads

Q42.json


```
1     {
2       "entities": {
3         "Q42": {
4           "pageid": 138,
5           "ns": 0,
6           "title": "Q42",
7           "lastrevid": 1065557227,
8           "modified": "2019-11-29T18:46:12Z",
9           "type": "item",
10          "id": "Q42",
11 ›        "labels": {…},
637 ›       "descriptions": {…},
967 ›       "aliases": {…},
1325 ›      "claims": {…},
10236 ›     "sitelinks": {…}
10916     }
10917   }
10918 }
```

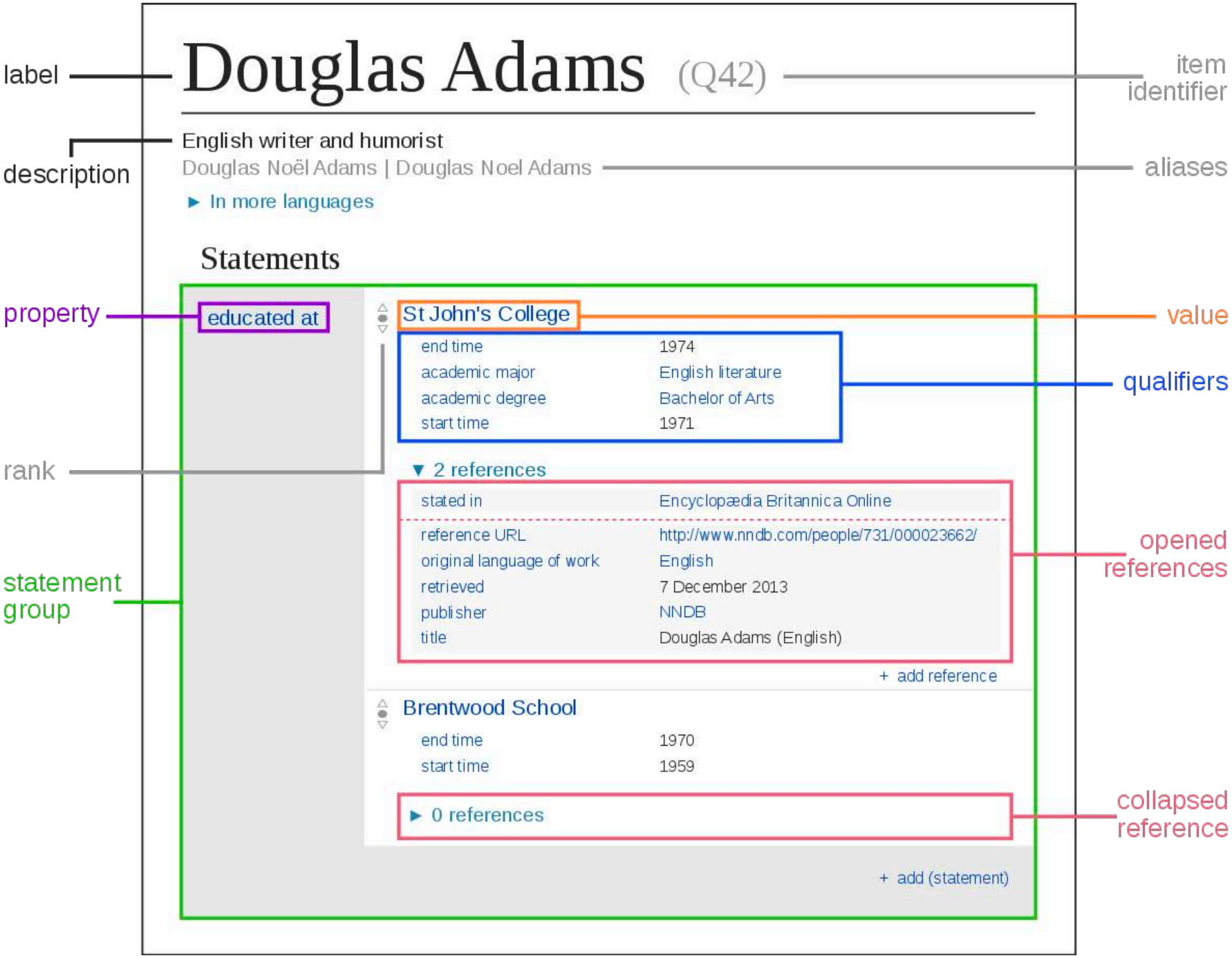

~/Downloads/Q42.json* 11:7 1 LF UTF-8 JSON GitHub Git (0) 1 update

https://wikidata.org/wiki/Special:EntityData/Q42.ttl

wikidata.org

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix ontolex: <http://www.w3.org/ns/lemon/ontolex#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix wikibase: <http://wikiba.se/ontology#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix schema: <http://schema.org/> .
@prefix cc: <http://creativecommons.org/ns#> .
@prefix geo: <http://www.opengis.net/ont/geosparql#> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix v: <http://www.wikidata.org/value/> .
@prefix wd: <http://www.wikidata.org/entity/> .
@prefix data: <https://www.wikidata.org/wiki/Special:EntityData/> .
@prefix s: <http://www.wikidata.org/entity/statement/> .
@prefix ref: <http://www.wikidata.org/reference/> .
@prefix wdt: <http://www.wikidata.org/prop/direct/> .
@prefix wdtn: <http://www.wikidata.org/prop/direct-normalized/> .
@prefix p: <http://www.wikidata.org/prop/> .
@prefix ps: <http://www.wikidata.org/prop/statement/> .
@prefix psv: <http://www.wikidata.org/prop/statement/value/> .
@prefix psn: <http://www.wikidata.org/prop/statement/value-normalized/> .
@prefix pq: <http://www.wikidata.org/prop/qualifier/> .
@prefix pqv: <http://www.wikidata.org/prop/qualifier/value/> .
@prefix pqn: <http://www.wikidata.org/prop/qualifier/value-normalized/> .
@prefix pr: <http://www.wikidata.org/prop/reference/> .
@prefix prv: <http://www.wikidata.org/prop/reference/value/> .
@prefix prn: <http://www.wikidata.org/prop/reference/value-normalized/> .
@prefix wdno: <http://www.wikidata.org/prop/novalue/> .

data:Q42 a schema:Dataset ;
        schema:about wd:Q42 ;
        cc:license <http://creativecommons.org/publicdomain/zero/1.0/> ;
        schema:softwareVersion "1.0.0" ;
```


label
Douglas Adams
(Q42)
item identifier
description
English writer and humorist
Douglas Noël Adams | Douglas Noel Adams
aliases
In more languages
Statements
property
educated at
St John's College
value
end time 1974
academic major English literature
academic degree Bachelor of Arts
start time 1971
qualifiers
rank
2 references
stated in Encyclopædia Britannica Online
reference URL http://www.nndb.com/people/731/000023662/
original language of work English
retrieved 7 December 2013
publisher NNDB
title Douglas Adams (English)
opened references
statement group
add reference
Brentwood School
end time 1970
start time 1959
0 references
collapsed reference
add (statement)

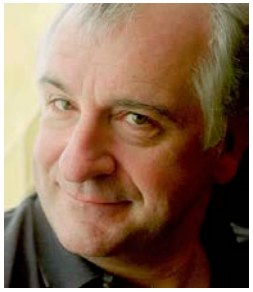
Wikibase Data Model

- **Item** = subjects = entities
- **Property** = properties
- **Value** = entities or datatypes (string, number,…)
- **Snak** = basic assertion about item, i.e. a Property-Value pair -- ”small, but more than a byte”
  - Some are simple claims: *population of Berlin is 3,499,879*
  - Others (e.g., type assertions) are structural: *type Berlin City*
  - Others include a claim an qualifiers
    Population of Berlin is 3,499,879, considering only territory of city, as estimated on 30 November 2011

Items have

- **Item identifier** (number prefixed with *Q*)
- **Fingerprint**, consisting of:
  - Multilingual **label***
  - Multilingual **description***
  - Multilingual **aliases**
- **Statements**, each consisting of:
  - **Claim**, consisting of:
    - Property
    - Value
    - Qualifiers (additional property-value pairs)
  - **References** (each with one or more property-value pairs)
  - **Rank**
- **Site links**

Statements…

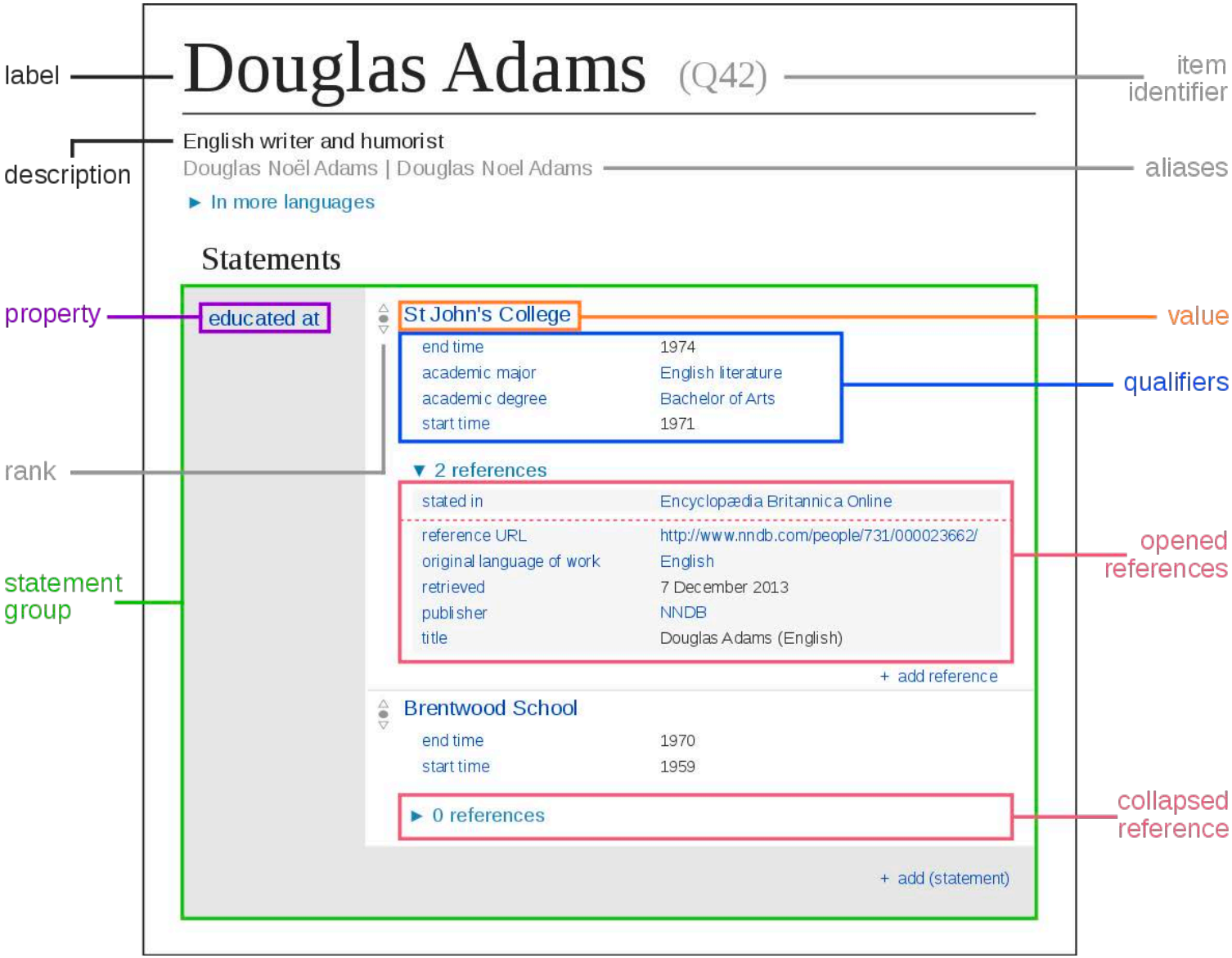


- A statement may have:
  - one property (in the example, P551 “residence")
  - one value (Q84 “London”)
  - optionally one or more qualifiers (e.g, property:P582, “end time” 11 May 2011)
  - optional reference(s) (e.g., property:P143 “imported from Wikipedia”)
- The property, value, and qualifiers together are also called the **claim**, which together with any source references forms a statement.

Properties have …

- **Property identifier** (number prefixed with *P*)
- **Fingerprint**, consisting of:
  - Multilingual **label***
  - Multilingual **description***
  - Multilingual **aliases**
- **Statements**, each consisting of:
  - **Claim**, consisting of:
    - Property
    - Value
    - Qualifiers (additional property-value pairs)
  - **References** (each with one or more property-value pairs)
  - **Rank**
- **Datatype**

Example of Data Model



Statements…

- Requirement: "Wikibase will not be about the truth, but about statements and their references"
- Doesn’t model items, but statements about them
- Not “Daulgas Adams residence is London”
- But “There’s a statement of Douglas Adams having a residence of London prior to 11 May 2011 according to WIkipedia”

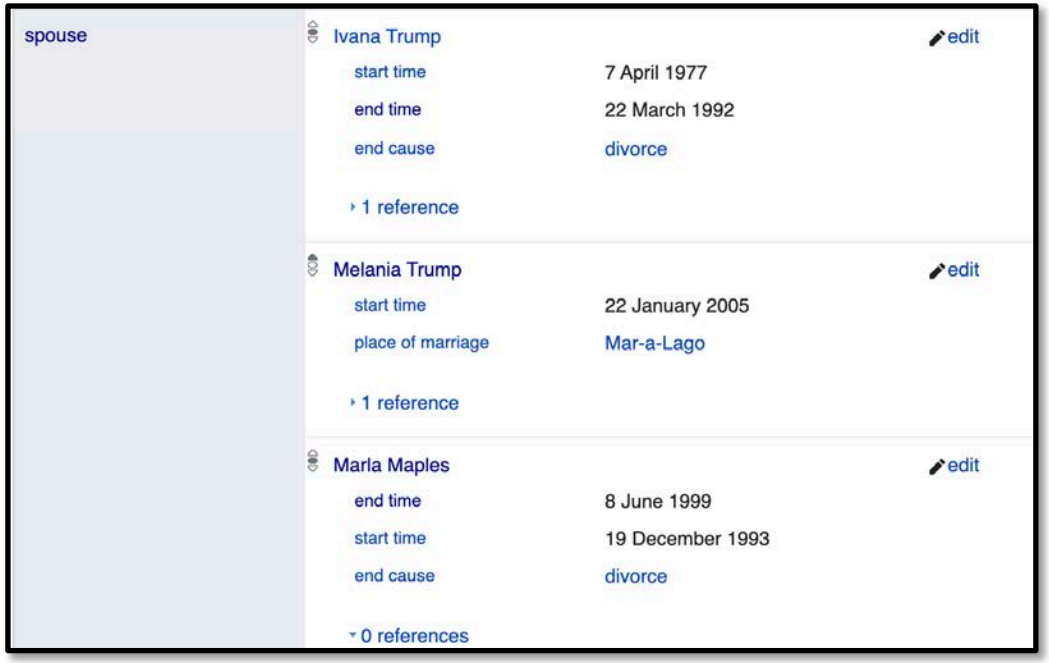
Example: Trumps spouses

- Who are Donald Trump's spouses?
- We must identify the IDs for
  - Donald Trump
  - Spouse relation
- And then write and run a simple SPARQL query

Let's give it a try

Well….

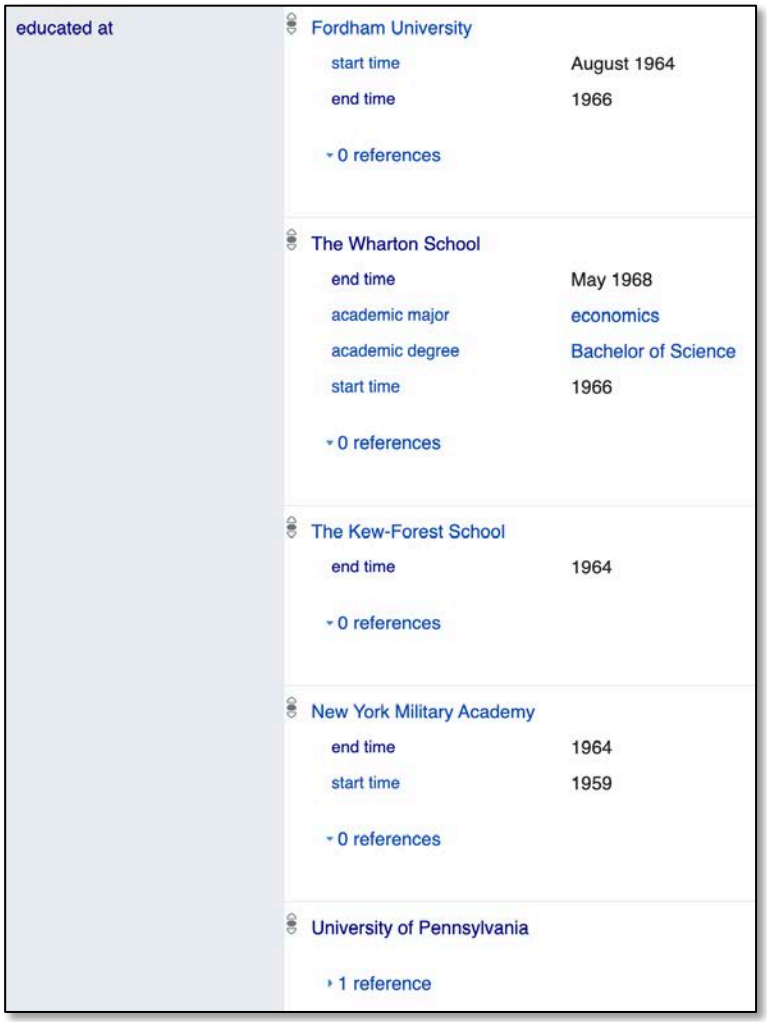
- It only returns one answer: his *current* spouse
- Other values have an end time
- Maybe that's a feature!



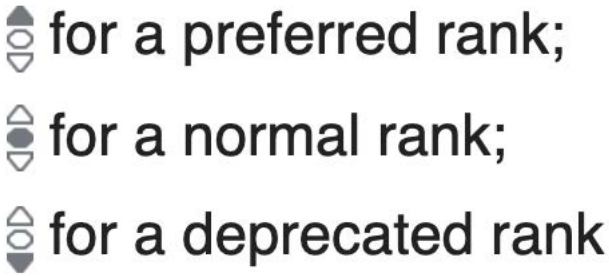
- Let's try another query: what schools did Donald Trump attend?

Property Rank

- We get four schools, even though all have end dates
  (we might quibble that Penn and Wharton are the same)
- Does Wikidata's ontology know that *spouse* (P26) is a temporal quality and *educated at* (P69) is not?
- No, though property has some constraints that might be useful
- The mechanism used is to give each value a rank



Ranking claims



- **Preferred:** most current or represent consensus
- **Normal:** default; no judgement of a value's accuracy and currency
- **Deprecated:** errors or outdated

For DT’s spouses, Melania has preferred rank and the others normal rank

All of DT’s schools had normal rank.

How are ranks represented in RDF and how does the Wikidata query service use them?

WDQS Procedure

What's matched for **?s wdt:Pxxx ?o**

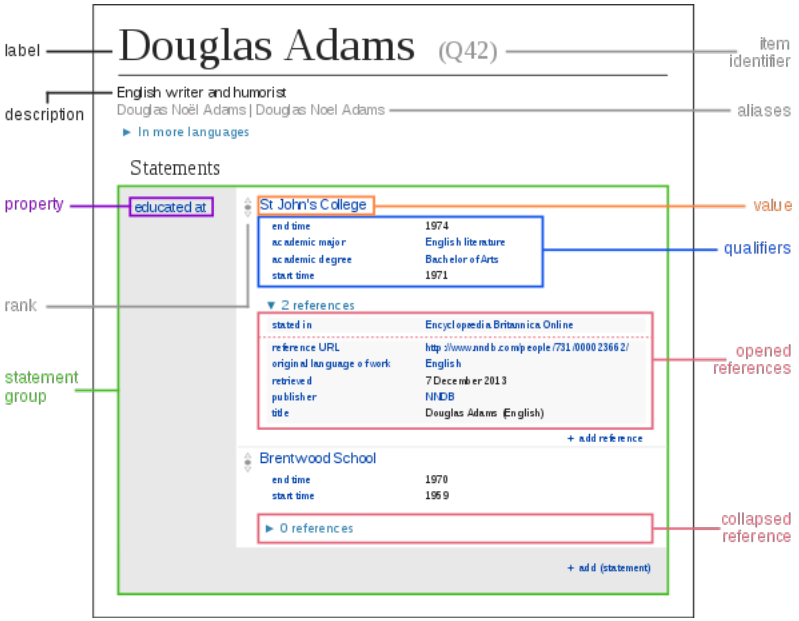
- If there's at least one ?v with preferred rank, only values preferred values are returned
- If there are no preferred values, all values with normal rank are returned
- Deprecated values are never returned.

The humans or bots populating the graph must figure out how to assign ranks

Qualifiers, rank and references

Wikidata uses special namespaces to access a reified node with claim's qualifiers, rank & references

- **prefix p:** points not to object, but to statement node
- It is then subject of other triples
- Within a statement node:
  - **ps:** gets the object
  - **pq:** gets qualifier information
  - **wikibase:rank** gets rank information
  - **prov:wasDerivedFrom/pr:P248** gets reference values



Example (1)

```
SELECT ?education ?educationLabel ?starttime ?endtime  WHERE {
  wd:Q42 p:P69 ?statement.
  ?statement ps:P69 ?education.
  ?statement pq:P580 ?starttime.
  ?statement pq:P582 ?endtime.
SERVICE wikibase:label { bd:serviceParam wikibase:language "en". } }
ORDER BY ?starttime
```

Try it

Example (2)

We can simplify this with the [ ] syntax to eliminate the ?statement variable.

```
SELECT ?education ?educationLabel ?starttime ?endtime  WHERE {
  wd:Q42 p:P69
    [ ps:P69 ?education;
      pq:P580 ?starttime;
      pq:P582 ?endtime ].
SERVICE wikibase:label { bd:serviceParam wikibase:language "en". } }
ORDER BY ?starttime
```

Try it

Example (3)

Here’s an example getting rank information

```
SELECT ?education ?educationLabel ?rank WHERE {
  wd:Q42 p:P69
    [ps:P69 ?education;
     wikibase:rank ?rank; ].
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en". } }
```

Try it

Trumps Spouses

```
# Get Donald Trump’s spouses, current and former and deprecated
SELECT ?spouse ?spouseLabel ?rank
WHERE {
 wd:Q22686 p:P26
   [ps:P26 ?spouse;
    wikibase:rank ?rank; ].
 SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}
```

Try it

Deprecated values

- See this page on deprecation
- Honoré de Balzac (Q9711) has two values for date of death (P570): 18 and 19 August 1850
- The August 19 claim is tagged as deprecated, with the reason incorrect value (Q41755623)

Getting the software and data

- Open source software to run an instance
  - Uses a RDBMS (e.g., mysql) for storage
  - Provides a SPARQL interface
- Data dumps in JSON or RDF
  - 33GB for JSON (compressed)
  - 43GB for TTL (compressed)