

Wikidata on MARS, or turning Wiki-data into Wiki-knowledge



Peter F. Patel-Schneider

PARC, Palo Alto, CA, USA pfps@parc.com

Abstract

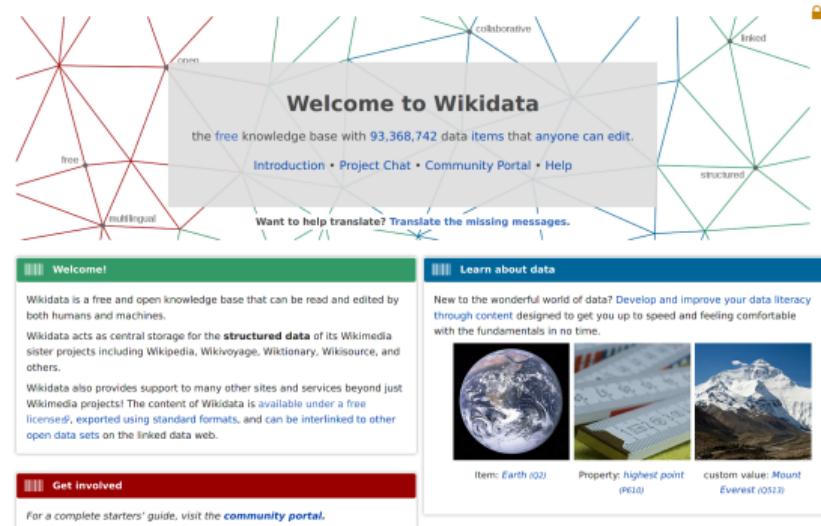
Users (particularly programs) need to know far too much to reliably get the information they need from Wikidata. So Wikidata matches its name in that it contains data, not knowledge. An underlying formal basis can be provided for Wikidata, turning Wiki-data into Wiki-knowledge. Getting these formal bases used, so that Wikidata and other Knowledge Graphs contain knowledge, is a separate matter.

Purpose of This Talk

- Show that there is more to knowledge than just Knowledge Graphs
- Show a way of turning Knowledge Graphs into knowledge
- I'm using Wikidata as my example Knowledge Graph.
 - Wikidata is useful, available, and examinable.
 - Although Wikidata has differences from other Knowledge Graphs, it has much more in common with them, particularly when support for knowledge is considered.
- I'm being a bit harsh on Wikidata.
 - Wikidata is a community effort, which makes organization hard.
 - Wikidata is improving, but not in the way I think it should.

What is Wikidata

- Large community-driven repository
 - Over 93 million entities
 - Anyone can update
 - Anyone can use freely



What is Wikidata

- Large community-driven repository
 - Over 93 million entities
 - Anyone can update
 - Anyone can use freely
- Display interface
Elizabeth Taylor

Elizabeth Taylor (Q34851)

British-American actress, businesswoman, and humanitarian

Elizabeth Rosemond Taylor | Liz Taylor | Dame Elizabeth Rosemond Taylor | Dame Elizabeth Taylor

[edit](#)

» [Recoin: Most relevant properties which are absent](#)

▼ In more languages

Language	Label	Description	Also known as
English	Elizabeth Taylor	British-American actress, businesswoman, and humanitarian	Elizabeth Rosemond Taylor Liz Taylor Dame Elizabeth Rosemond... Dame Elizabeth Taylor
German	Elizabeth Taylor	britisch-US-amerikanische Schauspielerin	Dame Elizabeth Rosemond... Elizabeth Rosemond Taylor Liz Taylor
French	Elizabeth Taylor	actrice britannique	Elizabeth Rosemond Taylor Elizabeth Taylor Liz Taylor

[All entered languages](#)

Statements

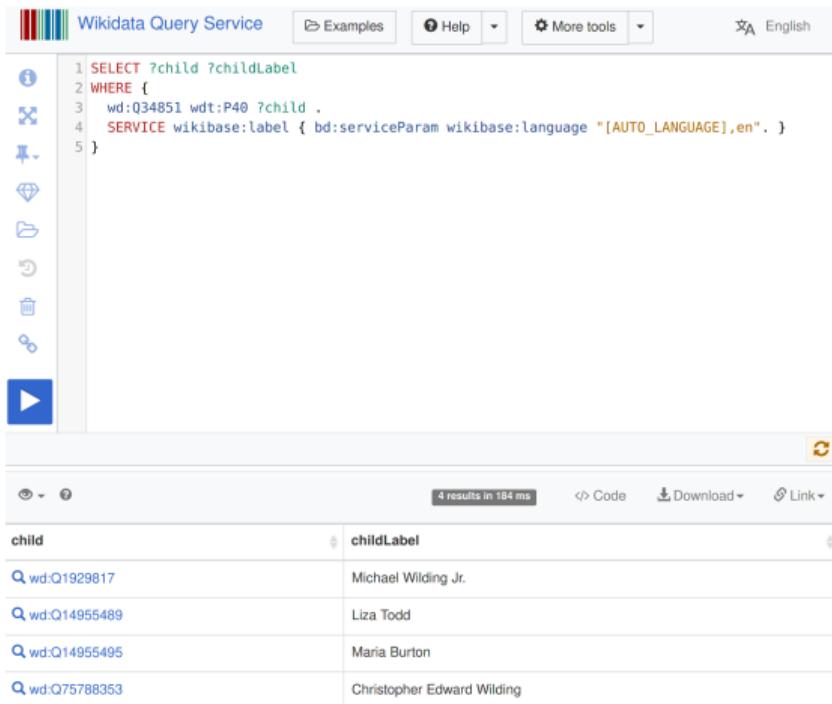
instance of	 human	edit
» 3 references		

[+ add value](#)

image		edit
-------	-------------------------------------------------------------------------------------	----------------------

What is Wikidata

- Large community-driven repository
 - Over 93 million entities
 - Anyone can update
 - Anyone can use freely
- Display interface
- SPARQL query interface
children of Elizabeth Taylor
 - (Annoying to use because of numeric identifiers)



The screenshot shows the Wikidata Query Service interface. The top navigation bar includes links for Examples, Help, More tools, and a language selector set to English. The main area is a code editor containing a SPARQL query:

```
1 SELECT ?child ?childLabel
2 WHERE {
3   wd:Q34851 wdt:P40 ?child .
4   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en" . }
5 }
```

Below the code editor is a large blue play button. The results table shows four rows of data:

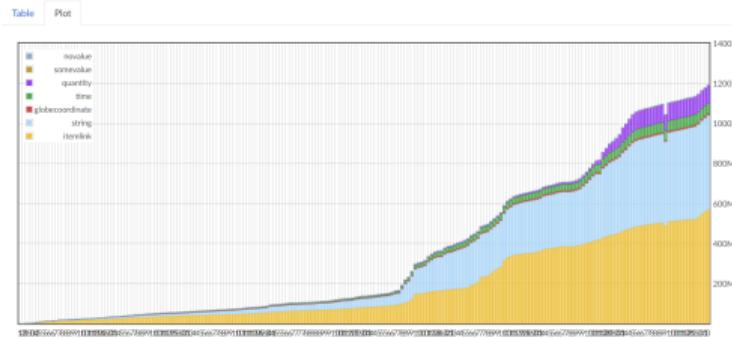
child	childLabel
Q wd:Q1929817	Michael Wilding Jr.
Q wd:Q14955489	Liza Todd
Q wd:Q14955495	Maria Burton
Q wd:Q75788353	Christopher Edward Wilding

At the bottom of the results table, there are links for Code, Download, and Link.

Is Wikidata Useful?

- Absolutely
 - Large
1.3 billion “facts”

Referenced statements by statement type



Is Wikidata Useful?

- Absolutely
 - Large
 - High Quality

both adopted and natural children of Elizabeth Taylor

child	<p> Maria Burton kinship to subject </p> <p> adopted daughter</p> <p> + 1 reference</p>
	<p> Michael Wilding Jr. </p> <p> + 1 reference</p>
	<p> Christopher Edward Wilding </p> <p> + 1 reference</p>
	<p> Liza Todd </p> <p> + 1 reference</p>
	<p> + add value</p>
number of children	<p> 4 </p> <p> + 1 reference</p>
	<p> + add value</p>

Is Wikidata Useful?

■ Absolutely

- Large
- High Quality
- Connections to other sources

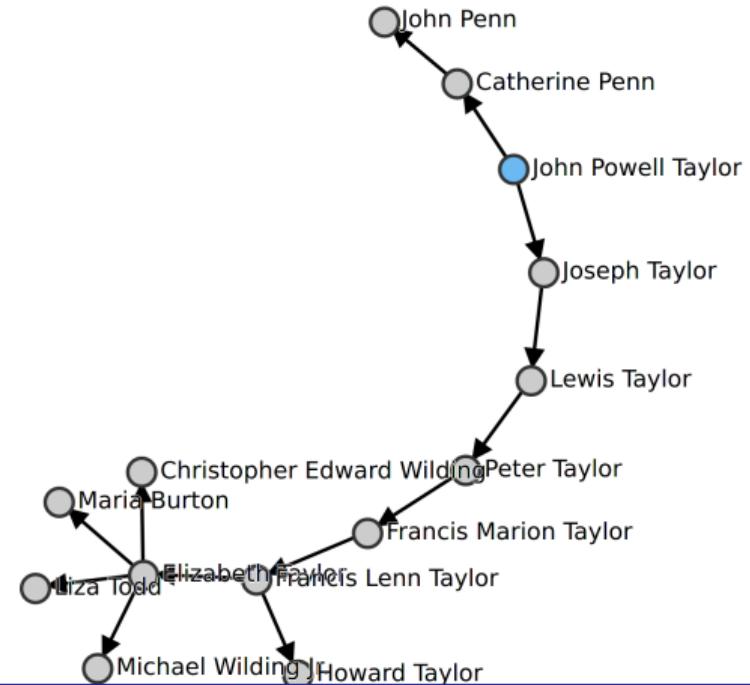
Elizabeth Taylor in other repositories

Identifiers		
VIAF ID	24624716 » 3 references	edit + add value
ISNI	0000 0001 2124 4985 » 1 reference	edit + add value
BNMM authority ID	000055929 » 0 references	edit + add reference + add value
BIBSYS ID	90544192 » 1 reference	edit + add value
CANTIC ID	a11572152 » 1 reference	edit + add value
Biblioteca Nacional de España ID	XX1130457	edit » 0 references

Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
Descendants of John Powell Taylor

Abstracting from definitions in “Towards a Definition of Knowledge Graphs”, Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.



Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
- It has information about real-world entities and relationships
People, places, etc.

Abstracting from definitions in “Towards a Definition of Knowledge Graphs”, Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.

Elizabeth Taylor (Q34851)

British-American actress, businesswoman, and humanitarian

Elizabeth Rosemond Taylor | Liz Taylor | Dame Elizabeth Rosemond Taylor | Dame Elizabeth Taylor



↳ Recoin: Most relevant properties which are absent

↳ In more languages

Language	Label	Description	Also known as
English	Elizabeth Taylor	British-American actress, businesswoman, and humanitarian	Elizabeth Rosemond Taylor Liz Taylor Dame Elizabeth Rosemond... Dame Elizabeth Taylor
German	Elizabeth Taylor	britisch-US-amerikanische Schauspielerin	Dame Elizabeth Rosemond... Elizabeth Rosemond Taylor Liz Taylor
French	Elizabeth Taylor	actrice britannique	Elizabeth Rosemond Taylor Elisabeth Taylor Liz Taylor

All entered languages

Statements

instance of	human	
	↳ 3 references	+ add value

image		
-------	--	--

Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
- It has information about real-world entities and relationships
- It defines interrelated classes and relations

Definition of the class human

Abstracting from definitions in “Towards a Definition of Knowledge Graphs”, Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.

human (Q5)			
		common name of Homo sapiens, unique extant species of the genus Homo	edit
English	human	common name of Homo sapiens, unique extant species of the genus Homo human being humankind people homosapiens person mankind peoplekind personkind persons » Recoin: Most relevant properties which are absent	
French	être humain	individu appartenant à l'espèce Homo sapiens, la seule espèce restante du genre Homo - distinct de « humain fictif » et de « humain possiblement fictif »	être humain Homme Homo sapiens sapiens Homo sapiens espèce humaine humaine humain

Statements			
instance of	of	organisms known by a particular common name	edit
		Homo sapiens Homo sapiens sapiens Homo	
		+ 0 references	+ add reference + add value

subclass of	of	edit
	person	+ 0 references

Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
- It has information about real-world entities and relationships
- It defines interrelated classes and relations

Definition of the class person, a superclass of human

Abstracting from definitions in “Towards a Definition of Knowledge Graphs”, Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.

person (Q215627)

being that has certain capacities or attributes constituting personhood (avoid use with P31; use Q5 for humans)
persons

↳ Recoin: Most relevant properties which are absent

▼ In more languages

Language	Label	Description	Also known as
English	person	being that has certain capacities or attributes constituting personhood (avoid use with P31; use Q5 for humans)	persons
French	personne	individu ou entité ayant certaines capacités et attributs constitutifs d'une personnalité. (Note: les éléments se rapportant à des êtres humains doivent utiliser « humain [Q5] » comme « nature de l'élément [P31] ».)	

All entered languages

Statements

subclass of	subject	edit + 0 references
	agent	edit + 0 references
	item with given name property	edit + 0 references
	individual	edit + 0 references

Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
- It has information about real-world entities and relationships
- It defines interrelated classes and relations

Definition of the relation spouse

Abstracting from definitions in “Towards a Definition of Knowledge Graphs”, Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.

spouse (P26)

the subject has the object as their spouse (husband, wife, partner, etc.). Use “unmarried partner” (P451) for non-married [companions](#) [edit](#)

wife | married to | marry | marriage partner | married | wedded to | wed | wives | husbands | spouses | husband | martial partner

* In more languages

Language	Label	Description	Also known as
English	spouse	the subject has the object as their spouse (husband, wife, partner, etc.). Use “unmarried partner” (P451) for non-married companions	wife married to marry marriage partner married wedded to wed wives husbands spouses husband martial partner
French	conjoint	personne liée par le mariage ou une union civile	épouse mari femme conjointe marié à compagne compagnon époux

[All entered languages](#)

Data type

Item

Statements

Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
- It has information about real-world entities and relationships
- It defines interrelated classes and relations

Definition of the relation child

Abstracting from definitions in “Towards a Definition of Knowledge Graphs”, Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.

child (P40)

subject has object as child. Do not use for stepchildren

son | daughter | kid | has child | children | sons | daughters | kids | has children | has son | has sons | has daughter | has daughters | has kid | has kids | offspring | progeny | issue | parent of | descendants

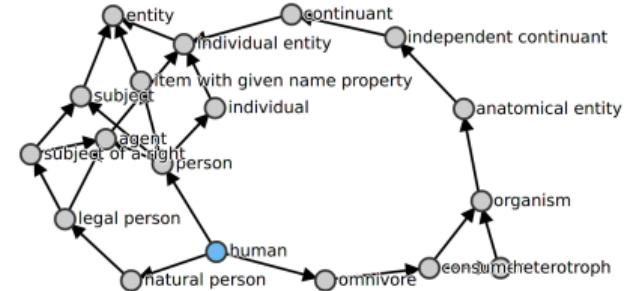
« In more languages

Language	Label	Description	Also known as
English	child	subject has object as child. Do not use for stepchildren	son daughter kid has child children sons daughters kids has children has son has sons has daughter has daughters has kid has kids offspring progeny issue parent of descendants
French	enfant	lien familial direct indiquant que l'objet du lien est un fils ou une fille du sujet. (indépendamment de son âge)	fils fille descendant filles descendants a pour fils Charles Antoine B... a pour fille a pour filles a pour descendants a pour descendant père de mère de

Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
- It has information about real-world entities and relationships
- It defines interrelated classes and relations
Generalizations of human

Abstracting from definitions in “Towards a Definition of Knowledge Graphs”, Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.

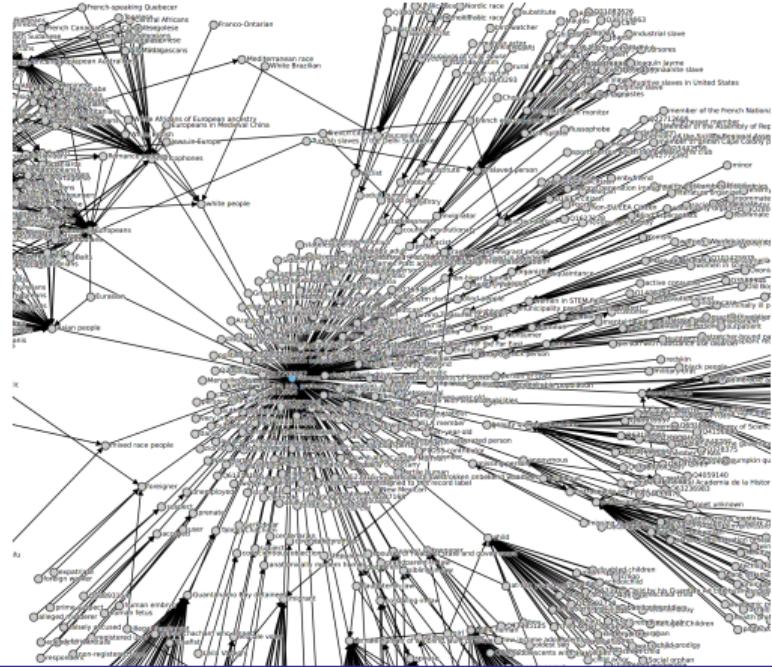


Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
- It has information about real-world entities and relationships
- It defines interrelated classes and relations

Some specializations of human

Abstracting from definitions in “Towards a Definition of Knowledge Graphs”, Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.



Is Wikidata A Knowledge Graph?

- It can be considered to be a graph
- It has information about real-world entities and relationships
- It defines interrelated classes and relations

Abstracting from definitions in “Towards a Definition of Knowledge Graphs”, Lisa Ehrlinger and Wolfram Wöß, SEMANTiCS 2016, Leipzig.

So, even though its name includes *data*, Wikidata meets requirements for Knowledge Graphs.

Does Wikidata Have Knowledge?

- How Can It Not?

Well . . .

Does Wikidata Have Knowledge?

- How Can It Not? *Well...*
- Information in Wikidata doesn't mean what you think it means
 - Consequences of information are not computed

How many humans in Wikidata?
9,067,016



The screenshot shows a SPARQL query interface. The query is:

```
1 SELECT { count(?item) as ?count }
2 WHERE {
3   ?item wdt:P31 wd:Q5 .
4 }
```

The results table shows one row:

count
9067016

Below the table, there are buttons for 'Code', 'Download', and 'Link'.

Does Wikidata Have Knowledge?

- How Can It Not? *Well...*
- Information in Wikidata doesn't mean what you think it means
 - Consequences of information are not computed

How many humans in Wikidata?
9,067,016 . . . or 9,088,310?
Need to explicitly count transitive closure of subclasses.
Also affects properties.



A screenshot of a SPARQL query results interface. The query is:

```
1 SELECT { count(?item) as ?count }
2 WHERE {
3   ?item wdt:P31/wdt:P279* wd:Q5 .
4 }
```

The results table shows one row:

count
9088310

At the bottom, there are buttons for 'Code', 'Download', and 'Link'.

Does Wikidata Have Knowledge?

- How Can It Not? *Well...*
- Information in Wikidata doesn't mean what you think it means
- Descriptions of classes are not correct
woman is described as female adult human

woman (Q467)

female adult human

female human | women | lady | womyn

» Recoin: Most relevant properties which are absent

+ In more languages

Language	Label	Description	Also known as
English	woman	female adult human	female human women lady womyn
French	femme	être humain adulte de sexe féminin	

All entered languages

Statements

subclass of

- adult
 - + 0 references
 - + add reference
- female human
 - + 0 references
 - + add reference
 - + add value

Image



Does Wikidata Have Knowledge?

- How Can It Not? *Well . . .*
- Information in Wikidata doesn't mean what you think it means
- Descriptions of classes are not correct
*But no (really 11) women in Wikidata
(At one time there were 42)*



A screenshot of a SPARQL query results interface. The query is:

```
1 SELECT ( count(?item) as ?count )
2 WHERE {
3   ?item wdt:P31 wd:Q467.
4   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en" . }
5 }
```

The results table shows one row with the following data:

count
0

Below the table, the status bar indicates "1 result in 10 ms".

Does Wikidata Have Knowledge?

- How Can It Not? *Well . . .*
- Information in Wikidata doesn't mean what you think it means
- Descriptions of classes are not correct
*But no (really 11) women in Wikidata
(At one time there were 42)*



A screenshot of a SPARQL query results interface. The query is:

```
1 SELECT { count(?item) as ?count }
2 WHERE {
3   ?item wdt:P31/wdt:P279* wd:0467.
4   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
5 }
```

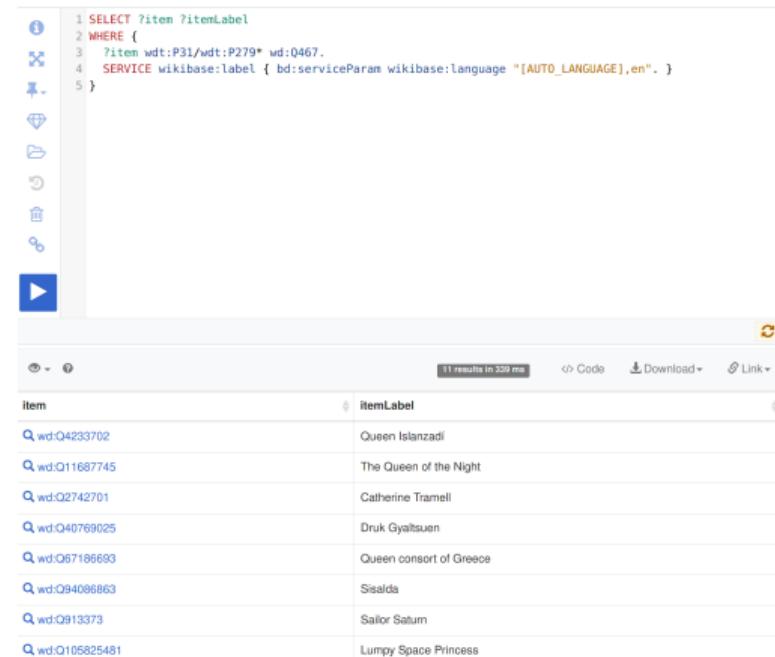
The results table shows one row:

count
11

At the bottom, it says "1 result in 164 ms".

Does Wikidata Have Knowledge?

- How Can It Not? *Well . . .*
- Information in Wikidata doesn't mean what you think it means
- Descriptions of classes are not correct
*But no (really 11) women in Wikidata
(At one time there were 42)*

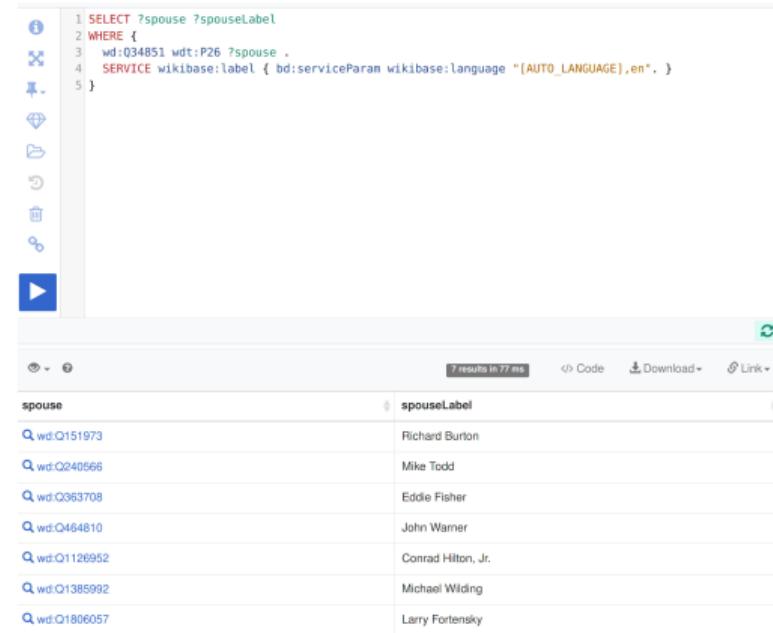


```
1 SELECT ?item ?itemLabel
2 WHERE {
3   ?item wdt:P31/wdt:P279* wd:0467.
4   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
5 }
```

item	itemLabel
wd:Q4233702	Queen Isilanzadi
wd:Q11687745	The Queen of the Night
wd:Q2742701	Catherine Tramell
wd:Q40769025	Druk Gyalsuuen
wd:Q67186693	Queen consort of Greece
wd:Q94086863	Sisalda
wd:Q913373	Sailor Saturn
wd:Q105825481	Lumpy Space Princess

Does Wikidata Have Knowledge?

- How Can It Not? *Well...*
- Information in Wikidata doesn't mean what you think it means
- Descriptions of classes are not correct
- Context is not taken into account
 - Who are the spouses of Elizabeth Taylor?



The screenshot shows a SPARQL query interface with the following details:

Query:

```
1 SELECT ?spouse ?spouseLabel
2 WHERE {
3   wd:Q34851 wdt:P26 ?spouse .
4   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en" . }
5 }
```

Results:

spouse	spouseLabel
wd:Q151973	Richard Burton
wd:Q240566	Mike Todd
wd:Q363708	Eddie Fisher
wd:Q464810	John Warner
wd:Q1126952	Conrad Hilton, Jr.
wd:Q1385992	Michael Wilding
wd:Q1806057	Larry Fortensky

Does Wikidata Have Knowledge?

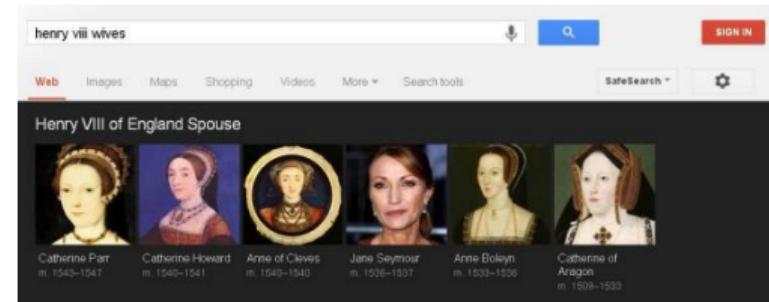
- How Can It Not? *Well . . .*
- Information in Wikidata doesn't mean what you think it means
- Descriptions of classes are not correct
- Context is not taken into account
 - What context (qualifiers) needs to be taken into account is unclear

spouse	 Richard Burton	
	start time	15 March 1964
	end time	26 June 1974
	end cause	divorce
	series ordinal	5
	 1 reference	
	 Richard Burton	
	start time	10 October 1975
	end time	29 July 1976
	end cause	divorce
	series ordinal	6
	 1 reference	
	 Conrad Hilton, Jr.	
	end time	29 January 1951
	start time	6 May 1950
	end cause	divorce
	series ordinal	1
	 1 reference	
	 Michael Wilding	
	end time	30 January 1957
	start time	21 February 1952
	end cause	divorce
	series ordinal	2
	 1 reference	
	 Mike Todd	

Does Wikidata Have Knowledge?

- How Can It Not? *Well ...*
- Information in Wikidata doesn't mean what you think it means
- Descriptions of classes are not correct
- Context is not taken into account
- Few sanity checks are made

*A picture of a wife of Henry VII?
(from Google)*



[The Six Wives of Henry VIII - Tudor History](#)
[Tudorhistory.org/wives/](#) «
Provides short biographies of each of the women to whom the king was married.
Kathryn Howard - Anne of Cleves - Anne Boleyn - Catherine Parr

[Wives of Henry VIII - Wikipedia, the free encyclopedia](#)
[en.wikipedia.org/wiki/Wives_of_Henry_VIII](#) «
Catherine Howard and Anne Boleyn, Henry's two queens who were ... Henry VIII actually had only two wives (Jane Seymour and Catherine Parr), despite having ... Anne of Cleves (22 September 1515 – 16 July 1557) was Henry's fourth wife, ... Catherine of Aragon - Anne Boleyn - Jane Seymour - Anne of Cleves

[The Six Wives of Henry VIII. Meet the Wives | PBS](#)
[www.pbs.org/wnet/sixwives/meet/](#) «



What is Knowledge?

Adapted from Russell Ackoff, "From Data to Wisdom" (1989).

- Data: Simple findings and figures
- Information: Collected and analyzed data
- Knowledge: Information organized and ready to put to use
- Wisdom: Proper use of knowledge

See also T. S. Eliot, *The Rock* (1934).

Does Wikidata Have Knowledge

Using Wikidata requires knowledge that is not in the data

- Humans can (often? sometimes?) determine what is needed
- Programs generally can't

So information in Wikidata isn't ready to use.

Thus Wikidata doesn't really have knowledge.

Turning Wikidata Into Wiki-knowledge

- Add in consequences of data (but what, exactly are they?)
 - Design a logic for Wikidata to define consequences
 - Implement a reasoner for the logic
- Incorporate information from descriptions
 - By turning them into expressions in the logic
- Take context into account
 - By providing simple ways to combine qualifiers in the logic
 - By augmenting query interfaces with a current context
 - not considered further here*

MARS, MARPL, eMARS, eMARPL

- A family of logics for Knowledge Graphs
 - Augment regular logic with qualifiers on facts
 - Add Wikidata datatypes
 - Add rules for handling qualifiers
 - Add constraints
- MARPL and eMARPL can be implemented using forward-chaining rules
 - Complexity of inference is not too bad

Turning Wikidata Into Wiki-knowledge

- Use eMARPL as the logic for Wikidata

```
instance_of(Elizabeth_Taylor, human)
```

```
spouse(Elizabeth_Taylor, Richard_Burton)@{start_time : '15 March 1964', ...}
```

Turning Wikidata Into Wiki-knowledge

- Use eMARPL as the logic for Wikidata
- Prohibit context qualifiers in the ontology
 - Not absolutely necessary, but makes the process easier

Turning Wikidata Into Wiki-knowledge

- Use eMARPL as the logic for Wikidata
- Prohibit context qualifiers in the ontology
- Add rules that determine consequences of ontology properties

$$\text{subclass_of}(c, d) \wedge \text{subclass_of}(d, e) \rightarrow \text{subclass_of}(c, e)$$
$$\text{instance_of}(y, c)@X \wedge \text{subclass_of}(c, d) \rightarrow \text{instance_of}(y, d)@X$$
$$\text{subproperty_of}(c, d) \wedge \text{subproperty_of}(d, e) \rightarrow \text{subproperty_of}(c, e)$$

Turning Wikidata Into Wiki-knowledge

- Use eMARPL as the logic for Wikidata
- Prohibit context qualifiers in the ontology
- Add rules that determine consequences of ontology properties
- Add rules that determine consequences of domain properties

$$\text{spouse}(x, y)@X \rightarrow \text{spouse}(y, x)@X$$

Turning Wikidata Into Wiki-knowledge

- Use eMARPL as the logic for Wikidata
- Prohibit context qualifiers in the ontology
- Add rules that determine consequences of ontology properties
- Add rules that determine consequences of domain properties
- Convert Wikidata constraints to logic constraints

$\text{property_constraint}(p, \text{distinct_values_constraint}) \wedge p(s1, o1) \wedge p(s2, o2) \wedge s1 \neq s2 \rightarrow o1 \neq o2$

Turning Wikidata Into Wiki-knowledge

That was the easy part, now for the hard part

Turning Wikidata Into Wiki-knowledge

That was the easy part, now for the hard part

- Provide rules for combining qualifiers

`start_time : max(start_time)`

`end_time : min(end_time)`

Turning Wikidata Into Wiki-knowledge

That was the easy part, now for the hard part

- Provide rules for combining qualifiers

`start_time : max(start_time)`

`end_time : min(end_time)`

Difficult to determine desired meaning of some combinations

Turning Wikidata Into Wiki-knowledge

That was the easy part, now for the hard part

- Provide rules for combining qualifiers
- Provide recognition rules for many classes

$$\text{instance_of}(\text{human}, h) \wedge \text{sex_or_gender}(h, \text{female}) \rightarrow \text{female_human}(h)$$

Turning Wikidata Into Wiki-knowledge

That was the easy part, now for the hard part

- Provide rules for combining qualifiers
- Provide recognition rules for many classes

$$\text{instance_of}(\text{human}, h) \wedge \text{sex_or_gender}(h, \text{female}) \rightarrow \text{female_human}(h)$$

Very many such classes, some may not have good recognition conditions

Status (Implementation)

- Just a proposal ...
- There are implementations of property graphs that can store Wikidata statements
- Some graph databases have rules, e.g., Ontotext GraphDB
- No system implements rules as described above.
- Considerable work would be required to implement the logic here
- But it's just forward chaining rules
 - Could implement by encoding in RDF and using RDF rule engine

Status (Community)

- Wikidata community is fragmented
- Portions of Wikidata community not friendly to this kind of effort
 - Seem to want to put more work on consumers of Wikidata
- Support group for Wikidata (Wikimedia Deutschland) is resource constrained

Summary

- Wikidata is a useful knowledge graph
- Like other knowledge graphs, Wikidata doesn't quite qualify as providing knowledge
- A formal account can be given for (most of) Wikidata that adds the missing pieces
 - Needs implementation
 - Not too much pain to implement
 - Computational aspects not horrible
 - Consumers see knowledge, much less pain
- Buy-in for this approach is missing

References

Peter F. Patel-Schneider and David Martin. Wikidata on MARS. DL Workshop, 2020.

David Martin and Peter F. Patel-Schneider. Wikidata Constraints on MARS. Wikidata workshop at ISWC 2020.

Marx, M., Krötzsch, M., Thost, V. Logic on MARS: Ontologies for generalised property graphs. IJCAI 2017.

Krötzsch, M., Marx, M., Ozaki, A., Thost, V. Attributed description logics: Ontologies for knowledge graphs. ISWC 2017.

Krötzsch, M., Marx, M., Ozaki, A., Thost, V. Attributed description logics: Reasoning on knowledge graphs. IJCAI 2018.

R.L. Ackoff. From data to wisdom. *Journal of applied systems analysis*, 1989.
<https://pdfs.semanticscholar.org/ea36/744970968bc18f9cd89592a86c6d2a366ba7.pdf>

T. S. Eliot. *The Rock*. Harcourt, Brace, and Company, 1934.
<https://books.google.com/books?id=Qmr8AgAAQBAJ>