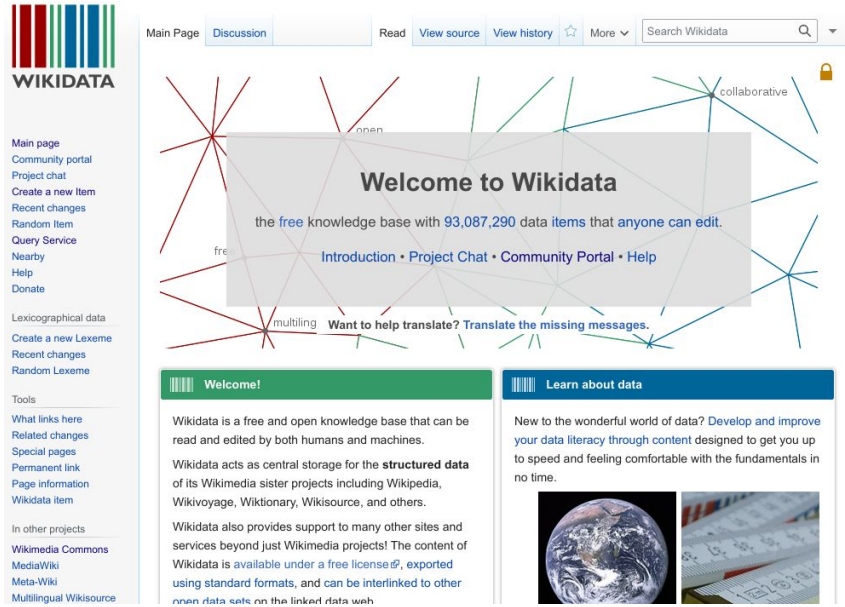
InChI and InChIKey in Wikidata and Scholia

Egon Willighagen
NIH Virtual Workshop on InChI
March 22-24, 2021

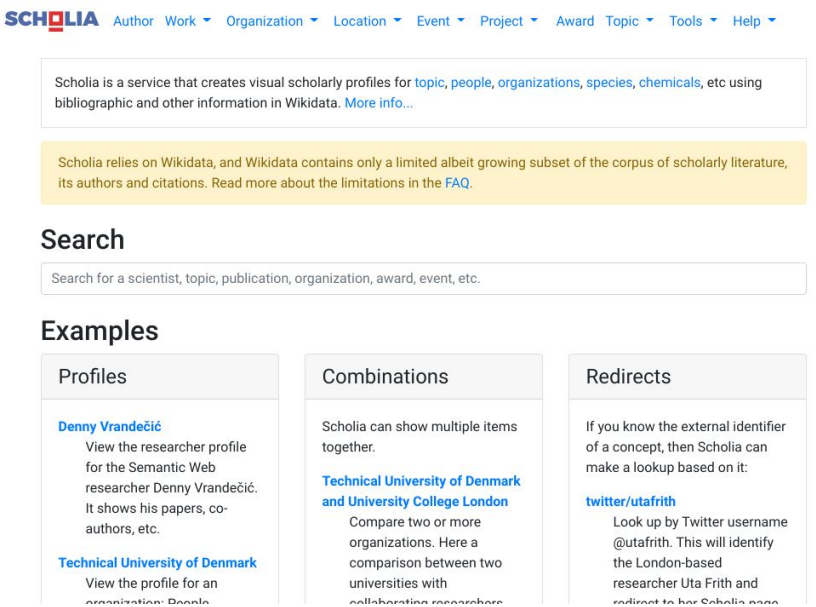
@egonwillighagen
0000-0001-7542-0286

NUTRIM
School of Nutrition and Translational Research in Metabolism
Maastricht University | Maastricht UMC+

Wikidata and Scholia



wikidata.org



scholia.toolforge.org

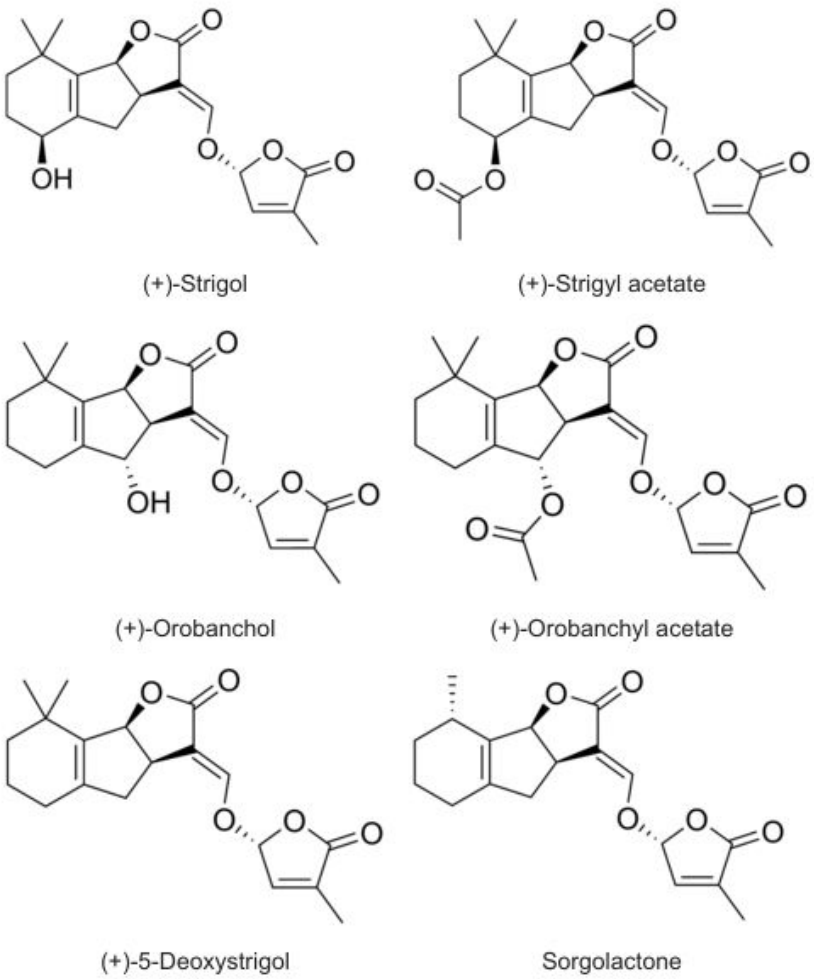
NUTRIM
School of Nutrition and Translational Research in Metabolism

Maastricht University | Maastricht UMC+

Strigolactones (in Wikipedia and Wikidata?)

Chemical structures [edit]

Some examples of strigolactones include:



strigolactones (Q2157332)

Strigolactones are a group of chemical compounds produced by a plant's roots. Due to their mechanism of actio
plant hormones or phytohormones. So far, strigolactones have been identified to be responsible for three differer
promote the germination of parasitic organisms that grow in the host plant's roots, such as Striga lutea and othe
English Wikipedia)

Class Hierarchy

WikiProject Chemistry

WIKIDATA

Project page | Discussion | Read | Edit | View history | More | Search Wikidata

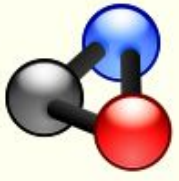
Wikidata:WikiProject Chemistry

Translate this page

Other languages: Deutsch • **English** • Nederlands • català • dansk • español • français • polski • português do Brasil • svenska • čeština • русский • українська • հայերեն • اردو • العربية • 中文

Home | **Guidelines** | **Properties** | **References** | **Tools**

Welcome to WikiProject Chemistry



Contents [hide]

Goals [edit]

- Define properties for items related to chemistry and the rules of use for these properties (qualifiers, datatypes, ...)
- Define references policy and especially ranking for references in order to ensure a high quality for chemical related data

Main page
Community portal
Project chat
Create a new Item
Recent changes
Random Item
Query Service
Nearby
Help
Donate

Lexicographical data
Create a new Lexeme
Recent changes
Random Lexeme

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Wikidata item

In other projects
Wikimedia Commons

WikiProject Chemistry

English
Español
Français
日本語
Polski
Svenska
Türkçe

22 more

Edit links

Custom tools

Participants [edit]

The participants listed below can be notified using the following template in discussions:

[+ Add yourself to the list]

```
{{Ping project|Chemistry}}
```

- Saehrimnir
- Leyo
- Snipre
- Jasper Deng
- Dcirovic
- Walkerma
- Egon Willighagen
- Denise Slenter
- Daniel Mietchen
- Kopiersperre
- Emily Temple-Wood
- Pablo Busatto (*Almondega*)
- Antony Williams (EPA)
- TomT0m
- Wostr
- Devon Fyson
- User:DePiep
- User:DavRosen
- Benjaminabel
- 99of9
- Kubaello
- Fractaler
- Sebotic
- Netha
- Hugo
- Samuel Clark
- Tris T7
- Leiem
- Christianhauck
- SCIdude
- Binter
- Photocyte
- Robert Giessmann
- Cord Wiljes
- Jonathan Bisson
- GrndStt
- Ameisenigel
- Charles Tapley Hoyt
- ChemHobby
- Peter Murray-Rust
- Erfurth

Bots [edit]

- SamoaBot - task 6 - set a property "atomic number" based on Wikipedia) - **On hold**

WikiProject Chemistry

Wikidata talk:WikiProject Chemistry

Old discussions are archived in Archive 2013, Archive 2014, Archive 2015, Archive 2016, Archive 2017, Archive 2018, Archive 2019.

Main page
Community portal
Project chat
Create a new Item
Recent changes
Random Item
Query Service
Nearby
Help
Donate

Lexicographical data
Create a new Lexeme
Recent changes
Random Lexeme

Tools
What links here
Related changes
Special pages
Permanent link
Page information

Custom tools

Contents [hide]

Modelling the Chemistry in Wikidata

NUTRIM
School of Nutrition and Translational Research in Metabolism
Maastricht University
Maastricht UMC+

WIKIDATA

Item | Discussion | Read | View history | More | Search Wikidata

acetic acid (Q47512)

chemical compound

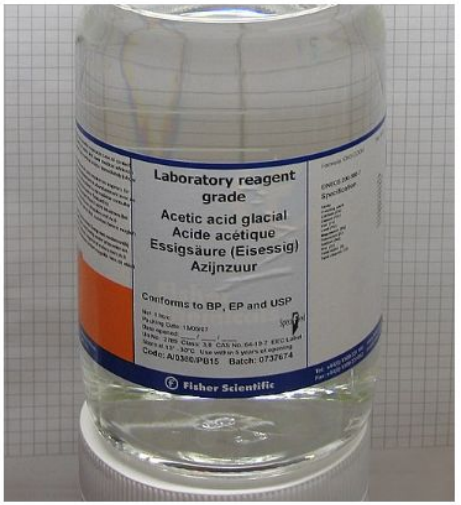
ethanoic acid | methanecarboxylic acid | CH3-COOH | Acetic acid, glacial | HOAc | Vinegar | Essigsäure | Glacial acetic acid | Ethanoate | acide acétique | Ethylic acid | Ethoic acid | Methanecarboxylic acid | Aceticum acidum | Ethanoic acid | Acetic acid | Ethanoat | E260 | CH3COOH

edit

▸ Recoin: Most relevant properties which are absent

▾ In more languages

Language	Label	Description	Also known as
English	acetic acid	chemical compound	ethanoic acid methanecarboxylic acid CH3-COOH Acetic acid, glacial HOAc Vinegar Essigsäure Glacial acetic acid Ethanoate acide acétique Ethylic acid Ethoic acid Methanecarboxylic acid Aceticum acidum Ethanoic acid Acetic acid Ethanoat E260 CH3COOH
German	Essigsäure	Ethansäure, einprotonige Carbonsäure	Haushaltsessig E260 Methancarbonsäure Acidum aceticum



Wikipedia (85 entries) edit [move]

af	Asynsuur
ar	حمض الخليك
ast	Ácidu acético
azb	استیک اسید
az	Sirkə turşusu
bcl	Asidong asetiko
be	Воцатная кіслата
bg	Оцетна киселина

Main page
Community portal
Project chat
Create a new Item
Recent changes
Random Item
Query Service
Nearby
Help
Donate

Lexicographical data
Create a new Lexeme
Recent changes
Random Lexeme

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Cite this page
Concept URI
Reasonator
Referee

Custom tools
List subclasses
List instances

Typing: chemical compound and more

Statements

instance of

by F705i and James Hare (NIOSH) and ProteinBoxBot and Egon Willighagen and Chire and Antoni Salvà and Thomas11 and Infovarius

- carboxylic acid — edit
  - 0 references
  - + add reference
- Class II combustible liquid — edit
  - 1 reference
- medication — edit
  - 1 reference

stated in	DrugBank
DrugBank ID	03166
language of work or name	English
title	Acetic acid (English)
publication date	17 November 2015

  - + add reference
- metabolite — edit

Chemical structure

chemical formula by ProteinBoxBot and Ivan A. Krestinin and The chemistds and Infovarius and Sebotic and SoCalChemBot and 87.68.189.9 and Wostr	$C_2H_4O_2$ ▸ 1 reference	edit + add value
canonical SMILES by ProteinBoxBot and Sebotic and SoCalChemBot	CC(=O)O ▸ 1 reference	edit + add value
InChI by Happy5214 and KrBot and ProteinBoxBot and Sebotic and SoCalChemBot and Scidudebot	InChI=1S/C2H4O2/c1-2(3)4/h1H3,(H,3,4) ▾ 1 reference stated in: PubChem PubChem CID: 176 language of work or name: English title: acetic acid (English) retrieved: 19 October 2016	edit + add reference + add value
InChIKey by Happy5214 and KrBot and ProteinBoxBot and Sebotic and SoCalChemBot	QTBSBXVTEAMEQO-UHFFFAOYSA-N ▸ 2 references	edit + add value

CDK DEPICT

astricht UMC+

Physicochemical properties

boiling point

by Emily Temple-Wood (NIOSH) and James Hare (NIOSH) and Egon Willighagen and Wostr

244±1 degree Fahrenheit — edit

pressure	760±1 torr

▾ 1 reference

reference URL	http://www.cdc.gov/niosh/npg/npgd0002.html

+ add reference

117.9±0.1 degree Celsius — edit

▾ 1 reference

stated in	Basic laboratory and industrial chemicals: a CRC quick reference handbook

+ add reference

117.9±0.2 degree Celsius — edit

pressure	101,325±1 pascal

▾ 1 reference

stated in	CRC Handbook of Chemistry and Physics (95th edition)
page(s)	3-4

+ add reference

+ add value

Wikidata Query Service | Examples | Help | More tools | English

```
SELECT ?typeLabel ?count WITH {
  SELECT ?type (COUNT(DISTINCT ?chemical) AS ?count) WHERE {
    ?chemical wdt:P31 ?type ;
              wdt:P235 [] .
  } GROUP BY ?type
} AS %TYPES {
  INCLUDE %TYPES
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
} ORDER BY DESC(?count)
```

2443 results in 30536 ms | Code | Download | Link

Search

typeLabel	count
chemical compound	1135341
group of stereoisomers	111523
chemical entity	14665
medication	2356
organic anion	826
carcinogen	487

cht UMC+

Chemical Types

typeLabel	count
chemical compound	1135341
group of stereoisomers	111523
chemical entity	14665
medication	2356
organic anion	826
carcinogen	487
diacylglycerophosphocholine	485
wax monoester	409
pair of enantiomers	409
intermetallic	376
lipid	342
flavonoid	342
essential medicine	302
heterocyclic compound	275
unsaturated fatty acids	263
fatty acyl-CoA	254
developmental toxicant	253
carboxylic acid	246
diacylglycerophosphoinositols	239
insecticide	212
herbicide	205

typeLabel	count
chemical compound	917519
group of stereoisomers	9647
medication	2656
intermetallic	677
carcinogen	495
chemical entity	459
wax monoester	393
pair of enantiomers	357
essential medicine	327
monoclonal antibody	325
flavonoid	292
heterocyclic compound	274
developmental toxicant	262
carboxylic acid	238
insecticide	217
family of isomeric compounds	207
herbicide	205
mixture	192
biopharmaceutical	173
unsaturated fatty acids	164
fungicide	162

Chemical type guidance

Compounds without fully defined isomerism or isotopic composition [edit]

In the following rules *structural need* is defined as having at least one external identifier to a reliable database (InChI, InChIKey or SMILES are not regarded as such) or at least one valid sitelink to a page on a Wikimedia project (cf. point 1 of Wikidata:Notability).

See discussion about this topic (2018) in WikiProject Chemistry.

Inclusion criteria [edit]

1. Every chemical compound with fully defined isomerism (*cis–trans* isomerism; *ortho*, *meta*, *para* isomerism; enantiomerism; etc.) or isotopic composition can be described in a separate item (hereafter **item A**).	• (S)-2-pentanol (Q20680358) describes a compound with fully defined stereochemistry (has one stereogenic centre and it is defined). Thus, it can be described in a separate item. • 2-Bromobenzaldehyde (Q33859440) describes a compound with defined positions of two substitutents of the benzene ring (*ortho* position), so it can be described in a separate item.
2. Chemical compound with fully (**item B**) or partially (**item C**) undefined isomerism or isotopic composition can be described in a separate item if it fulfils some structural need.	• DL-methamphetamine (Q44909815) describes a compound with fully undefined stereochemistry (has one stereogenic centre and it is undefined). However, it has some external identifiers in reliable databases, like ChemSpider ID (P661) or DSSTox substance ID (P3117), so it can be described in a separate item. • (2S)-homocystine (Q27161892) describes a compound with partially undefined stereochemistry (has two stereogenic centres and only one is defined). It has some external identifiers like ChEBI ID (P683) or PubChem CID (P662), so it can be described in a separate item. • Bromobenzaldehyde (Q33859433) describes a compound with undefined position of two substitutents of the benzene ring (i.e. it is a group of three isomers: *ortho*, *meta* and *para*). It has an article in Wikipedia, so it can be described in a separate item.
3. Item about a racemic mixture (Q467717) (or more generally: about mixture of isomers) can be described in a separate item (**item D**) if it fulfils some structural need.	• (±)-nicotine (Q56697247) describes a racemic mixture of (−)-nicotine (Q28086552) and (+)-nicotine (Q27119762) and it is different from nicotine (Q12144) that describes a compound with undefined stereochemistry. It can be described in a separate item, because there is an external identifier (ChEBI ID (P683)).
4. Item about a compound being an atropisomer (Q757764) can be described in a separate item if it can be isolated or if it fulfils some structural need.	• *no example yet*

WikiShape Entity Schema Property Query Help

Visualize Wikidata Schema

racemic mixture

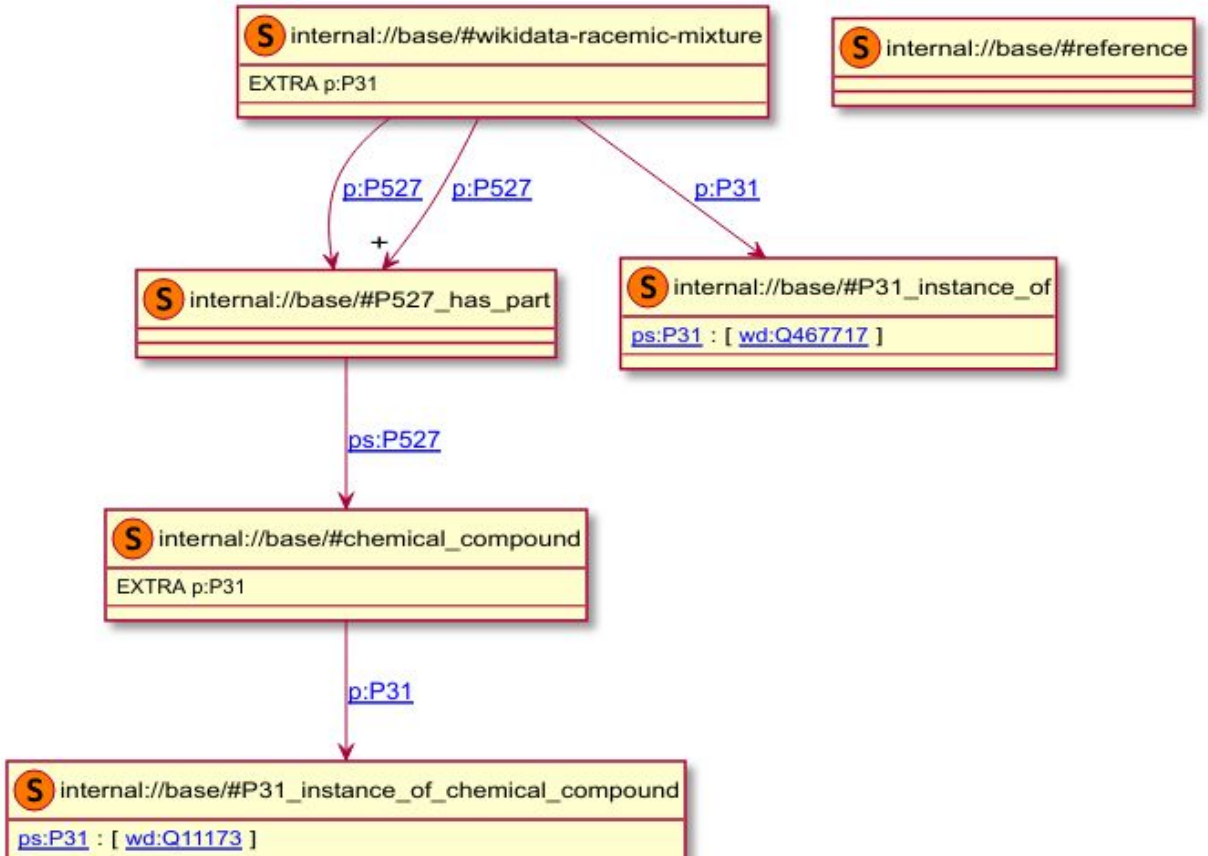
Language en

Info about schema entity

E47 - racemic mixture

mixture of chemicals with the same structure but different stereochemistry

https://www.wikidata.org/wiki/EntitySchema:E47



Methodology article | Open Access | Published: 22 January 2021

A protocol for adding knowledge to Wikidata: aligning resources on human coronaviruses

Andra Waagmeester, Egon L. Willighagen, Andrew I. Su, Martina Kutmon, Jose Emilio Labra Gayo, Daniel Fernández-Álvarez, Quentin Groom, Peter J. Schaap, Lisa M. Verhagen & Jasper J. Koehorst

BMC Biology **19**, Article number: 12 (2021) | Cite this article

Abstract

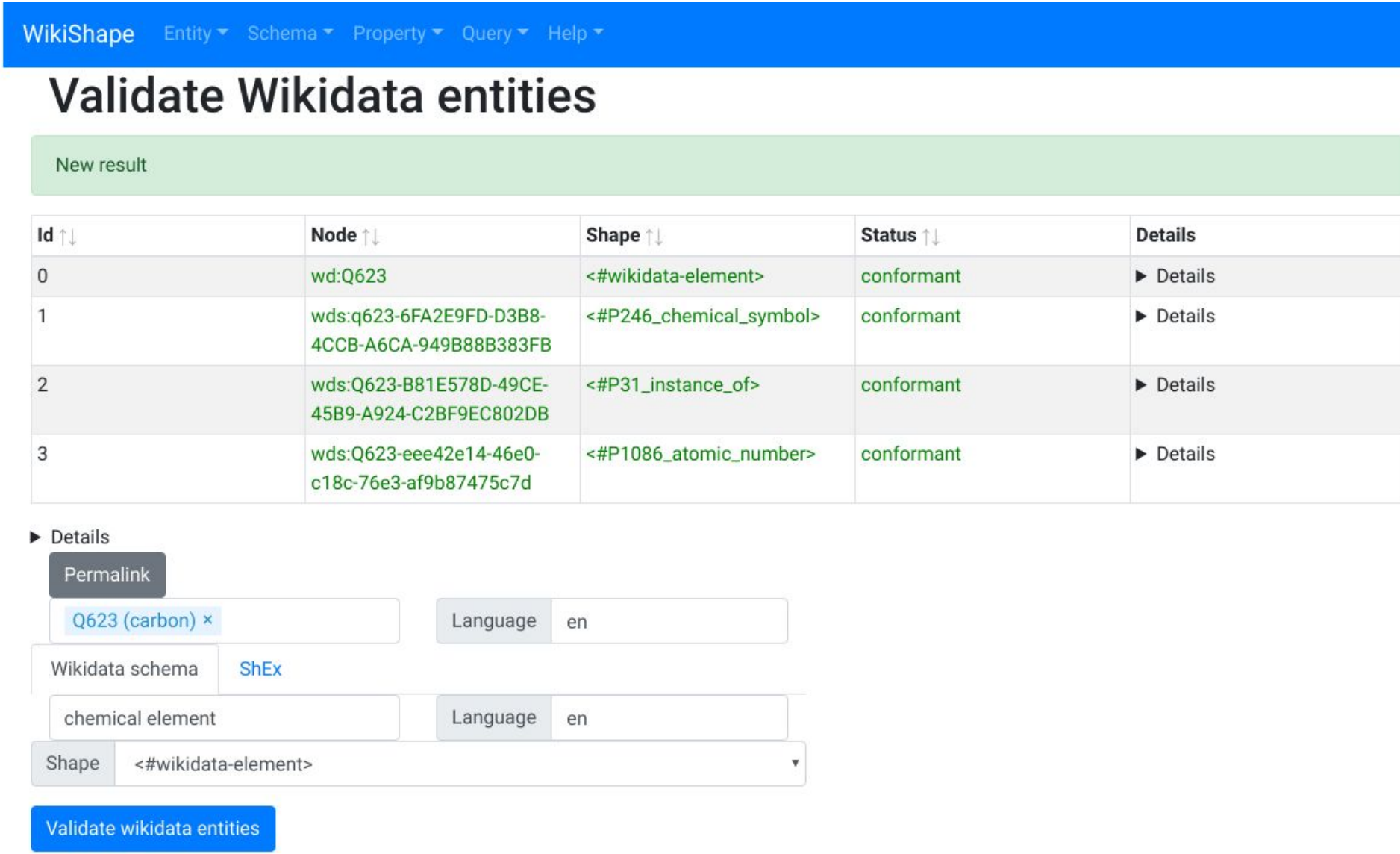
Background

Pandemics, even more than other medical problems, require swift integration of knowledge. When caused by a new virus, understanding the underlying biology may help finding solutions. In a setting where there are a large number of loosely related projects and initiatives, we need common ground, also known as a "commons." Wikidata, a public knowledge graph aligned with Wikipedia, is such a commons and uses unique identifiers to link knowledge in other knowledge bases. However, Wikidata may not always have the right schema for the urgent questions. In this paper, we address this problem by showing how a data schema required for the integration can be modeled with entity schemas represented by Shape Expressions.

Results

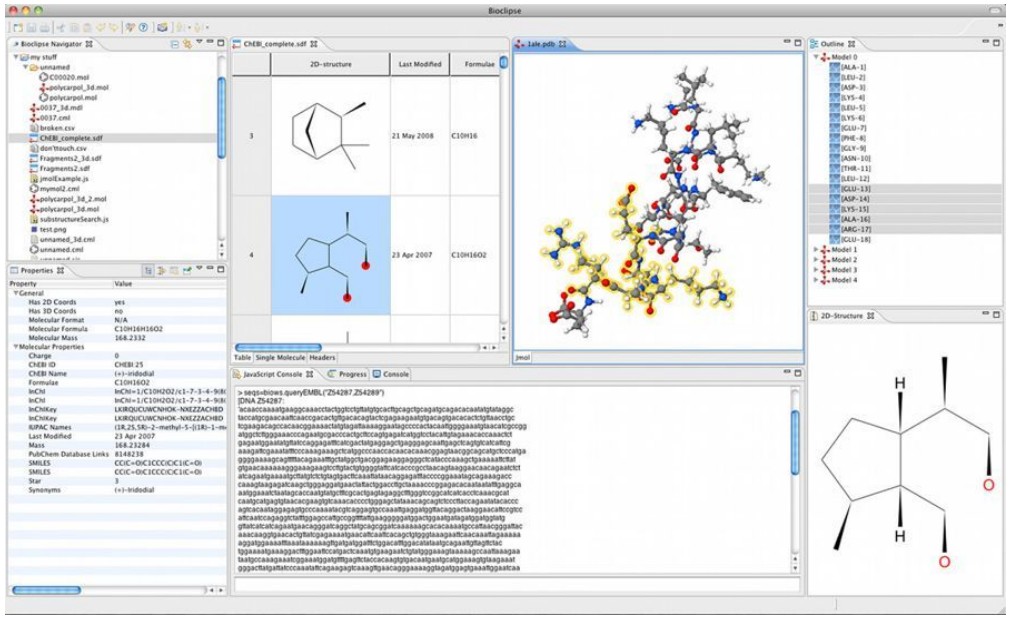
As a telling example, we describe the process of aligning resources on the genomes and proteomes of the SARS-CoV-2 virus and related viruses as well as how Shape Expressions can be defined for Wikidata to model the knowledge, helping others studying the SARS-CoV-2 pandemic. How this model can be used to make data between various resources interoperable

ShEx validation: E46 → chemical element

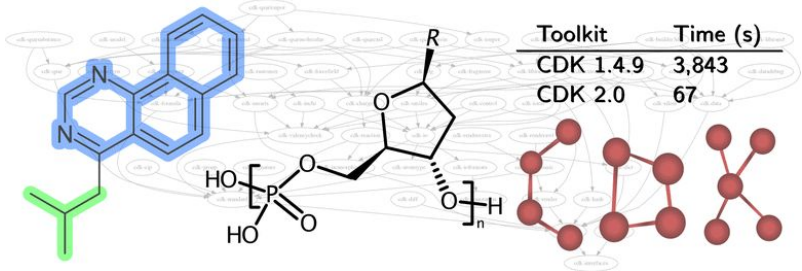


Adding chemical compounds to Wikidata

Workhorse: Bioclipse scripts + the CDK



10.1186/1471-2105-8-59,
10.1186/1471-2105-10-397



Toolkit	Time (s)
CDK 1.4.9	3,843
CDK 2.0	67

Journal of Cheminformatics

Home About Articles Submission Guidelines About The Editors Calls For Papers

Software | Open Access | Published: 06 June 2017

The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching

Egon L. Willighagen, John W. Mayfield, Jonathan Alvarsson, Arvid Berg, Lars Carlsson, Nina Jeliazkova, Stefan Kuhn, Tomáš Pluskal, Miquel Rojas-Chertó, Ola Spjuth, Gilleain Torrance, Chris T. Evelo, Rajarshi Guha & Christoph Steinbeck

Journal of Cheminformatics 9, Article number: 33 (2017) | Download Citation

Bacting: Bioclipse on the command line

```
@Grab(group='io.github.egonw.bacting', module='managers-cdk', version='0.0.9')

workspaceRoot = "."
def cdk = new net.bioclipse.managers.CDKManager(workspaceRoot);

println cdk.fromSMILES("COC")
```

- Wikicite/findConcepts.groovy
- Wikidata/createWDitemsFromSMILES.groovy
- LipidMaps/classifyLipids.groovy
- ExtIdentifiers/comptox.groovy
- MeltingPoints/createQuickStatements.groovy
- ...

https://github.com/egonw/bacting
https://github.com/egonw/ons-wikidata

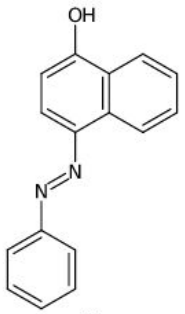
CDK DEPICT

NEXTMOVE SOFTWARE

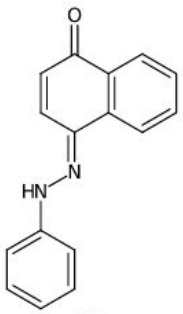
Generate depictions of molecules and reactions from SMILES or SDF.

```
c1(c2cccc1)C(O)=CC=C2\N=N\c1ccccc1
c1(c2cccc1)C(=O)C=C\C2=N/Nc1ccccc1
c1(c2cccc1)C(O)=CC=C2N=Nc1ccccc1
c1(c2cccc1)C(=O)C=CC2=NNc1ccccc1
```

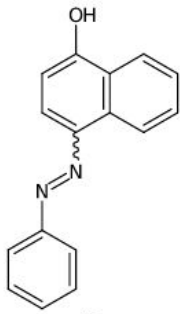
Black on Clear | No Annotation | Chiral Hydrogens (smart) | Do Not Abbreviate | Enter SMARTS pattern...



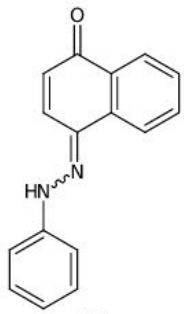
#1



#2



#3



#4

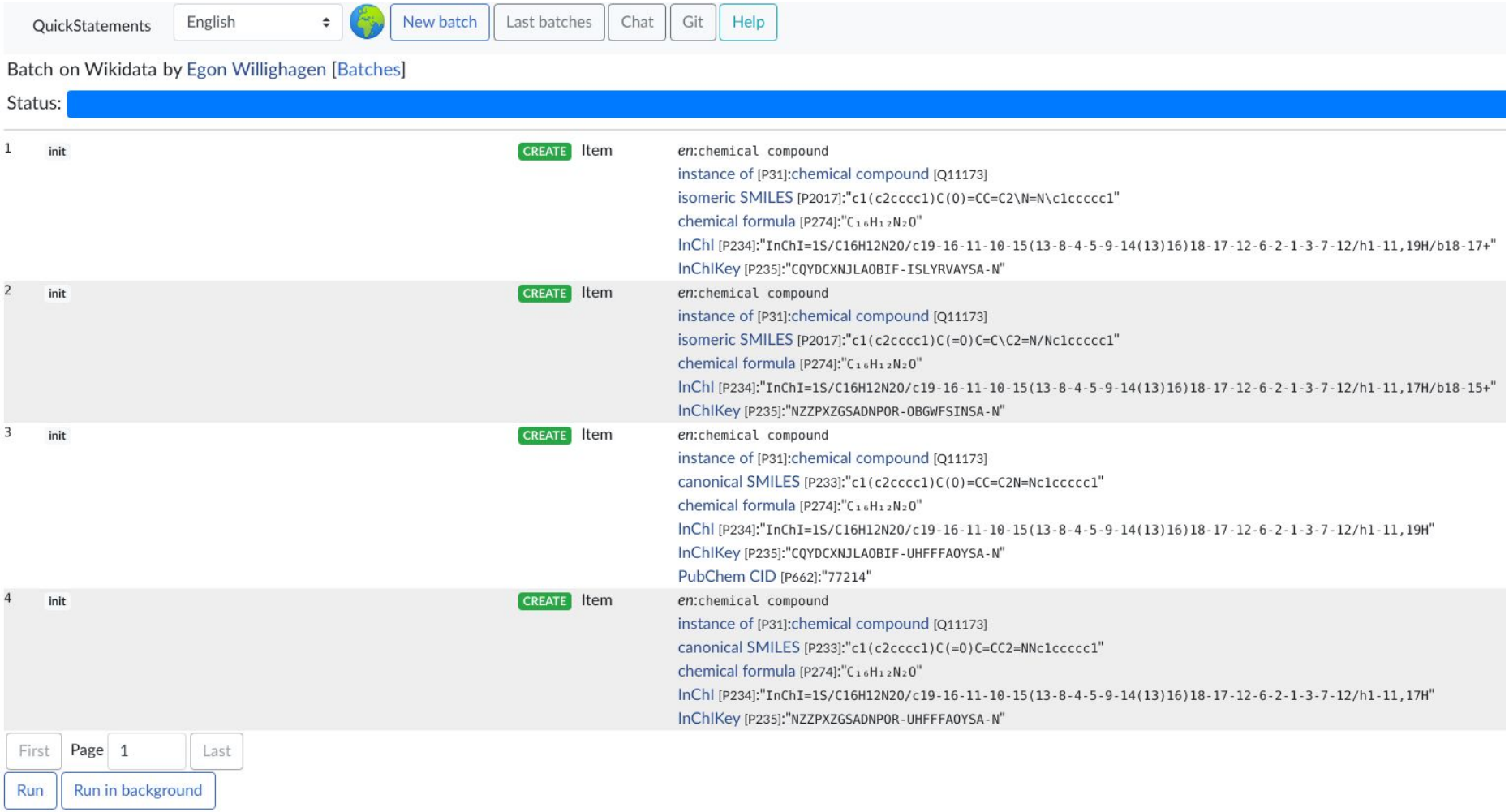
Built with the Chemistry Development Kit. Depict v1.6-SNAPSHOT, CDK v2.4-SNAPSHOT.

Compare against Wikidata (with InChIKey)

```
egonw@debian:~/var/Projects/hub/ons-wikidata/Wikidata$ groovy createWDitemsFromSMILES.groovy
====================
C16H12N2O is not yet in Wikidata
Full stereochemistry is defined
====================
====================
C16H12N2O is not yet in Wikidata
Full stereochemistry is defined
====================
====================
C16H12N2O is not yet in Wikidata
Compound has missing stereo on # of centers: 2
====================
====================
C16H12N2O is not yet in Wikidata
Compound has missing stereo on # of centers: 1
====================
```

```
egonw@debian:~/var/Projects/hub/ons-wikidata/Wikidata$ more output.quickstatements

      CREATE

      LAST      P31     Q11173
      LAST      Den     "chemical compound"
      LAST      P2017   "c1(c2cccc1)C(O)=CC=C2\N=N\c1ccccc1"
      LAST      P274    "C16H12N2O"
      LAST      P234    "InChI=1S/C16H12N2O/c19-16-11-10-15(13-8-4-5-9-14(13)16)18-17-12-6-2-
      LAST      P235    "CQYDCXNJLAOBIF-ISLYRVAYSA-N"



      CREATE

      LAST      P31     Q11173
      LAST      Den     "chemical compound"
```

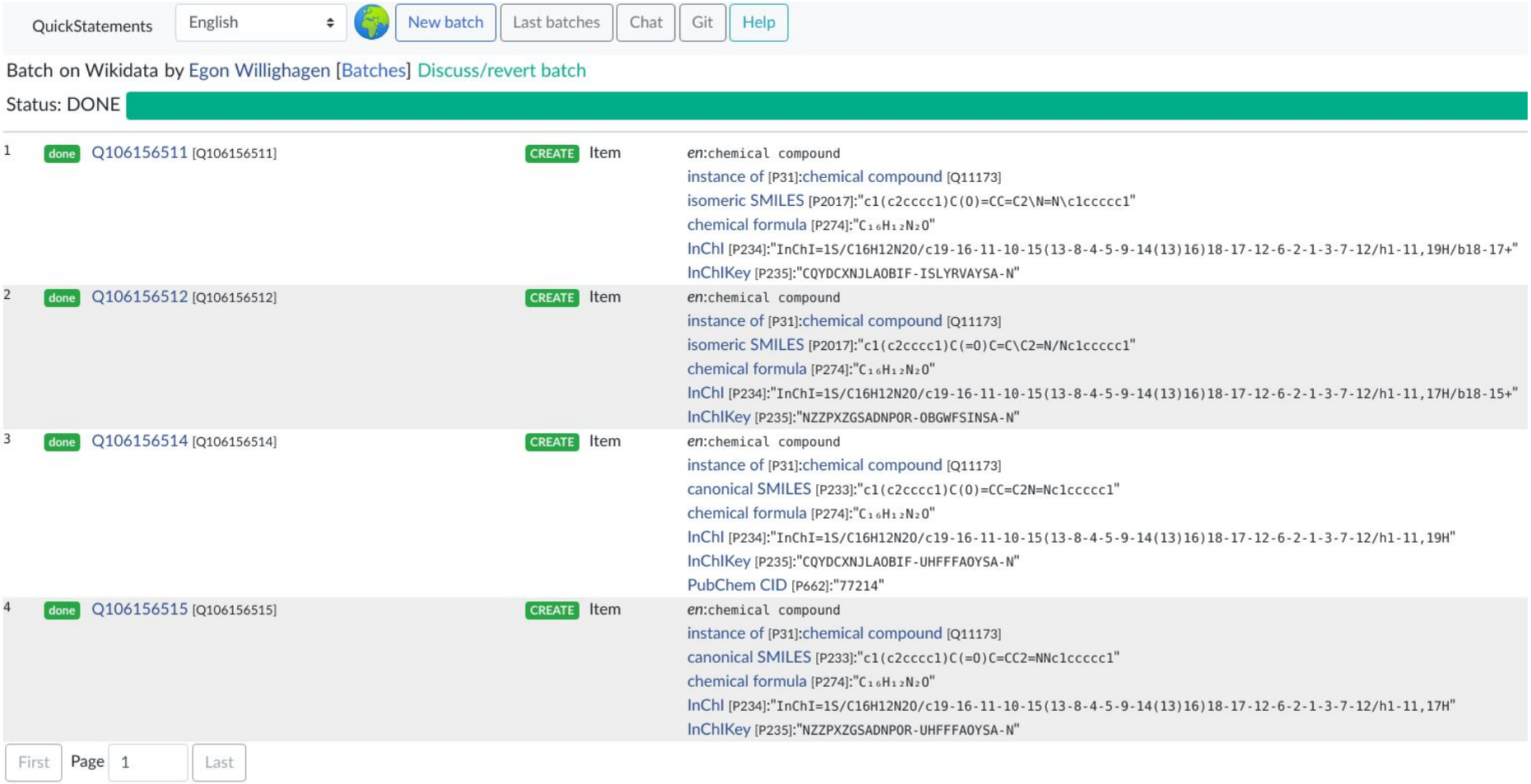
Use QuickStatements to add to Wikidata



NUTRIM
School of Nutrition and Translational Research in Metabolism

Maastricht University | Maastricht UMC+

Use QuickStatements to add to Wikidata



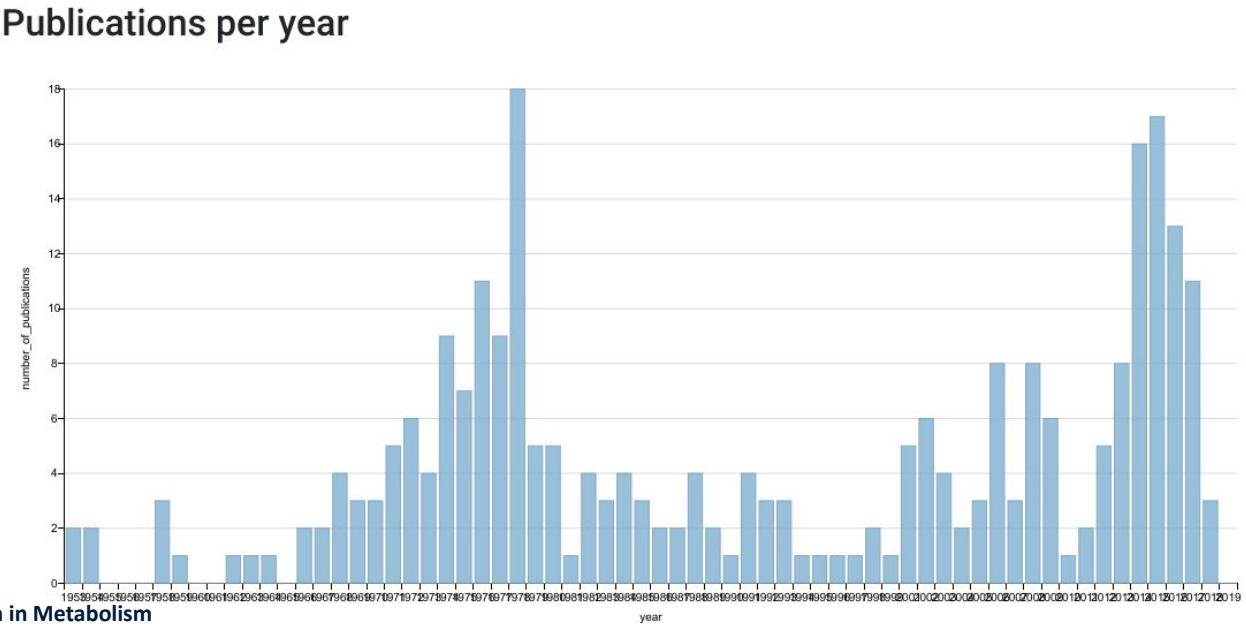
NUTRIM
School of Nutrition and Translational Research in Metabolism

Maastricht University | Maastricht UMC+

Wikidata Quickstatements v2

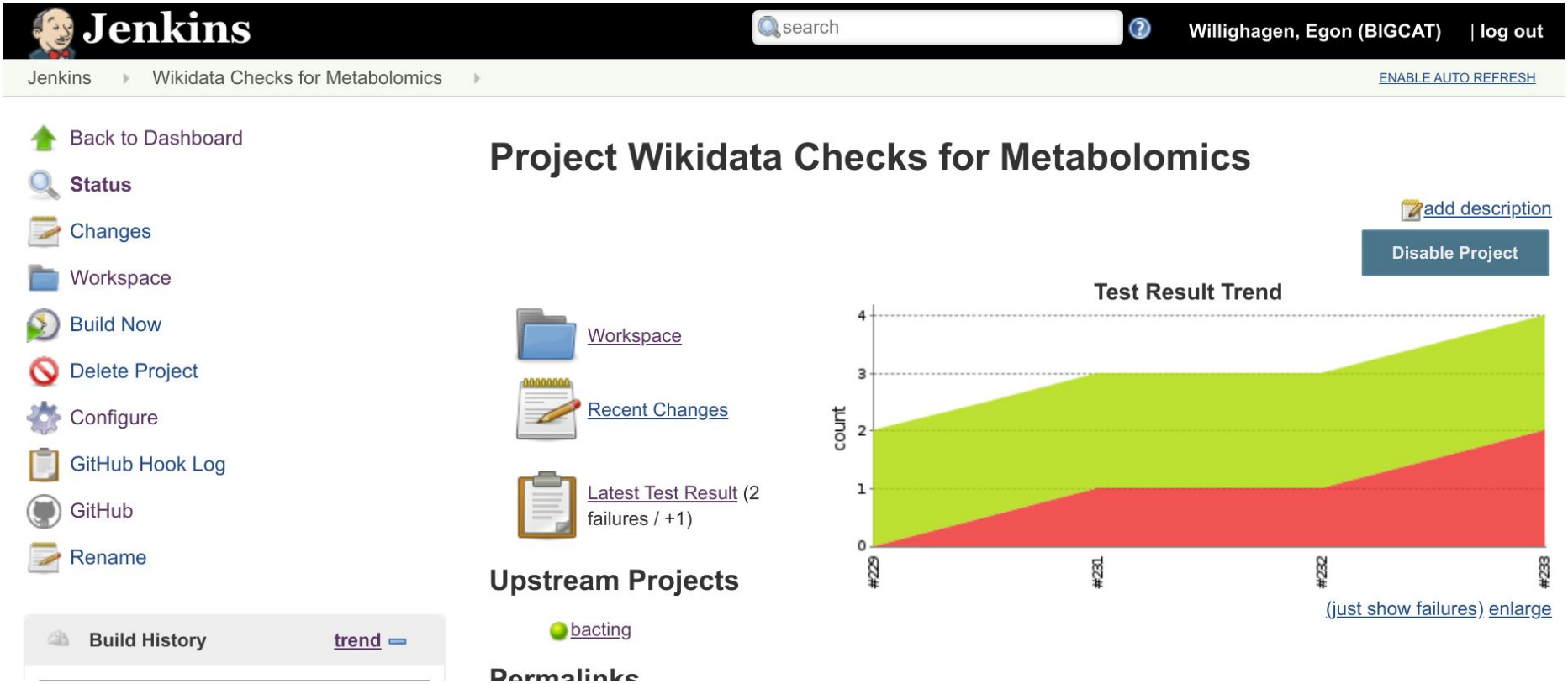
qid,P921,#
Q26801490,Q70828631,Activities and Effects of Ergot Alkaloids on ...
Q28082319,Q70828631,Diversification of ergot alkaloids in natural and ...
Q28214648,Q70828631,Biotechnology and genetics of ergot alkaloids
Q28276288,Q70828631,Ergot alkaloids--biology and molecular biology
Q28287164,Q70828631,Occurrence of peptide and clavine ergot alkaloids …
...

Dr. Magnus Manske
Sanger Institute



NUTRIM
School of Nutrition and Translational Research in Metabolism

Jenkins for Wikidata quality control



NUTRIM
School of Nutrition and Translational Research in Metabolism

Maastricht University | Maastricht UMC+

ChemCuration example: InChIKeys

Jenkins ▸ Wikidata Checks for Metabolomics ▸ #233 ▸ Test Results ▸ (root) ▸ InChITests ▸ InChIKeyMismatch ENABLE AUTO REFRESH

History
Git Build Data
No Tags
Test Result
Previous Build

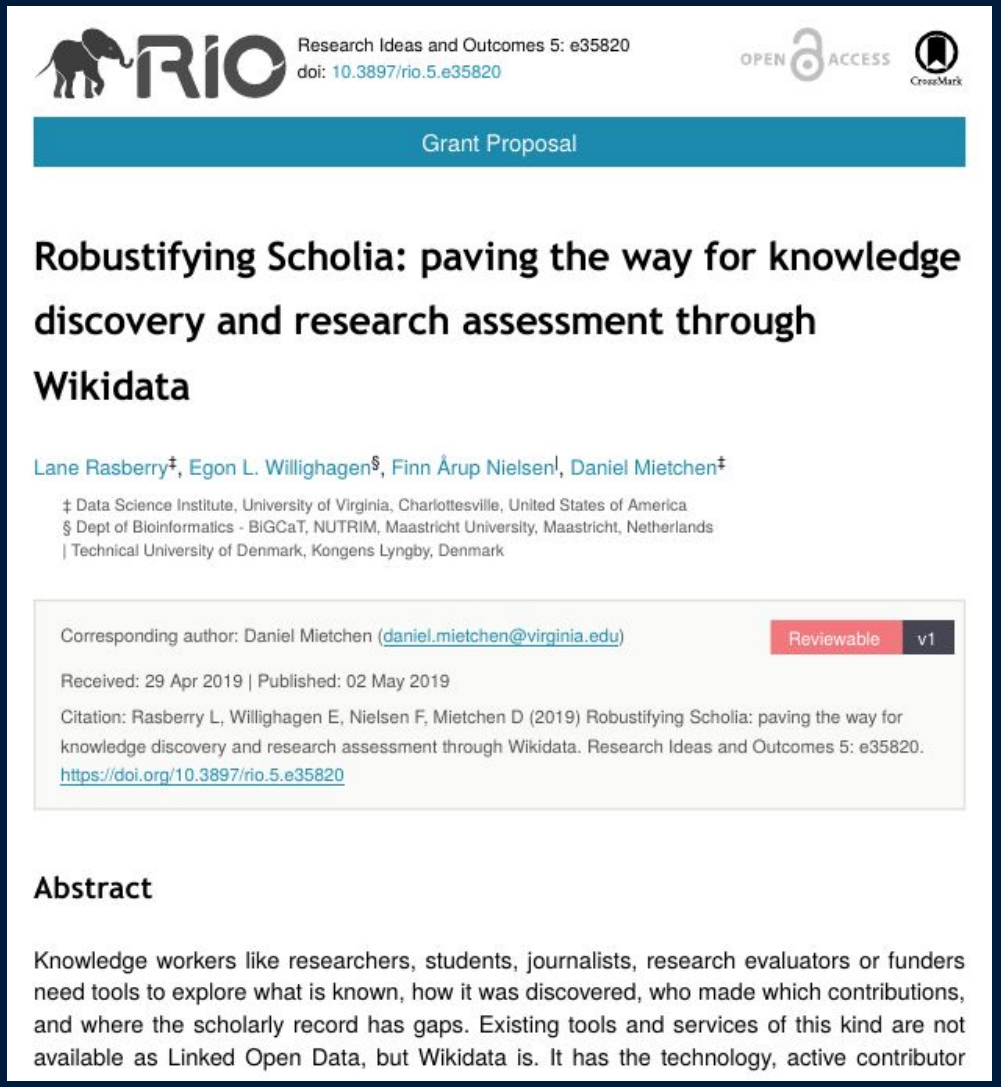
Error Message

The InChIKey computed from the isomeric SMILES and InChIKey in Wikidata does not match

Stacktrace

http://www.wikidata.org/entity/Q421291 with isomeric SMILES '[Fe+2].O[C@H]([C@@H](O)C([O-])=O)[C@H](O)[C@H](O)CO.[O-]C(=O)[C@H](O)[C@@H](O)[C@H](O)[C@H](O)CO' has a calculated InChIKey VRIVJOXICYMTAG-IYEMJOQQSA-L that does not match the given QDUZQOIJXPPTLY-GMBKLUGCSA-N
http://www.wikidata.org/entity/Q7777226 with isomeric SMILES 'Oc1cc(cc(O)c1O)C(=O)Oc5c(O)cc([C@H]2Oc3cc(O)cc(O)c3C[C@H]2O)c6\C=C(/C=C(/OC(=O)c4cc(O)c(O)c(O)c4)C(=O)c56)[C@H]7Oc8cc(O)cc(O)c8C[C@H]7O' has a calculated InChIKey FJYGFTHLNNSVPY-BBXLVSEPSA-N that does not match the given TUJOKWPTOVJHLY-JBJHRQGLSA-N
http://www.wikidata.org/entity/Q15427926 with isomeric SMILES 'CC1(C)C([C@@H](OC(C)=O)C[C@@]2(C)[C@](C([C@@H](OC(C)=O)CC2)=C)([H])[C@@H]3OC(C)=O)=C(C)[C@@H](OC([C@H](C)[C@H](C)O)=O)C[C@]13[H]' has a calculated InChIKey ULHOQEOTAJVICR-SJJKDWJASA-N that does not match the given FMPIEMVVEJGMCY-IRWPHOLZSA-N
http://www.wikidata.org/entity/Q568 with isomeric SMILES '[Li]' has a calculated InChIKey WHXSMMKQMYFTQS-UHFFFAOYSA-N that does not match the given SIAPCJWMELPYOE-UHFFFAOYSA-N
http://www.wikidata.org/entity/Q5278705 with isomeric SMILES 'C[C@@]13O[C@]1(/C=C/C(C)=C/C=C/C(C)=C/C=C/C=C(C)/C=C/C=C(C)/[C@@H]=C=C2C(C)(C)C[C@@H](OC(C)=O)C[C@]2(C)O)C(C)(C)C[C@H](O)C3' has a calculated InChIKey PVNVIBOWBAPFOE-RWNIHPGNSA-N that does not match the given GJFBHWJTMDTLNX-UWCSZFODSA-N

Scholia



Research Ideas and Outcomes 5: e35820
doi: 10.3897/rio.5.e35820

OPEN ACCESS

Grant Proposal

Robustifying Scholia: paving the way for knowledge discovery and research assessment through Wikidata

Lane Rasberry‡, Egon L. Willighagen§, Finn Årup Nielsen|, Daniel Mietchen‡

‡ Data Science Institute, University of Virginia, Charlottesville, United States of America
§ Dept of Bioinformatics - BiGCaT, NUTRIM, Maastricht University, Maastricht, Netherlands
| Technical University of Denmark, Kongens Lyngby, Denmark

Corresponding author: Daniel Mietchen (daniel.mietchen@virginia.edu)

Reviewable v1

Received: 29 Apr 2019 | Published: 02 May 2019

Citation: Rasberry L, Willighagen E, Nielsen F, Mietchen D (2019) Robustifying Scholia: paving the way for knowledge discovery and research assessment through Wikidata. Research Ideas and Outcomes 5: e35820. https://doi.org/10.3897/rio.5.e35820

Abstract

Knowledge workers like researchers, students, journalists, research evaluators or funders need tools to explore what is known, how it was discovered, who made which contributions, and where the scholarly record has gaps. Existing tools and services of this kind are not available as Linked Open Data, but Wikidata is. It has the technology, active contributor

Wikidata / Scholia

Redirecting

If you know the identifier then Scholia can make a lookup based on the identifier:

cas/50-00-0

Lookup CAS 50-00-0. This will identify formaldehyde and redirect to its Scholia page.

inchikey/QTBSBXVTEAMEQO-UHFFFAOYSA-N

Redirect also works for InChIKeys, here for acetic acid.

Show 10 entries Search:

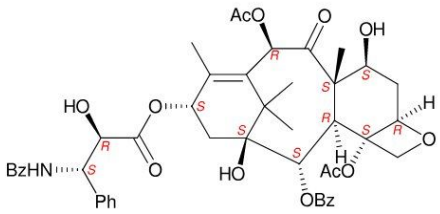
Mol	InChIKey	CAS	ChemSpider	PubChem CID
acetic acid	QTBSBXVTEAMEQO-UHFFFAOYSA-N	64-19-7	171	176
deuterated acetic acid	QTBSBXVTEAMEQO-GUEYOVJQSA-N	1186-52-3	2006083	2723903
acetic acid c-14	QTBSBXVTEAMEQO-HQMMCQRPSA-N	2845-03-6	144444	164769
acetic acid c-13	QTBSBXVTEAMEQO-VQEHIDDOSA-N	1563-79-7	8329490	10153982
acetic acid c-11	QTBSBXVTEAMEQO-JVVVGQRLSA-N	78887-71-5	396653	450349
acetate ion	QTBSBXVTEAMEQO-UHFFFAOYSA-M	71-50-1	170	175

Edit on query.Wikidata.org

Showing 1 to 6 of 6 entries Previous 1 Next

paclitaxel (Q423762)

Paclitaxel (PTX), sold under the brand name Taxol among others, is a chemotherapy medication used to treat a number of types of cancer. This includes ovarian cancer, breast cancer, lung cancer, Kaposi sarcoma, cervical cancer, and pancreatic cancer. It is given by injection into a vein. ... (from the English Wikipedia)



Identifiers

Show 10 entries Search:

IDpred	Id

2019: 10.3897/rio.5.e35820
2017: 10.6084/m9.figshare.6356027.v1

NUTRIM

UM SCHOLIA

InChIKey

Redirecting

If you know the identifier then Scholia can make a lookup based on the identifier:

cas/50-00-0

Lookup CAS 50-00-0. This will identify formaldehyde and redirect to its Scholia page.

inchikey/QTBSBXVTEAMEQO-UHFFFAOYSA-N

Redirect also works for InChIKeys, here for acetic acid.

Show 10 entries Search:

Mol	InChIKey	CAS	ChemSpider	PubChem CID
acetic acid	QTBSBXVTEAMEQO-UHFFFAOYSA-N	64-19-7	171	176
deuterated acetic acid	QTBSBXVTEAMEQO-GUEYOVJQSA-N	1186-52-3	2006083	2723903
acetic acid c-14	QTBSBXVTEAMEQO-HQMMCQRPSA-N	2845-03-6	144444	164769
acetic acid c-13	QTBSBXVTEAMEQO-VQEHIDDOSA-N	1563-79-7	8329490	10153982
acetic acid c-11	QTBSBXVTEAMEQO-JVVVGQRLSA-N	78887-71-5	396653	450349
acetate ion	QTBSBXVTEAMEQO-UHFFFAOYSA-M	71-50-1	170	175

Edit on query.Wikidata.org

Showing 1 to 6 of 6 entries Previous 1 Next

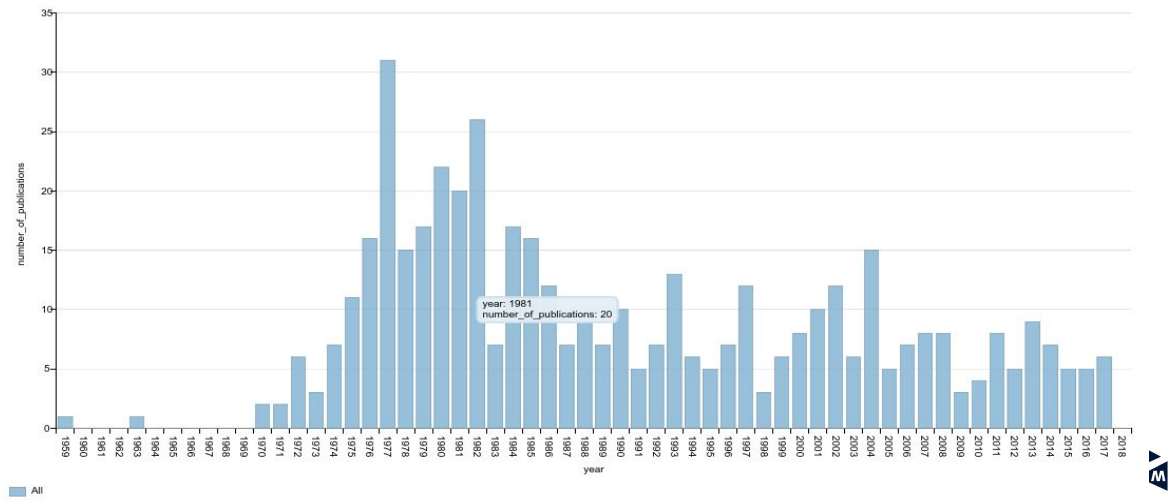
Wikidata / Scholia

Physchem Properties

Show 10 entries Search:

PropEntity	Value	Units	Qualifiers	Source	Doi
acid dissociation constant	4.74	1		Small Scale Determination of the pKa Values for Organic Acids	10.1021/ED071PA6
mass	60.021129	atomic mass unit		PubChem	
acid dissociation constant	4.756	1	temperature: 25	CRC Handbook of Chemistry and Physics (95th edition)	
boiling point	117.9	degrees Celsius	pressure: 101325	CRC Handbook of Chemistry and Physics (95th edition)	
density	1.0446	gram per cubic centimetre	temperature: 25	CRC Handbook of Chemistry and Physics (95th edition)	

Publications per year



Recently published works on the chemical

Show 10 entries

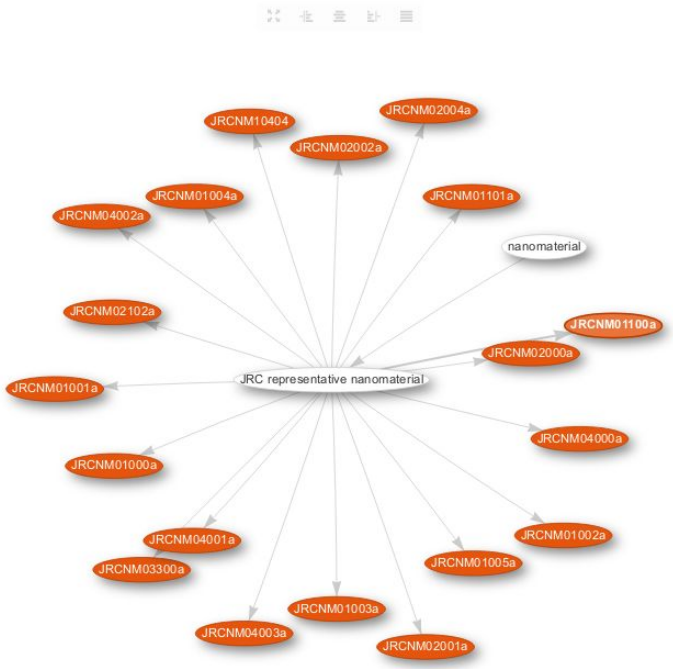
Date	Work	Type	Topics
2017-08-09	In vitro human skin permeation of benzene in gasoline: effects of concentration, multiple dosing and skin preparation	scholarly article	oil and gas extraction // benzene
2017-04-27	Nicotine, aerosol particles, carbonyls and volatile organic compounds in tobacco- and menthol-flavored e-cigarettes	scholarly article	toluene // benzene

Scholia: JRC representative industrial nanomaterials

topic | chemical

JRC representative nanomaterial (Q47461491)

Class Hierarchy



Recently published works on the chemical RSS

Show 10 entries Search:

Date	Work	Type	Topics
2017-09-28	Fish cell lines as a tool for the ecotoxicity assessment and ranking of engineered nanomaterials.	scholarly article	JRCNM02000a // JRCNM04000a // JRCNM01101a // JRCNM01100a // JRCNM02102a // nanomaterial // toxicology
2017-06-01	Graphistrength© C100 MultiWalled Carbon Nanotubes (MWCNT): thirteen-week inhalation toxicity study in rats with 13- and 52-week recovery periods combined with comet and micronucleus assays	scholarly article	JRCNM04002a // Brown Rat // toxicology
2017-05-19	Elucidating the Role of Dissolution in CeO2 Nanoparticle Plant Uptake by Smart Radiolabeling.	scholarly article	JRCNM02102a // general chemistry // catalysis // nanoparticle
2017-04-05	Multi-walled carbon nanotube-physicochemical properties predict the systemic acute phase response following pulmonary exposure in mice.	scholarly article	JRCNM04003a // JRCNM04001a // JRCNM04000a // carbon nanotube
2017-01-03	Negligible cytotoxicity induced by different titanium dioxide nanoparticles in fish cell lines.	scholarly article	JRCNM01005a // JRCNM01004a // JRCNM01003a
2016-11-01	The JRC Nanomaterials Repository: A unique facility providing representative test materials for nanoEHS research	scholarly article	JRC representative nanomaterial // Directorate-General for Joint Research Centre // nanomaterial // toxicology
2015-11-12	Towards the standardization of nanoecotoxicity testing: Natural organic matter 'camouflages' the adverse effects of TiO2 and CeO2 nanoparticles on green microalgae.	scholarly article	JRCNM02102a // JRCNM01003a

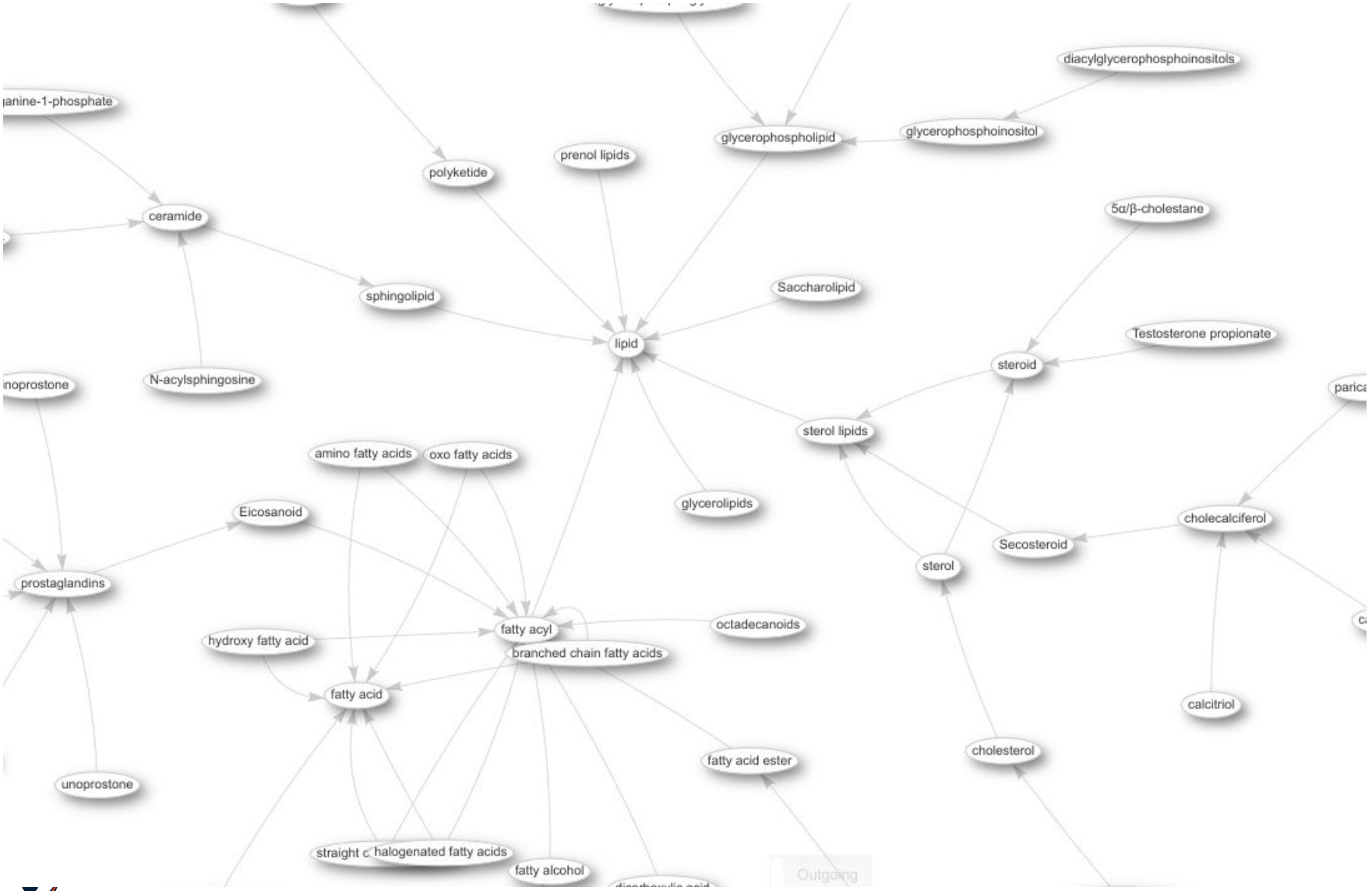
NUTRIM
School of Nutrition and Translational Research in Metabolism

SCHOLIA

wikicite

Maastricht University | Maastricht UMC+

The LIPID MAPS hierarchy (in Wikidata)



class	classLabel	lmid	count
wd:Q63433687	fatty acyl	LMFA	0
wd:Q63434442	straight chain fatty acids	LMFA0101	37
wd:Q24901874	branched chain fatty acids	LMFA0102	79
wd:Q61737535	unsaturated fatty acid	LMFA0103	279
wd:Q40211102	hydroxy fatty acid	LMFA0105	184
wd:Q63435564	oxo fatty acids	LMFA0106	56
wd:Q63436532	halogenated fatty acids	LMFA0109	24
wd:Q63434663	amino fatty acids	LMFA0110	39
wd:Q422050	dicarboxylic acid	LMFA0117	78
wd:Q61716319	octadecanoids	LMFA02	82
wd:Q407680	Eicosanoid	LMFA03	83
wd:Q209717	prostaglandins	LMFA0301	89
wd:Q4198767	isoprostane	LMFA0311	5
wd:Q378871	fatty alcohol	LMFA05	156

In which species is this lipid found?

lipid	lipidLabel	lmid	species	speciesLabel	source	sourceLabel	doi
wd:Q26840883	(-)-methyl jasmonate	LMFA02020010	wd:Q23501	Solanum lycopersicum	wd:Q33228063	Induced defences in plants reduce herbivory by increasing cannibalism	10.1038/S41559-017-0231-6
wd:Q27158341	quercetin 5,7,3',4'-tetramethyl ether	LMPK12112771	wd:Q22701	Sambucus nigra	wd:Q39812430	Elderberry flavonoids bind to and prevent H1N1 infection in vitro.	10.1016/J.PHYTOCHEM.2009.06.003
wd:Q55620521	(R)-1,7-Dioxaspiro[5.5]undecane	LMPK09000012	wd:Q2207329	olive fruit fly	wd:Q55645881	Sex-specific activity of (R)-(-)- and (S)-(+)-1,7-dioxaspiro[5.5]undecane, the major pheromone of Dacus oleae	10.1007/BF01012372
wd:Q55620476	(S)-1,7-Dioxaspiro[5.5]undecane	LMPK09000013	wd:Q2207329	olive fruit fly	wd:Q55645881	Sex-specific activity of (R)-(-)- and (S)-(+)-1,7-dioxaspiro[5.5]undecane, the major pheromone of Dacus oleae	10.1007/BF01012372
wd:Q27135687	geranylacetone	LMFA11000696	wd:Q16528	Nelumbo nucifera	wd:Q902623	ChEBI	
wd:Q27135687	geranylacetone	LMFA11000696	wd:Q16528	Nelumbo nucifera	wd:Q43240571	Comparative analysis of essential oil components and antioxidant activity of extracts of Nelumbo nucifera from various	10.1021/JF902643E

Poster

figshare Browse Search on figshare... Log in Sign up

Maastricht University

NUTRIM

NUTRIM School of Nutrition and Translational Research in Metabolism

Wikidata and Scholia as a hub linking chemical knowledge

Egon Willighagen[A], Denise Slenter[A], Daniel Mietchen[B], Chris Evelo[A,C], Finn Nielsen[D]

[A] Department of Bioinformatics - BiGCaT, Maastricht University, The Netherlands, [B] Data Science Institute, University of Virginia, Charlottesville, Virginia, USA,
[C] Maastricht Centre for Systems Biology - MaCSBio, Maastricht University, The Netherlands, [D] Cognitive Systems, DTU Compute, Technical University of Denmark, Denmark

Introduction

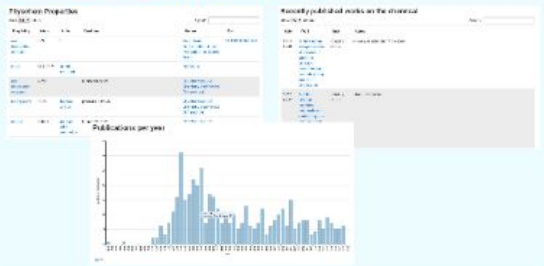
Making chemical databases more FAIR (findable, accessible, interoperable, and reusable) benefits computational chemistry and cheminformatics. We here discuss Wikidata, a young sister project of Wikipedia, with one key difference: it is a machine readable database, making it far more useful for interoperability of molecular databases in systems biology [1,2]. Thanks to the WikiProject Chemistry community on Wikidata, there is a growing amount of information about chemical compounds.

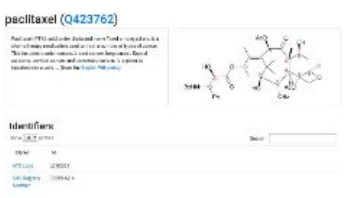
Results

We here introduce our contributions to the WikiProject Chemistry to support FAIR-ification of open chemical knowledge. For example, we proposed new Wikidata properties to annotate compounds with external database identifiers for the EPA CompTox Dashboard [3], the SPLASH [4], and MetaboLights. We also introduced a Scholia extension [5], visualizing data about chemicals and chemical classes:

https://tools.wmflabs.org/scholia/

Literature-backed (PhysChem) Facts





Provenance: "stated in"

Identifiers

Related compounds

Lookup by identifier

Methods

Scholia is a Python/Flask-based server system that creates webpages using a template approach [5]. It defines templates for concepts around knowledge exchange, such as publications, journals, publishers, but also topics. It uses SPARQL queries against the Wikidata Query Service (WDQS,

Linking Databases

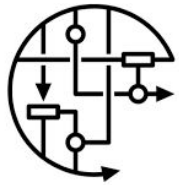
NUTRIM School of Nutrition and Translational Research in Metabolism

https://doi.org/10.6084/m9.figshare.6356027.v1

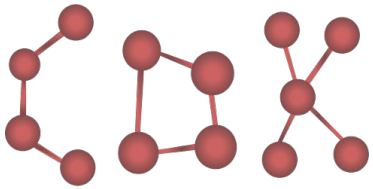
Maastricht University | Maastricht UMC+

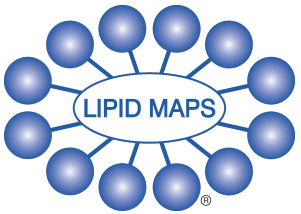
Acknowledgements

- the many people of WikiProject Chemistry
- Denise Slenter (PhD candidate, metabolites in WikiPathways)
- the Blue Obelisk community
  - Chemistry Development Kit
  - Bioclipse
  - InChI wrapping in Java
- Scholia team: Finn, Daniel, Lane, Ammar
- Roger Sayle (for sharing his slides)



WIKIPATHWAYS
Pathways for the People





List of publications RSS

Show 10 entries

Search: wikidata

Date	Work	Type	Pages	Venue	Authors
2021-03-01	Open Natural Products Research: Curation and Dissemination of Biological Occurrences of Chemical Structures through Wikidata	scholarly article	36	bioRxiv	Adriano Rutz, Jonathan Bisson, Jiří Vondrášek, Pierre-Marie Allard, Jean-Luc Wolfender, Guido F Pauli, Ralf Stephan, Maria Sorokina, James G Graham, Egon Willighagen, Daniel Mietchen, Christoph Steinbeck, Roderic D. M. Page, Jakub Galgonek
2021-01-22	A protocol for adding knowledge to Wikidata: aligning resources on human coronaviruses	scholarly article	14	BMC Biology	Martina Summer-Kutmon, Lisa M. Verhagen, Daniel Fernández-Álvarez, Jasper Koehorst, Quentin Groom, Peter J. Schaap, Andra Waagmeester, José Emilio Labra Gayo, Andrew I. Su, Egon Willighagen
2020-04-07	A protocol for adding knowledge to Wikidata, a case report	preprint		bioRxiv	Martina Summer-Kutmon, Lisa M. Verhagen, Daniel Fernández-Álvarez, Jasper Koehorst, Peter J. Schaap, Andra Waagmeester, José Emilio Labra Gayo, Andrew I. Su, Egon Willighagen
2020-03-17	Wikidata as a knowledge graph for the life sciences	scholarly article	15	eLife	Sebastian Burgstaller-Muehlbacher, Elvira Mitraka, Lynn Schriml, Kristina Hanspers, Henning Hermjakob, Katherine Thornton, Núria Queralt Rosinach, Gregory Stupp, Anders Riutta, Chunlei Wu, Alexander R. Pico, Toby Hudson, Ginger Tsueng, Andra Waagmeester, Kevin Hybiske, Sarah M Keating, Thomas Shafee, Sabah Ul-Hasan, Michael Mayers, Roger Tu, Ralf Stephan, Timothy Elliott Putman, Andrew I. Su, Benjamin M. Good, Egon Willighagen, Malachi Griffith, Daniel Mietchen, Magnus Manske, Obi Griffith, Denise Slenter

InChI and InChIKey in Wikidata and Scholia

Egon Willighagen
NIH Virtual Workshop on InChI
March 22-24, 2021

@egonwillighagen
0000-0001-7542-0286

NUTRIM
School of Nutrition and Translational Research in Metabolism
Maastricht University | Maastricht UMC+