

Using Wikidata Query and QuickStatements to automatically amend Wikidata Items

Run Wikidata Queries (SPARQL)

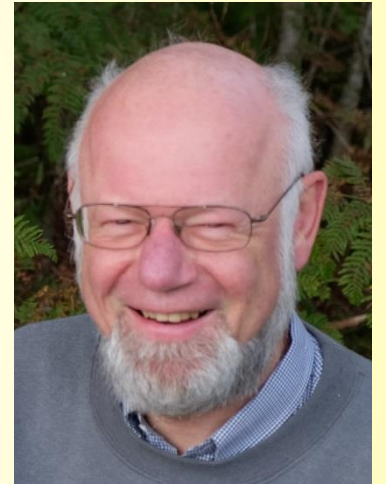
Detect anomalies


Correct them with QuickStatements

Geert Van Pamel

- Wikipedian since 2005
- Member of WMNL since 2012
- Using Wikidata since 2012
- Contributing to OpenStreetMap since 2013
- Board member of WMBE since 2014
- Living in Belgium (Vlaams Brabant)
- President of Wikimedia Belgium
- Professionally working with Geospatial and Big data systems
- Developer for Open source projects (Asterisk PBX); loving Linux

- CC BY-SA
- <https://commons.wikimedia.org/wiki/File:Wikidata-20171028-WMBE-Query-QuickStatements.pdf>





“ Mutual symbiosis --
an almost living ecosystem

Organising the content

”

Infrastructure: application, servers, network.

Content, structure, policy, procedures.

Volunteers, Organisations, Wikipedians.

Collaboration, creativity, build upon others, stronger together.

Free Licenses.



Generic and hierarchical database

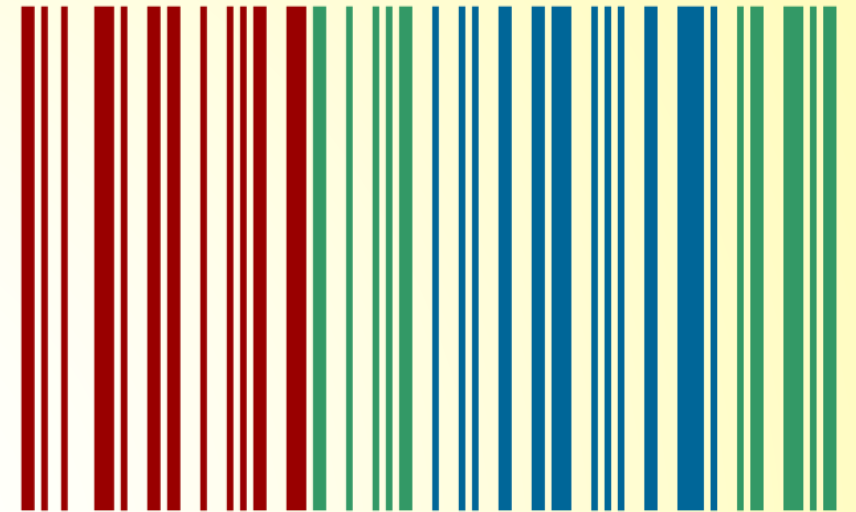
Linked objects and attributes

(Meta)data

Entities, Attributes

Relationships, Values

Inventories, Lists



WIKIDATA

Model the world

Acronyms

| Acronym | Description |
|--------------------------------------|--|
| RDF | Resource Description Framework |
| IPV, KVP | Triple store |
| Statements: | |
| Item, Instance, Subject | Anything existing in the real world |
| Property, Attribute | Defined Properties (that you can propose!) |
| Object, Value, (external) Identifier | Data about the Item |

The Wikidata license



| Acronym | Description | Non-commercial fallback |
|----------|---------------------|-------------------------|
| CC0 | Public domain | - |
| CC BY | Attribution | CC BY-NC |
| CC BY-SA | Share alike | CC BY-NC-SA |
| CC BY-ND | Non derivative | CC BY-NC-ND |
| © | All Rights Reserved | - |



Motivation and Goal



Wikidata is extremely open.

Anyone can edit.

The quality of the data needs to be continuously verified.

Wikidata query allows to code SPARQL queries in an interactive way.

It allows to generate reports on inconsistent or incomplete data.

QuickStatements allows to automatically amend Objects & Attributes & Relationships.

Data improvement process



Wikidata generic data model

| Context | Triples | | |
|---------------------|-------------|-----------------|-----------------|
| Wikidata term: | Item | Property | Value |
| Database term: | Entity | Attribute | Relationship |
| Object Label: | Person | Nationality | Country |
| Instance Label: | Eddy Merckx | Nationality | Belgium |
| Statement: | Eddy Merckx | has nationality | Belgian citizen |
| RDF encoded triple: | Q103756 | P27 | Q31 |

RDF in Wikidata

RDF Triple presented in Wikidata:

- <https://www.wikidata.org/wiki/Q103756>
- <https://www.wikidata.org/wiki/Property:P21>
- <https://www.wikidata.org/wiki/Q31>

Presented in Reasonator:

- <https://tools.wmflabs.org/reasonator/?q=Q103756>
 - Eddy Merckx (Q103756)
 - road bicycle racer
 - Eddy Merckx is a Belgian sport cyclist.
 - He was born on June 17, 1945 in Meensel-Kiezegem.
 - His children include Sabrina and Axel Merckx.



Completeness

In one or more languages:

- Missing Labels
- Missing Descriptions
- Missing Alias

- Missing Properties (profession)
- Missing (untranslated) professions
- Missing Interwiki links (how is a new page detected if it has a different label?)
- Links to other Wikimedia projects

Constraints

- Unique per language: Label – Description
 - No two Items can have the same Label + Description (per language)
 - Persons with the same first name & family name
 - Require birth date/date of death suffix

Enable Check Constraints:

- Duplicate or conflicting Properties (some is wrong; or has a different accuracy)
 - Duplicate statements (duplicate/conflicting birth dates/nationality)
- Reciprocal Properties missing: part of – has part (municipality – submunicipality)
- Symmetric Properties missing: neighbouring cities, sister cities



Quality



- Conflicting statements (person versus hotel)
- Non-matching property-value (woman – female, man – male, country – nationality)
- Old/new places (history versus current)
- Female functions (gender titles)
- Description not matching with “is an instance of” (persons can have multiple instances)
- Description same as Label
- Label equal to Alias
 - Alias and Description can be swapped amongst languages




Languages



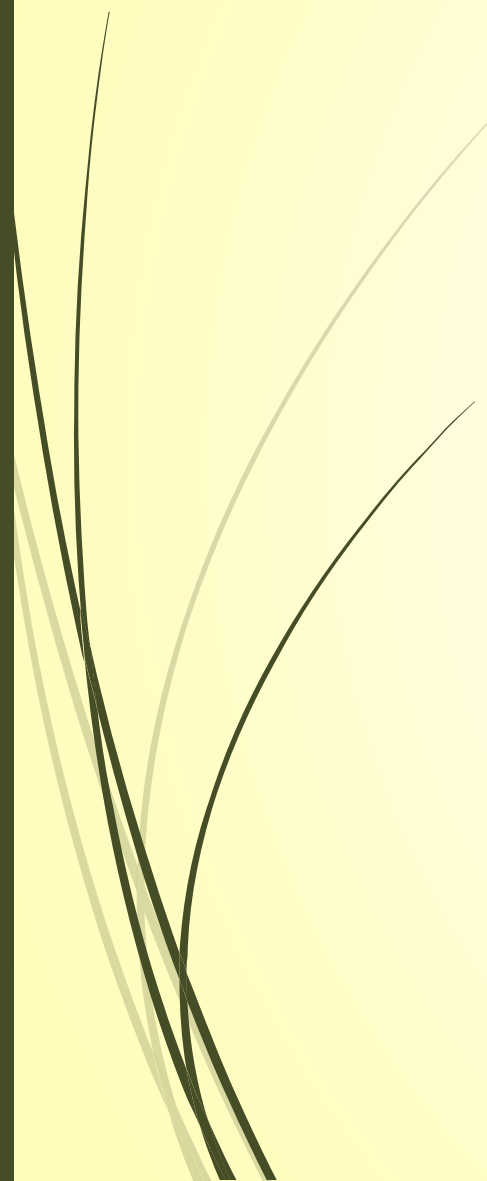
Preferred languages:

- EN: largest number of items (central exchange and reference)
- Your own language: cultural richness, search
- Multi-lingual countries (Belgium, Canada, Switzerland) ⇒ replicate to other language
- Other volunteers will amend their own language content

- Polyglot: be sure you use the right wording
 - Different languages have their own preferences/naming rules



Tools

- Wikidata
 - Wikidata Query
 - QuickStatements
 - Reasonator
- 

Wikidata Query

- Wikidata Query is a GUI-based query tool
- Using the SPARQL standard from W3C
- Running on RDF databases
- Multilingual, like Wikidata
- Easy to use, once you understand the IPV-system
- Interactive
- You can choose the output result format: Table, Chart, Graph, Map, Timeline, Tree
- <https://query.wikidata.org>



QuickStatements

- Automatic execution of Wikidata edits
- Prepare a flat transaction file
- Updates the Wikidata database
- Produces an error log ⇒ minor manual corrections could be required
- Provide user credentials via WiDaR
- https://tools.wmflabs.org/wikidata-todo/quick_statements.php
- Note: There is a small delay (minutes) between Wikidata updates and Wikidata Query
- A small User Guide (how to) is included



Reasonator

Can help you to verify the results

➤ <https://tools.wmflabs.org/reasonator/?q=Q103756>

Example with sister city

- ▶ This is a really simple mechanism

```
SELECT ?item ?itemLabel ?itemDescription ?sister_city ?sister_cityLabel WHERE {  
  ?item wdt:P31 wd:Q2039348.  
  ?item wdt:P190 ?sister_city.  
  SERVICE wikibase:label { bd:serviceParam wikibase:language "nl". }  
  FILTER(NOT EXISTS { ?sister_city wdt:P190 ?item. })  
}  
ORDER BY ?itemDescription
```

Transaction file (partial)

| Result preview | | | | |
|---|------------|--|---|------------------|
| item | itemLabel | itemDescription | sister_city | sister_cityLabel |
|  wd:Q10076 | Middelburg | Nederlandse gemeente in Zeeland |  wd:Q66013 | Głogów |
|  wd:Q9899 | Amsterdam | Nederlandse gemeente, omvattende de stad Amsterdam |  wd:Q1899 | Kiev |
|  wd:Q9899 | Amsterdam | Nederlandse gemeente, omvattende de stad Amsterdam |  wd:Q18125 | Manchester |

- Download TSV file
- Prepare Excel transaction file

| Item | Property | Value |
|------------------------|----------|------------------------|
| Q66013 | P190 | Q10076 |
| Q1899 | P190 | Q9899 |
| Q18125 | P190 | Q9899 |

Result

- Copy/paste from Excel
- Execute the transaction file
- https://tools.wmflabs.org/wikidata-todo/quick_statements.php

| Item | Property | Value |
|------------------------|----------|------------------------|
| Q66013 | P190 | Q10076 |
| Q1899 | P190 | Q9899 |
| Q18125 | P190 | Q9899 |

Click on “Do it”

You are logged into WiDaR as (Your user name)

Processing Q66013 (Q66013 P190 Q10076)

Processing Q1899 (Q1899 P190 Q9899)

Processing Q18125 (Q18125 P190 Q9899)

All done!.

Verify the results

User contributions

- 16:02, 21 October 2017 (diff | hist) . . (+407) . . Manchester (Q18125) (Created claim: sister city (P190): Amsterdam (Q9899), #quickstatements) (**current**) (Tag: Widar [1.4])
- 16:02, 21 October 2017 (diff | hist) . . (+406) . . Kiev (Q1899) (Created claim: sister city (P190): Amsterdam (Q9899), #quickstatements) (**current**) (Tag: Widar [1.4])
- 16:02, 21 October 2017 (diff | hist) . . (+408) . . Głogów (Q66013) (Created claim: sister city (P190): Middelburg (Q10076), #quickstatements) (**current**) (Tag: Widar [1.4])

Revision history of "Manchester" (Q18125)

- (cur | prev) 16:02, 21 October 2017 Geertivp (talk | contribs) . . (71,546 bytes) (+407) . . (Created claim: sister city (P190): Amsterdam (Q9899), #quickstatements) (undo) (Tag: Widar [1.4])

Other examples available from:

- <https://www.wikidata.org/wiki/Special:PrefixIndex/User:Geertivp/training>



Review errors

- Double check before you start updates
- Do not swap Label/Description Lxx – Dxx
- Pay attention to UTF-8 accents (download does not work properly in WD Query)
- Verify if the Item/Property does not already exist
- Conflicting updates (already exist, constraint violation)
- When creating Objects, use at least “is an instance of”
- Otherwise it is considered an empty Item that will be deleted by the moderators (even if it contains a Label/Description in one language...)



Prepare yourself

As a result of your automatic corrections/updates/inserts:

- You might make (mass) errors
- You might receive (angry) remarks on your discussion page
- People do not always understand what you are doing
- Describe clearly what you are doing (what your intentions are)
- Excuse for your mistakes when appropriate
- Verify your updates before executing
- Sometimes (small) manual updates might be better/more easy



Other tools

- Wikidata GUI itself
- The Wikidata built-in search engine
- Mix'n match
- Wikidata Game
- Reasonator
- ...
- WiDaR - central authorisation Oauth protocol
- Wikipedia
- https://www.wikidata.org/wiki/Wikidata:Tools/External_tools



Training and Support



- <https://commons.wikimedia.org/wiki/File:Wikidata-20171028-WMBE-Query-QuickStatements.pdf>
- <https://www.wikidata.org/wiki/Special:PrefixIndex/User:Geertivp/training>

Contact

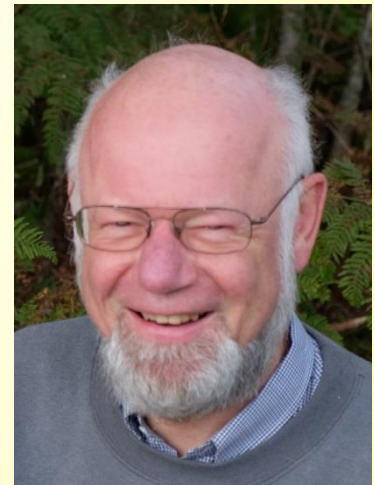
Geert Van Pamel

<https://nl.wikipedia.org/wiki/Gebruiker:Geertivp>

<mailto:geert@wikimedia.be>

Twitter @Geertivp

Was ik metselaar dan was Linux mijn truweel





Acknowledgements



- Tim Berners-Lee to provide us with the wonderful technique of interlinking web pages with URL
 - Otherwise Wikidata would not have been possible
- Wikidata for stimulating my imagination
- My Linux servers, for allowing me to experiment:
 - Tru64 UNIX on AlphaServer for running my company MediaWiki since 2005
 - Raspberry Pi 3/Raspbian running a local MediaWiki server



References



1. https://commons.wikimedia.org/wiki/File:Wikimedia_Belgium.svg
2. https://commons.wikimedia.org/wiki/File:Wikidata_barcode.svg
3. <https://commons.wikimedia.org/wiki/File:CreativeCommonsSpectrum.PNG>