

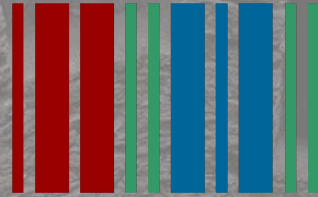


Multilingual Structured (Climate) Research Data in Wikidata

The Data Perspective

Daniel Mietchen & Cristina Sarasua

DOI: [10.5281/zenodo.3994266](https://doi.org/10.5281/zenodo.3994266)



WIKIDATA

*The free knowledge base that
anyone can use and edit.*



- Main page
- Community portal
- Project chat
- Create a new Item
- Create a new Lexeme
- Recent changes
- Random Item
- Query Service
- Nearby
- Random Primary Sources item
- Help
- Donate

Tools

- What links here
- Related changes
- Special pages
- Permanent link
- Page information
- Cite this page
- Concept URI
- Reasonator
- Ancestors
- Automatic addition
- VIP's labels

Browse Primary Sources

Item [Discussion](#) [Read](#) [Labels list](#) [View history](#) [☆](#) [←](#) [→](#) [Page](#) [More](#)

Hurricane Katrina (Q16422)...

Atlantic hurricane in 2005

[edit](#)

[In more languages](#)

[Configure](#)

Language	Label	Description	Also known as
English	Hurricane Katrina	Atlantic hurricane in 2005	
French	ouragan Katrina	ouragan de la saison cyclonique 2005 de l'océan Atlantique Nord	
Spanish	Huracán Katrina	huracán Categoría 5	Huracan Katrina
Traditional Chinese	颶風卡特里娜	No description defined	

[All entered languages](#)

Statements

instance of	category 5 hurricane ... edit
	determination method Saffir–Simpson hurricane wind scale ...
0 references	
+ add reference	
+ add value	



- 首页
- 社区门户
- 互助客栈
- 创建新项
- 创建新词位
- 最近更改
- 随机项目
- 查询服务
- 附近
- Random Primary Sources item
- 帮助
- 资助

工具

- 链入页面
- 相关更改
- 特殊页面
- 固定链接
- 页面信息
- 引用本页
- 概念URI
- Reasonator
- Ancestors
- 自动添加内容
- VIP's labels

Browse Primary Sources

▲ back to top ▲

<https://www.wikidata.org/wiki/Q16422?uselang=zh>

项目 [讨论](#)

[阅读](#)

[标签列表](#)

[查看历史](#)



Page ▾

更多 ▾

颶風卡特里娜 (Q16422) …

2005年出現的一個五級颶風

[编辑](#)

▾ [更多语言](#)

[配置](#)

语言	标签	描述	别名
中文	颶風卡特里娜	2005年出現的一個五級颶風	
英语	Hurricane Katrina	Atlantic hurricane in 2005	
德语	Hurrikan Katrina	Tropischer Wirbelsturm mit verheerenden Folgen, 2005	Hurricane Katrina
法语	ouragan Katrina	ouragan de la saison cyclonique 2005 de l'océan Atlantique Nord	

所有输入的语言

声明

性質	<div><div><div> category 5 hurricane 英语 …</div><div> 断定方法 萨菲尔-辛普森颶風风力等级 …</div><div> 0个来源</div></div></div> <div>编辑</div> <div>+ 添加参考文献</div> <div>+ 添加值</div>
-----------	--

所属实体	<div><div><div> 2005年大西洋颶風季 …</div></div></div> <div>编辑</div>
-------------	---

Default data formats

<http://www.wikidata.org/entity/Q1>.json

.php

.n3

.ttl

.nt

.rdf

.html

.json-ld

universe (Q1)...

totality consisting of space, time, matter and energy

[Our Universe](#) | [The Universe](#) | [The Cosmos](#) | [cosmos](#) | [Heaven and earth](#) | [Yin and](#)

[Yang](#) | [space](#) | [world](#) | [creation](#) | [system](#) | [metagalaxy](#) | [existence](#) | [all](#) | [macrocosm](#) | [outer space](#) | [Un](#)

▼ [In more languages](#)

[Configure](#)

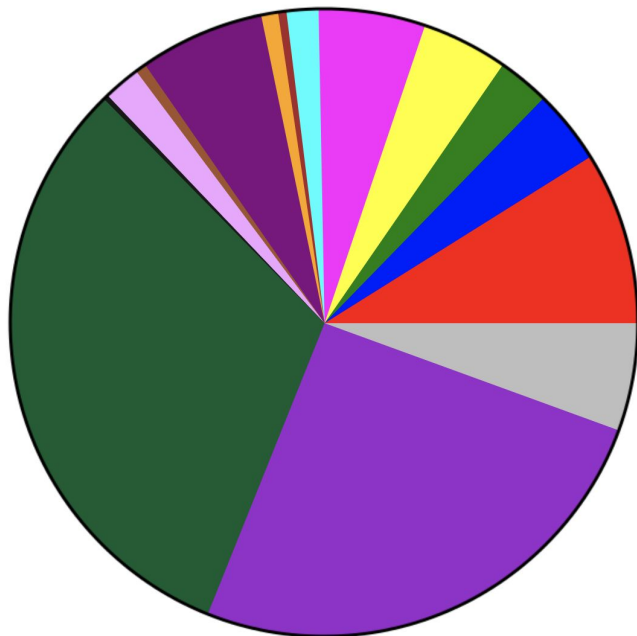
Language	Label	Description
English	universe	totality consisting of space, time, matter and energy

Data content



What is in Wikidata?

71,611,020 items as of 2020-02-16



	human: 6,376,879 (8.9%)
	taxon: 2,726,046 (3.8%)
	administrative territorial entity: 1,943,285 (2.7%)
	architectural structure: 3,159,472 (4.4%)
	occurrence: 3,898,674 (5.4%)
	chemical compound: 1,188,724 (1.7%)
	film: 294,370 (0.4%)
	thoroughfare: 630,794 (0.9%)
	astronomical object: 4,601,733 (6.4%)
	Wikimedia list article: 404,454 (0.6%)
	Wikimedia disambiguation page: 1,358,230 (1.9%)
	Wikinews article: 195,900 (0.3%)
	scholarly article: 22,574,314 (31.5%)
	other P31/P279: 18,284,676 (25.5%)
	no P31/P279: 3,973,469 (5.5%)

7,841 properties as of 2020-08-20

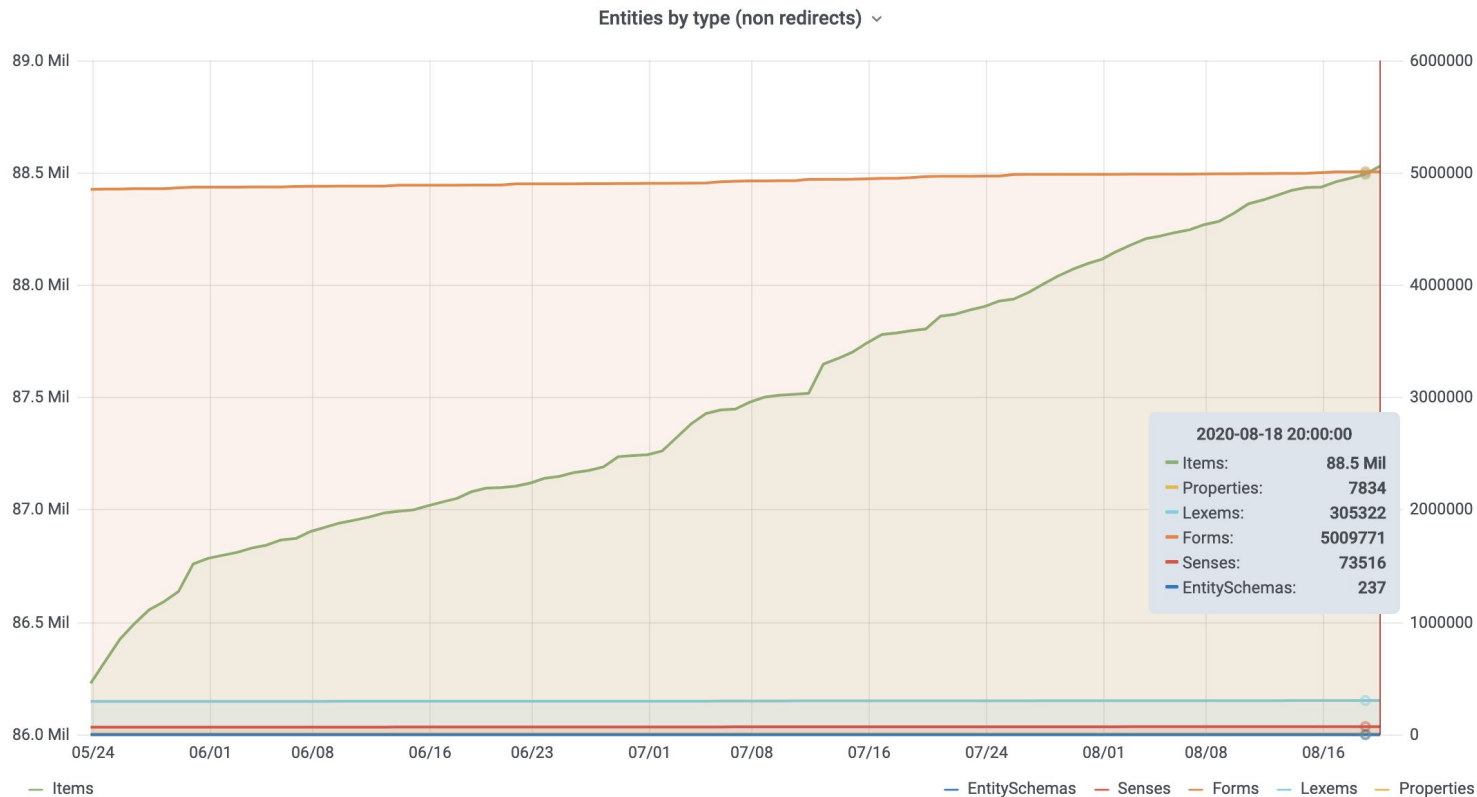
Property	Quantity of item pages
cites work (P2860)	174377685
series ordinal (P1545)	151037942
author name string (P2093)	132282330
instance of (P31)	93773388
stated in (P248)	67709756
retrieved (P813)	66095593
PubMed ID (P698)	59230552
reference URL (P854)	47319398
publication date (P577)	40828990
title (P1476)	40260850
published in (P1433)	35909547
page(s) (P304)	34571714
volume (P478)	34052572
apparent magnitude (P1215)	33123972
astronomical filter (P1227)	33123919
issue (P433)	30999754
catalog code (P528)	28698782
DOI (P356)	25855935
catalog (P972)	23771397
author (P50)	18814323
language of work or name (P407)	15201689
country (P17)	12999974

Supported data types

- Commons media
- Globe coordinate
- Item
- Property
- String
- Monolingual text
- External identifier
- Quantity
- Time
- URL
- Mathematical expression
- Geographic shape
- Musical Notation
- Tabular data
- Lexemes
- Forms
- Senses

https://www.wikidata.org/wiki/Help:Data_type

Data types over time



Most recent page creations

The ten most recently created items [edit]

- 00:03, 21 August 2020 Category:Religious performing arts (Q98534767) (hist) [9,049 bytes] **M** (Created a new Item: *Bot: New item from enwiki/arwiki.*)
- 00:03, 21 August 2020 Category:Religious fiction (Q98534766) (hist) [8,995 bytes] **Mr.Ibrahe** (Created a new Item: *Bot: New item from enwiki/arwiki.*)
- 00:03, 21 August 2020 Category:Religious plays (Q98534765) (hist) [8,991 bytes] **Mr.Ibrahe** (Created a new Item: *Bot: New item from enwiki/arwiki.*)
- 00:02, 21 August 2020 Category:Events in Iran by subject (Q98534764) (hist) [3,023 bytes] **Cr** (Created a new Item: *Category:Events in Iran by subject, Wikimedia category*)
- 00:01, 21 August 2020 Christian Meyer (Q98534763) (hist) [2,886 bytes] **PKM** (talk | contri**B**) (Created a new Item: *New item based on Christian Meyer (#10173877)*) (Tag: Mix'n'match [1.0])
- 00:01, 21 August 2020 Biography of an American Bondman (Q98534762) (hist) [2,467 bytes] **Dsp13** (talk | contri**B**) (Created a new Item: *New item based on Biography of an American Bondman (#83213643)*) (Tag: Mix'n'match [1.0])
- 00:01, 21 August 2020 Kim Jinai (Q98534761) (hist) [494 bytes] **Pi bot** (talk | contri**B**) (Created a new Item: *BOT - Creating item for Kim Jinai (en): human female*)
- 00:01, 21 August 2020 Marcos Leonardo Santos Almeida (Q98534760) (hist) [4,271 bytes] **Pi bot** (talk | contri**B**) (Created a new Item: *BOT - Creating item for Marcos Leonardo Santos Almeida (en): human male*)
- 00:00, 21 August 2020 XII tabulae : testo e commento (Q98534759) (hist) [5,439 bytes] **JerfBTaylor** (talk | contri**B**) (Created a new Item: *XII tabulae : testo e commento, 2018 academic book*)
- 00:00, 21 August 2020 Category:Savo Karjala building (Q98534758) (hist) [4,871 bytes] **Yzarg** (talk | contri**B**) (Created a new Item)

The ten most recently created lexemes [edit]

See also *Wikidata:Lexical Masks*.

- 18:36, 20 August 2020 μρμηγήκι (L309631) (hist) [698 bytes] **Simplificationalizer** (talk | contri**B**)
- 18:36, 20 August 2020 μρμηγκιον (L309630) (hist) [270 bytes] **Simplificationalizer** (talk | contri**B**)
- 18:34, 20 August 2020 κρηφίνας (L309629) (hist) [696 bytes] **Simplificationalizer** (talk | contri**B**)
- 18:34, 20 August 2020 κρηφίν (L309628) (hist) [262 bytes] **Simplificationalizer** (talk | contri**B**) (Created a new entity: *κρηφίν*)
- 18:30, 20 August 2020 σκουλήκι (L309627) (hist) [698 bytes] **Simplificationalizer** (talk | contri**B**) (Created a new entity: *σκουλήκι*)
- 18:29, 20 August 2020 σκωλήκιον (L309626) (hist) [270 bytes] **Simplificationalizer** (talk | contri**B**) (Created a new entity: *σκωλήκιον*)
- 18:27, 20 August 2020 স্মৃষ্ণ (L309625) (hist) [1,281 bytes] **Titodutta** (talk | contri**B**) (Created a new Lexeme: *bengali-noun-inanimate-consonants*) (Tag: Wikidata Lexeme Forms [1.2])
- 18:26, 20 August 2020 timpuriu (L309624) (hist) [689 bytes] **Simplificationalizer** (talk | contri**B**) (Created a new entity: *timpuriu*)
- 18:22, 20 August 2020 ঔগ্ধ (L309623) (hist) [700 bytes] **Simplificationalizer** (talk | contri**B**) (Created a new entity: *ঔগ্ধ*)
- 18:21, 20 August 2020 मेजा (L309622) (hist) [693 bytes] **Simplificationalizer** (talk | contri**B**) (Created a new entity: *मेजा*)

seconds
to
minutes

minutes
to
hours

hours
to
days

The ten most recently created properties [edit]

- 19:23, 20 August 2020 recording location (P8546) (hist) [7,150 bytes] **Pamputt (A)** (talk | contri**B**) (Property: *recording location, location where a media item (sound/video/etc) was recorded*)
- 09:56, 20 August 2020 Nasjonalt skoleregister ID (P8545) (hist) [13,025 bytes] **Thierry Caro** (talk | contri**B**) (New Property: *Nasjonalt skoleregister ID, identifier of an educational institution in the Nasjon. School Registry) from the Norwegian Directorate for Education and Training*)
- 06:08, 20 August 2020 Frauen in Bewegung 1848–1938 ID (P8544) (hist) [11,584 bytes] **Pam** (Created a new Property: *Frauen in Bewegung 1848–1938 ID, identifier for entries in the Pam. Bewegung 1848–1938® database, published by the Austrian National Library*)
- 19:23, 19 August 2020 Monumentbrowser ID (P8543) (hist) [7,751 bytes] **Tobias1984** (talk | contri**B**) (Created a new Property: *Monumentbrowser ID, ID im Verzeichnis der Baudenkmläer in der autonomen italienischen Provinz Südtirol*)
- 17:20, 19 August 2020 Macedonian cadastral municipality ID (P8542) (hist) [10,480 bytes] **Pamputt (A)** (talk | contri**B**) (Created a new Property: *Macedonian cadastral municipality ID, identifier of Macedonian cadastral municipalities*)
- 16:54, 19 August 2020 BHF magazine ID (P8541) (hist) [9,060 bytes] **Pamputt (A)** (talk | contri**B**) (Created a new Property: *BHF magazine ID, identifier for a magazine on the online Bibliographie de l'histoire de France*)
- 05:59, 19 August 2020 American Folklore Society Ethnographic Thesaurus ID (P8540) (hist) [7,779 bytes] **Pamputt (A)** (talk | contri**B**) (Created a new Property: *American Folklore Society Ethnographic Thesaurus ID, controlled vocabulary of terms about folklore, ethnomusicology, ethnology, and related fields*)
- 21:45, 18 August 2020 ERIC Thesaurus ID (P8539) (hist) [10,528 bytes] **Pamputt (A)** (talk | contri**B**) (Created a new Property: *ERIC Thesaurus ID, identifier for a term in the Education Resources Information Center (ERIC) Thesaurus*)
- 18:00, 18 August 2020 raga (P8536) (hist) [9,364 bytes] **Pamputt (A)** (talk | contri**B**) (Created a new Property: *raga, melodic framework for Indian music*)
- 17:59, 18 August 2020 tala (P8535) (hist) [9,353 bytes] **Pamputt (A)** (talk | contri**B**) (Created a new Property: *tala,*

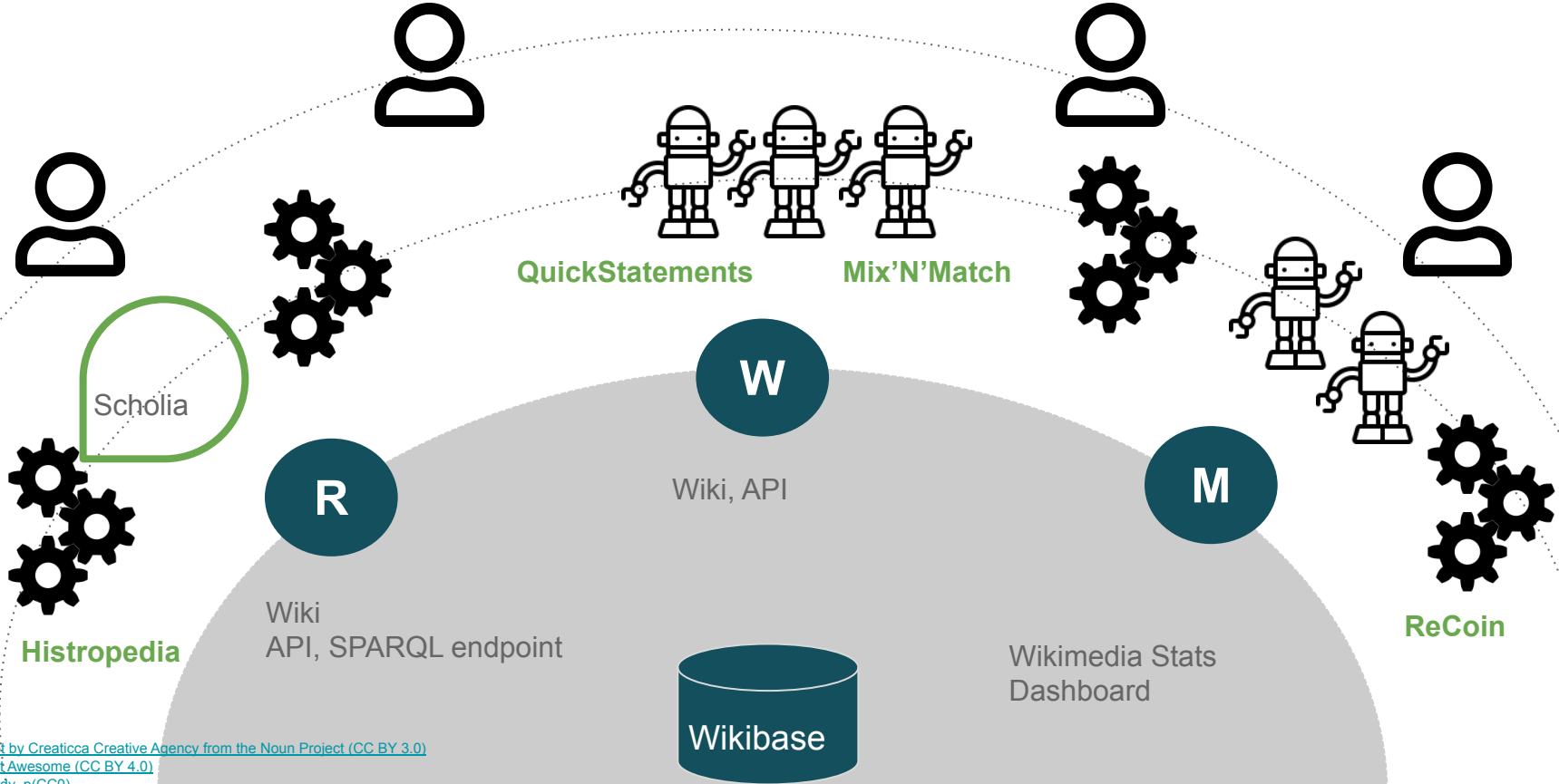
The ten most recently created pages in the Entity schema namespace [edit]

See also *Special:NewEntitySchema*, a list of schemas and *Schema visualization@*.

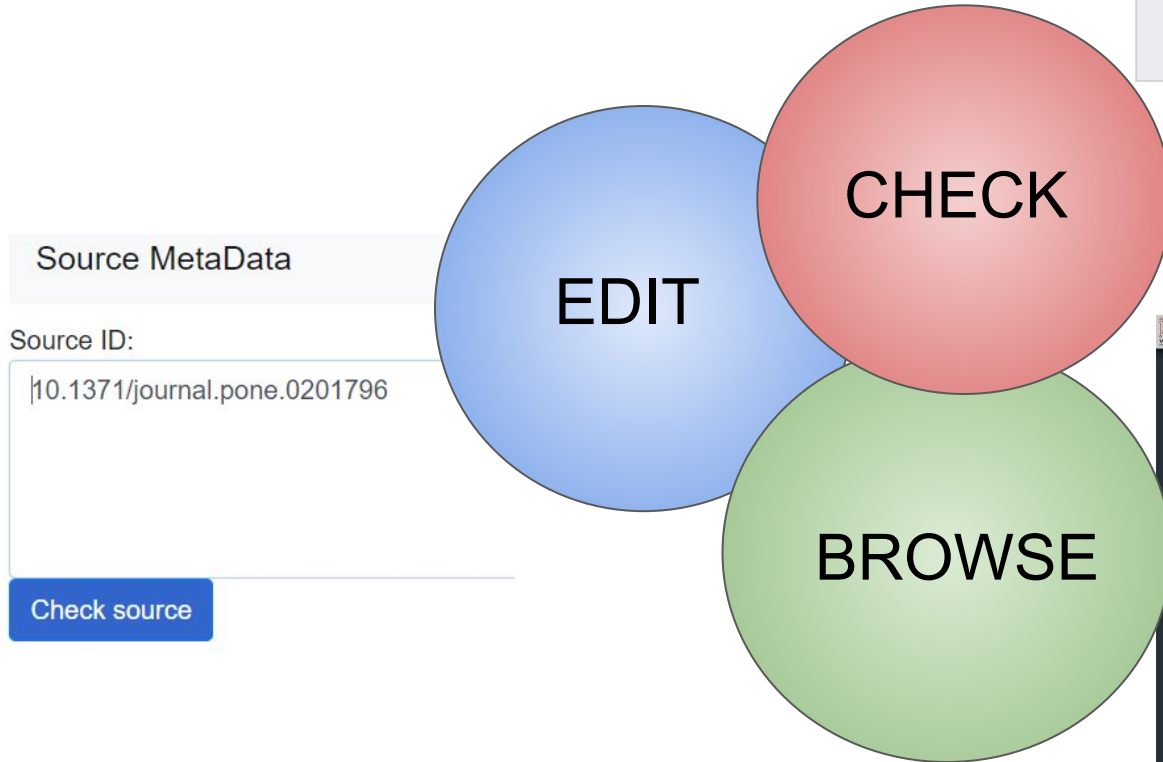
- 14:48, 19 August 2020 EntitySchema:E236 (hist | edit) [1,188 bytes] **Teester** (talk | contri**B**) (Member of the *Oireachtas*)
- 10:52, 6 August 2020 EntitySchema:E235 (hist | edit) [1,509 bytes] **Teester** (talk | contri**B**) (*auto racing event*)
- 11:22, 1 August 2020 EntitySchema:E234 (hist | edit) [1,168 bytes] **Teester** (talk | contri**B**) (*auto race track*)
- 08:57, 25 July 2020 EntitySchema:E233 (hist | edit) [563 bytes] **Egon Willighagen** (talk | contri**B**) (Created a new Schema: *protein family*)
- 08:51, 25 July 2020 EntitySchema:E232 (hist | edit) [569 bytes] **Egon Willighagen** (talk | contri**B**) (Created a new Schema: *lipid*)
- 05:57, 23 July 2020 EntitySchema:E231 (hist | edit) [1,510 bytes] **Salgo60** (talk | contri**B**) (Created a new Schema: *Svenska Akademin - medlem*)
- 05:57, 23 July 2020 EntitySchema:E230 (hist | edit) [1,574 bytes] **Salgo60** (talk | contri**B**) (Created a new Schema: *Svenska Akademin - medlem*)
- 19:41, 22 July 2020 EntitySchema:E229 (hist | edit) [2,990 bytes] **Belleshassar** (talk | contri**B**) (Created a new Schema: *decision from the Supreme Court of Sweden*)
- 16:21, 22 July 2020 EntitySchema:E228 (hist | edit) [2,862 bytes] **Piecesofuk** (talk | contri**B**) (Created a new Schema: *United Kingdom Member of Parliament*)

days
to
weeks

Tools in the Wikidata Ecosystem



Overview of Wikidata tools



symptoms

temporary blindness

edit

Potential issues

citation needed constraint Help Discuss

Statements for **symptoms** should have at least one reference.

+ add statement

WIKIDATA FOR SOCIAL PROVISION

about reports news

support

try a new search...

Python (Q28865)
— An interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. —
also referred to as: py / python 2 / python 3

Python (Q271218)
— genus of reptiles —

Python (Q15721)
— earth-dragon of Delphi —

<https://www.wikidata.org/wiki/Wikidata:Tools>

Browsing Wikidata

Crotos
Cosmos
Callisto
Lab

Language: English Images/page: 20
Contribution mode

Everything -40000 2016

Crotos is a search and display engine for visual artworks powered by Wikidata and Wikimedia Commons

- 279 809 artworks, including 177 039 with HD image
- Links to Wikipedia, reference websites, multilingual, free and open, Lab

Cosmos Callisto Random Chronological order

Christ on the Cross with Three Angels
Albrecht Dürer

At Work by the Greenhouse
Edward Munch - Stavanger kunstmuseum

Standbeeld koningin Wilhelmina

Still life with four shells
Adriaen Coen - private collection

[Crotos](#), focused on works of art

WIKIDATA FOR DIGITAL PRESERVATION

about reports news support

Python (Q28865)

An interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace.

Language: English

Current Data About Python

object-based language (Q389923)	aliases	py python 2 python 3
reference URL (P854) https://www.python.org/doc/essays/comparisons/		
retrieved (P813) Friday, August 19, 2016		
dynamic programming language (Q1268880)		An interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of
reference URL (P854) http://www.oznetherd.com/python-strong-dynamically-typed-language/		
retrieved (P813) Thursday, July 5, 2018		
functional programming language (Q3839207)		
reference URL (P854) https://docs.python.org/2/faq/faq_1.html#faq-1.1		

other details

property checklist

- instance of (P179) : 10
- based on (P178) : 0
- developer (P177) : 2
- subclass of (P176) : 0
- software version (P175) : 86
- publication date (P174) : 0
- reference URL (P854) : 0
- described at (P173) : 0
- media type (P172) : 3
- file extension (P171) : 6
- described by (P170) : 0
- relative to (P169) : 0
- PRONOM file (P168) : 0
- Library of Congress (P167) : 0
- encoding (P166) : 0

[Wikidata for Digital Preservation](#),
focused on software

See also: [List of Wikidata front ends](#)

Showcase



<https://scholia.toolforge.org/>

Scholarly profiles of

- People
- Institutions
- Topics
- Events
- Publication venues
- Individual works
- Publishers
- Awards
- Genes
- Proteins
- Clinical trials
- Wikidata
- Scholia
- etc.

based on

- Open data
- Open source
- Open collaboration
- Community curation
- Work in progress



<https://scholia.toolforge.org/>



[Author](#) [Work](#) [Organization](#) [Location](#) [Event](#) [Project](#) [Award](#) [Topic](#) [Tools](#) [Help](#)

Scholia is a service that creates visual scholarly profiles for [topic](#), [people](#), [organizations](#), [species](#), [chemicals](#), etc using bibliographic and other information in Wikidata.

Scholia relies on Wikidata, and Wikidata contains only a limited albeit growing subset of the corpus of scholarly literature, its authors and citations. Read more about the limitations in the [FAQ](#).

Search

Workshop on data

[Workshop on Data Quality Management in Wikidata](#) - workshop in Berlin in January 2019

[Workshop on Data Science in Climate and Climate Impact Research](#) - virtual Workshop at ETH Zurich on 20-21 August 2020

Examples

Showcase examples

- Event: [Workshop on Data Science in Climate and Climate Impact Research](#)
- Author: [Markus Reichstein](#) / [Reto Knutti](#) / [Sonia I. Seneviratne](#)
- Topic: [Hurricane Katrina](#) / [Permafrost](#) / [Remote sensing](#)
- Institution: [US Geological Survey](#) / [University of Nairobi](#)
- Journals: [Global Change Biology](#) / [Biogeosciences](#)
- Works: [The Diplommatinidae of Fiji - a hotspot of Pacific land snail biodiversity \(Caenogastropoda, Cyclophoroidea\)](#)
- Pathway: [Aryl Hydrocarbon Receptor Pathway](#)
- Curation: [Ngrams related to climate impacts](#)

- More via <https://www.wikidata.org/wiki/Wikidata:Scholia> .
- Code at <https://github.com/fnielsen/scholia> .

Thanks!

- Providers of open infrastructure
 - Wikidata community
 - Scholia team
- Alfred P. Sloan Foundation ([G-2019-11458](#))
 - The organizers of the workshop
 - YOU

Questions?

Cristina Sarasua <sarasua@ifi.uzh.ch> Daniel Mietchen <dm7gn@virginia.edu>



University of
Zurich^{UZH}



UVA

SCHOOL of DATA SCIENCE

[Image by Addshore, CC0](#)