

## Everything you need to become a data quality hero on Wikidata!



#### Why does this matter?

Wikidata is a central piece of infrastructure in today's technology world. Our data is used in the Wikimedia Projects and far beyond to help give more people more access to more knowledge. We want to provide verifiable, reliable and up-to-date information to everyone.

#### What types of data quality issues are there?

- Vandalism
- Ontology issues
- Unintentionally bad edits
- Missing data (especially references)
- ...

### Your secret super power: Queries query.wikidata.org

Many data quality tools rely on queries to help find wrong, missing, etc. data. You can query directly with SPARQL in the **Wikidata Query Service** (e.g. "politicians born within the last century missing a picture"). Or you can use the **Query Builder** for an easier (but slightly

Wikidata Query Service 🕞 Examples 🛛 Help 🗸 🌣 More tools 🗸	Wikidata Query Builder The Wikidata Query Builder provides a visual interface for building a simple Wikidata query. It is ideal for users with line or no experience in <u>SRAPQ</u> , the powerful query language. The Query Builder doesn't offer SRAPQI: Is ful interfaceasy genry part query up in the Query Service, where you can view, edit or expand it via the link above the results. <u>Excelback is welcome here</u> .
<pre>1 #Map of Libraries in Canada 2 # Canadian libraries on a map (must have coordinates!) 3 #defaultView:Map 4 SELECT ?library ?libraryLabel ?coords WHERE { 4 SELECT ?library ?libraryLabel ?coords WHERE { 5 SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en,fr". } 6 ?library (wdt:P31/wdt:P279*) wd:Q7075. 7 ?library wdt:P17 wd:Q16.</pre>	Query         Find al items         Value o       References o         With Without       instance of         matching       superhero         include related values in o       with and without         ite search (recommended)
<pre>8 ?library wdt:P625 ?coords. 9 } 3 </pre>	and or Property O Value O References O =
<b>逾</b>	With     Without     from narrative universe     matching     Marvel Cinematic Universe     with and without references       Include related values in the search (recommended)
	Add condition Settings
	Limit the number of results to     100     Show IDs instead of labels (may prevent timeout)

Documentation Wikimedia Toolforge	Source code Logged in	as Lydia Pintscher (WMDE)					
1049838 (history)			Wikidata vandalism	<b>dashboard</b> Hom	e Toolforge Source code		
Unatasful dded link to [viwiki]: Động đất Tônankai 19	Eiyafirayoku 944 Added link to [commonswiki]: Categ Tonankai earthquake	gory:1944	Language code(s de, en	*			
nks / commonswiki / name	links / commonswiki / name		Limit 20	0			
	+ Category:1944 Tonankai earthquake		✓ Changes in	descriptions 🖌 Ch	anges in labels and aliases 🕑 Sitelink removals		
	۹ 🖍	2 2	Search				
	_		Edit ID	Entity ID	Username	Entity title	Edit summ
Gira Submit your program	proposal for the GLAM Wiki Conferenc	e	1942895466	Q29963486	2A02:908:3611:8280:25D0:633E:5E65:BEEA	Escape Room	Changed la
warcoweases			1942892486	Q67221078	84.9.238.191	château de Cros	Added des
	[Help with translation:	s!]	1942892483	Q67221078	84.9.238.191	château de Cros	Added labe
1044 Tāpapkai aarthau			1942892270	Q67221071	84.9.238.191	château Cornudet	Added deso
1944 Tōnankai earthqu	(Q1049838)		1942892265	Q67221071	84.9.238.191	château Cornudet	Added labe
Item Discussion			1942891261	Q114310637	84.9.238.191	château de Beauregard	Added des
☆ <sup>©</sup>		:	1942891257	Q114310637	84.9.238.191	château de Beauregard	Added labe
			1942891172	Q56160823	84.9.238.191	château de Beauregard	Added desc
Revision as of 13:50, 30 July 2023 by to [commonswiki]: Category:1944 Tone	Eiyafirayoku (talk   contribs   block) (Addee	d link	1942890837	Q83711874	84.9.238.191	château de Beauregard	Added desc
	a   Latest revision (diff)   Newer revision →	(diff)	1942890835	Q83711874	84.9.238.191	château de Beauregard	Added labe
			1942890801	Q83711911	84.9.238.191	château de Beauregard	Added desc
RES predicted quality: C (2.98 🖄)			1942890799	Q83711911	84.9.238.191	château de Beauregard	Added labe
		🖍 edit	1942890717	Q83711855	84.9.238.191	château de Beauregard	Added desc
December 1944 earthquake in Japan							

#### Vandalism fighting

speedpatrolling.toolforge.org pltools.toolforge.org/rech wdvd.toolforge.org

- A lot of problematic edits are automatically tagged and you can **filter Recent Changes** by them to quickly find issues to fix.
- **Speed Patrolling** shows you an edit and gives you quick options to mark it as patrolled or revert it. Great for a boring subway ride!
- **reCh** gives you a quick overview of recent edits incl. options to revert problematic ones.
- The Wikidata Vandalism Dashboard is especially useful to monitoring edits to labels, descriptions, aliases and sitelinks for a specific language.

#### Finding ontology issues and other problematic data

Help:Constraints WD:Schemas mismatch-finder.toolforge.org

- **Constraint Checks** and **EntitySchemas** are both ways to encode how data should be modeled to make it possible to find statements that do not match the agreed upon modeling. When a Constraint is violated there will be a little warning next to the statement explaining what is wrong.
- The **Mismatch Finder** is there to compare Wikidata's data against other sources. When the data isn't the same you can review it and make changes as necessary.

				ATA					Mattia_Capozzi_WMDE	Lo
			Mismate	h Finder						
			← Refine Items	Selection o with the mismatch	nes?					
	Suggestions × citation needed constraint Help Discuss	+ add reference	You can fix them! entry on the exter Wikidata or the ex	The table below presen nal source. You can con cternal data source. Afte the options from the dr	its you with links to th npare the values and er investigating and fi	manually fix the misma xing the mismatch, you	atch, editing			
Statements for date of birth should have at least one reference.		+ add value	Jane Goodall (Q	184746)						
f birth	🗟 -> 27 February 1978 •••• 🏲	🎤 edit	Property	Value on Wikidata	Value on external source	External Source	Status		Upload Info	
	• 0 references		Date of birth	4 September 1990	3 April 1914	VIAF	To be defined	~	WikiMacaroons 20.04.2021 Batch_Upload_1	
		+ add reference + add value	Place of birth	Hampstead, UK	Sicily, Italy	BnF Authorities	To be defined	~	ToBeFree 11.06.2021 Upload_final	
			Sex or gender	female	male	VIAF	To be defined	~	Llywrch 29.01.2021 Upload_12	
			Language Spoken	English	Japanese	Who's Who	To be defined	~	Fjjulien 01.11.2020 Batch_Upload_190	
									Apply	chan

- Refine Item selection	< Wikidata:Lists This list is periodically updi	Wikidata:Lists/List of canonical She - WikidataLits The let is periodsity updated by a bot, Manual changes to the let will be removed on the next WOO39 (Proceed) The Manuard (Prod magenting Research changes (Prod magenting Resea				
verage quality score based on the 100 Items selected: 2.54 Scores can go from 1 (worst quality) to 5 (best quality)		image	Work •	descriptio		
scores for 100 Items	لي Download as CSV	mis flowenites as Data Wilson				
Label	♦ Score ♦		The Adventures of Sherlock Holmes	collection of short stories b		
Daredevil (Q327553)	3.37	<b>GCi</b> 7 A Stran Door				
-alcon (Q672591)	3.19	A STUDY IN				
olaris (Q2071256)	3.09	SCARLET	A Study in Scarlet	first Sherlock Holmes nove Doyle		
ean Grey (Q5350072)	3.06	The second s		a signal		
iles Morales (Q59826)	3.06					
Gambit (Q2639150)	3.06	1 Dour	The Sign of the Four	ur novel		
uke Cage (Q2267760)	3.06	63	ine sign of the rote			
Captain Britain (Q2405949)	3.04	A Court Only				
X-23 (Q1872439)	3.04	100 A 100	The Adventure of the Red-Headed League	short story by Arthur Conar		
Psylocke (Q973050)	3.03	the second	reague			
Cable (Q2267017)	3.03					

# Which is a constructed Sherblock Holmes note by Statute of the statu

#### Monitoring and evaluating

item-quality-evaluator.toolforge.org integraality.toolforge.org WD:Listeria

- The **Item Quality Evaluator** gives every Item a quality score. You can create a list of Items to find the lowest-quality ones for improvement. Great for an editathon!
- Integraality and Listeria are tools to put query results in a wiki page. The wiki page is updated regularly so you can put it on your watchlist and see any changes in the data you care about.

#### WikiProjects

Here you can find your fellow heroes and coordinate your next mission:

- WD:WikiProject Counter-Vandalism
- WD:WikiProject Data Quality

#### Contact

Lydia Pintscher lydia.pintscher@wikimedia.de @wikidata

cc-by-sa 4.0