

# Adding chemical compounds to Wikidata

Egon Willighagen 

Published January 27, 2016

## Citation

Willighagen, E. (2016, January 27). Adding chemical compounds to Wikidata. *Chem-bla-ics*.  
<https://doi.org/10.59350/q5a38-51n24>

## Keywords

Wikidata, Chemistry, Bioclipse

## Abstract

Adding chemical compounds to Wikidata is not difficult. You can store the chemical formula (P274), (canonical) SMILES (P233), InChIKey (P235) (and InChI (P234), of course), as well various database identifiers (see what I wrote about that here [\]\]](#)). It also allows storing of the provenance, and has predicates for that too.

## Copyright

Copyright © Egon Willighagen 2016. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.


Adding chemical compounds to [Wikidata](#) is not difficult. You can store the chemical formula ([P274](#)), (canonical) [SMILES](#) ([P233](#)), InChIKey ([P235](#)) (and InChI ([P234](#)), of course), as well various database identifiers (see what I wrote about that [here](#) ]). It also allows storing of the provenance, and has predicates for that too.

So, to enter a new structure for a compound, you should enter the compound information to Wikidata. Of course, make sure to create the needed accounts, particularly one for Wikidata ([create account](#)) (not sure if the next steps needs a more general Wikimedia account too).

### Entering the research paper:

[Magnus Manske](#) pointed me to [this helper tool](#). If you have the DOI of the paper, it is easy to add a new paper. This is what the tool shows for doi:[10.1128/AAC.01148-08](#) (but no longer when you try!):

Source MetaData

Tools Git Talk 

You can create a new item, or update an existing one, for this source in QuickStatements:

CREATE

LAST	P698	"19223625"
LAST	P932	"2681551"
LAST	P356	"10.1128/AAC.01148-08"
LAST	P31	Q13442814
LAST	P1476	en:"A small-molecule dengue virus entry inhibitor"
LAST	Len	"A small-molecule dengue virus entry inhibitor"
LAST	P364	Q1860
LAST	P1433	Q578004
LAST	P478	"53"
LAST	P433	"5"
LAST	P577	+2009-05-00T00:00:00Z/10
LAST	P304	"1823-31"
LAST	P2093	"Qing-Yin Wang" P1545 "1"

Open in QuickStatements

You need permission to run this script and the tool will alert you about that, and give the instructions how to get permission. After I clicked the Open in QuickStatements I get this output, showing me an entry in Wikidata was created for this paper:

## chem-bla-ics

1. Processing [Q22309806](#) (Q22309806 P698 "19223625")
2. Processing [Q22309806](#) (Q22309806 P932 "2681551")
3. Processing [Q22309806](#) (Q22309806 P356 "10.1128/AAC.01148-08")
4. Processing [Q22309806](#) (Q22309806 P31 Q13442814)
5. Processing [Q22309806](#) (Q22309806 P1476 en:"A small-molecule dengue virus entry inhibitor")
6. Processing [Q22309806](#) (Q22309806 Len "A small-molecule dengue virus entry inhibitor")
7. Processing [Q22309806](#) (Q22309806 P364 Q1860)
8. Processing [Q22309806](#) (Q22309806 P1433 Q578004)
9. Processing [Q22309806](#) (Q22309806 P478 "53")
10. Processing [Q22309806](#) (Q22309806 P433 "5")
11. Processing [Q22309806](#) (Q22309806 P577 +2009-05-00T00:00:00Z/10)
12. Processing [Q22309806](#) (Q22309806 P304 "1823-31")
13. Processing [Q22309806](#) (Q22309806 P2093 "Qing-Yin Wang" P1545 "1")
14. Processing [Q22309806](#) (Q22309806 P2093 "Sejal J Patel" P1545 "2")
15. Processing [Q22309806](#) (Q22309806 P2093 "Eric Vangrevelinghe" P1545 "3")
16. Processing [Q22309806](#) (Q22309806 P2093 "Hao Ying Xu" P1545 "4")
17. Processing [Q22309806](#) (Q22309806 P2093 "Ranga Rao" P1545 "5")
18. Processing [Q22309806](#) (Q22309806 P2093 "Deana Jaber" P1545 "6")
19. Processing [Q22309806](#) (Q22309806 P2093 "Wouter Schul" P1545 "7")
20. Processing [Q22309806](#) (Q22309806 P2093 "Feng Gu" P1545 "8")
21. Processing [Q22309806](#) (Q22309806 P2093 "Olivier Heudi" P1545 "9")
22. Processing [Q22309806](#) (Q22309806 P2093 "Ngai Ling Ma" P1545 "10")
23. Processing [Q22309806](#) (Q22309806 P2093 "Mee Kian Poh" P1545 "11")
24. Processing [Q22309806](#) (Q22309806 P2093 "Wai Yee Phong" P1545 "12")
25. Processing [Q22309806](#) (Q22309806 P2093 "Thomas H Keller" P1545 "13")
26. Processing [Q22309806](#) (Q22309806 P2093 "Edgar Jacoby" P1545 "14")
27. Processing [Q22309806](#) (Q22309806 P2093 "Subhash G Vasudevan" P1545 "15")

---

All done!.

Later, I can use the new Q-code ([Q22309806](#)) to use as source for statements about the compound (formula, etc).

### Draw your compound and get an InChIKey:

The next step is to draw a compound and get an InChIKey. This can be done with many tools, including [Bioclipse](#). Rajarshi opted for alternatives:

- [@collabchem](#) [@egonwillighagen](#) OSRA or <https://t.co/ZIQdgrYsmr?>  
— Rajarshi Guha (@rguha) [January 27, 2016](#)

Then check if the compound is not already in Wikidata. You can use this SPARQL query for that using the InChIKey of the compound (it's for acetic acid, so it will be found):

The screenshot shows the Wikidata Query Service interface. At the top, there's a header with 'Wikidata Query Service', a 'Beta' badge, and navigation links for 'Prefixes', 'Tools', and 'Help'. A red 'Examples' button is on the right. The main area contains a SPARQL query:

```

1 PREFIX wdt: <http://www.wikidata.org/prop/direct/>
2 SELECT ?compound WHERE {
3   ?compound wdt:P235 "QTBSBXVTEAMEQO-UHFFFAOYSA-N" .
4 }
5

```

Below the query, a green status bar indicates: 'Press [CTRL-SPACE] to activate auto completion. Data last updated: 6:33:21 PM GMT+1, Jan 27, 2016'. The results section shows '1 Results in 203 ms'. A red 'Execute' button, a 'Clear' button, and 'Download' and 'Link' icons are present. The results table has a header 'compound' and one row with the value 'wd:Q47512'.

For convenience, here the copy/pastable SPARQL:

```

PREFIX wdt:
SELECT ?compound WHERE {
  ?compound wdt:P235 "QTBSBXVTEAMEQO-UHFFFAOYSA-N" .
}

```

### Entering the compound:

So, the compound is not already in Wikidata, so time to add it. The minimal information you should provide is the following:

- mark the new entry as 'instance of' (P) 'chemical compound' (Q)
- the chemical formula and SMILES (use as reference the paper)
  - add the reference to the paper you entered above
- add the InChIKey and/or InChI

The first step is to create a new Wikidata entry. The Create new item menu in the left side panel can be used, showing a page like this:

## Create a new item

---

Please make sure that your item complies with the [notability policy](#).

**Note:** Don't create an item for your userpage. It *will* be deleted.

**Make sure the item does not already exist!** (If you make a mistake, you can request your item's deletion [here](#).)

You should create a [label](#) and a [description](#) for all new items. The first letter of your label should be capitalized only if it is a proper noun, and your description should *not* be phrased as a sentence.

By clicking "Create", you agree to the [terms of use](#), and you irrevocably agree to release your contribution under the [Creative Commons CC0 License](#).

**Create a new item**

Language:

English ▼

Label:

enter a label in English

Description:

enter a description in English

Aliases, pipe-separated:

enter some aliases in English

Create

As a label you can use the name used in the paper for the compound, even if a code, and as description 'chemical compound' will do for now; it can be changed later.

Feel free to add as much information about the compound as you can find. There are some chemically rich entries in Wikidata, such as that for acetic acid ([Q47512](#)).