

State of WikiCite 2020 (Version 3)

Backchannels for questions & comments

Please put questions and comments into the Etherpad at

<https://w.wiki/iGK>

For issues requiring a deeper discussion, consider joining the
[WikiCite group on Telegram](#).

Structure of the talk

- Hello WikiCite!
- Previous versions of State of WikiCite 2020
- Overview of WikiCite
- WikiCite developments since the 2018 meeting in Berkeley
 - Content
 - Infrastructure, tools and workflows
 - Grants & fellowships
 - Events, including WikiCite 2020 sessions and talks
- Getting involved
- Q & A

Hello WikiCite!

◻ I am

- Daniel Mietchen ([User:Daniel Mietchen](#))
- thankful to organizers of WikiCite activities
- a member of the [WikiCite Steering Committee](#)
- a biophysicist by training
- a data scientist at the University of Virginia's [School of Data Science](#)
- [@EvoMRI](#) on Twitter
- looking forward to questions and comments



State of WikiCite 2020

Previous versions

Source: <https://meta.wikimedia.org/wiki/WikiCite/administration>

Daniel Mietchen - 26 October 2020 - DOI: [10.5281/zenodo.4121846](https://doi.org/10.5281/zenodo.4121846) - Q&A: <https://www.wiki/iGK>

State of WikiCite 2020 - version 1

Annual reports

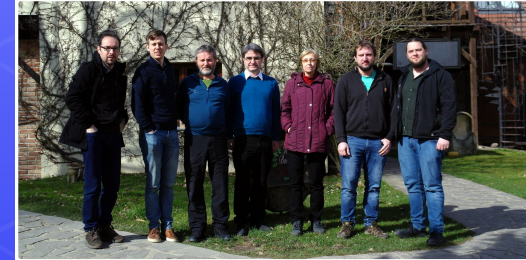


Source: <https://meta.wikimedia.org/wiki/WikiCite/administration>

Daniel Mietchen - 26 October 2020 - DOI: [10.5281/zenodo.4121846](https://doi.org/10.5281/zenodo.4121846) - Q&A: <https://www.wiki/iGK>

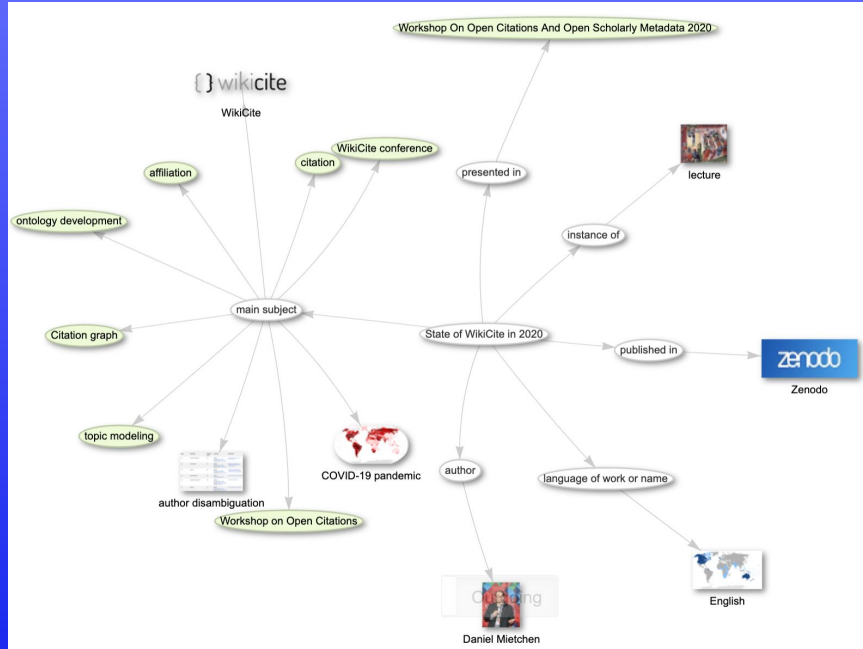
Key takeaways from the annual report

- ⬡ The event series goes decentral & multilingual
 - Australia
 - Czechia
 - ...
- ⬡ The COVID-19 pandemic
 - Modified activities planned in Czechia, Ireland, Finland
 - Canceled events in Ghana, Germany, India
 - Postponed events in Haiti, Ghana
 - Triggered WikiProject COVID-19, with lots of WikiCite in it
- ⬡ The ecosystem around WikiCite is evolving



Datová spolupráce našeho spolku s knihovnami se dynamicky rozvíjí. Poté, co jsme loni zpracovali datovou sadu *personálních národních autorit*, si teď brousíme zuby na další knihovnická data a pilně je propojujeme se záznamy ve Wikidatech. Rozhodli jsme se tedy na březen 2020 svolat knihovnický workshop, který by naší činnost i plány do budoucna uchopil systematicky. Akce podpořená z grantu *WikiCite Satellite Events* původně měla trvat celý víkend a mělo se jí zúčastnit více než 15 knihovníků i wikidatistů. Bohužel zasáhl koronavirus, a tak jsme se sešli pouze na půl den a v poněkud omezeném počtu. I tak se našťastí podařilo probrat mnoho z původně plánovaných témat a účastníci odjžděli s pocitem, že se podařilo odstartovat něco nového a významného na poli knihovnických dat.

State of WikiCite 2020 - version 2



- A talk at the Workshop On Open Citations And Open Scholarly Metadata 2020
- Key takeaways:
 - Despite the pandemic, WikiCite is moving forward, though not sure about the citation part

Source: <https://scholia.toolforge.org/topic/Q99196756>

State of WikiCite 2020

So now on to version 3

Overview of WikiCite (i)

- ⬡ A community
- ⬡ A sociotechnical platform
- ⬡ An ecosystem of projects, tools and other activities
- ⬡ A collection of datasets
- ⬡ A series of events
- ⬡ ...

Overview of WikiCite (ii)

- Community with
 - The (2016) vision
 - [to curate] “the sum of all human citations, as linked open data”
 - the mission
 - “to develop open citations and linked bibliographic data to serve free knowledge” using Wikimedia platforms, especially Wikidata

Overview of WikiCite (iii)

- Community with
 - The goals to
 - Lay the foundations for building a repository of all Wikimedia references as structured data in Wikidata
 - Design data models and technology to improve the coverage, quality, standards-compliance and machine-readability of citations and source metadata in Wikipedia, Wikidata and other Wikimedia projects

Overview of WikiCite (iv)

- Community with
 - A (2018) roadmap that is rarely used

1. Centralized	2. Namespace	3. Sister site	4. Federated
All bibliographic data stays in Wikidata	All bibliographic data stays in Wikidata, but in a dedicated namespace with a dedicated data model (similar to lexemes)	Bibliographic data (other than what's considered notable in Wikidata) moves to an official Wikimedia sister project, powered by Wikibase, and tightly integrated with other Wikimedia projects.	Bibliographic data is handled and curated through a federation of specialized wikibases, with no central hub or data store.

Overview of WikiCite (v)

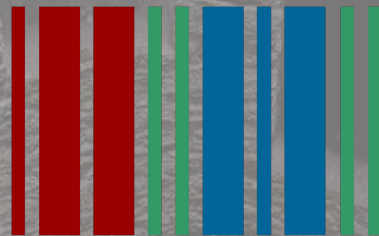
⬡ Homebase

- on (Wikimedia) Meta: [WikiCite](#)
- on Wikidata: [WikiProject Source Metadata](#)
 - Oversees about $\frac{1}{3}$ of Wikidata

⬡ Series of events

- 2016 Berlin, 2017 Vienna, 2018 Berkeley, 2020 decentral

What is Wikidata?



WIKIDATA

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



**Sorry,
what is
Wikidata?**



**Try this
7-minute
intro**

Wikidata in WikiCite numbers (i)

Source: [Scholia homepage](#)

<u>Description</u>	<u>Count</u>
Triples	11,636,281,340
Citations	174,483,508
Scholarly articles	36,635,929
Items with a DOI	26,201,937
ORCIDs	1,651,914

Wikidata in WikiCite numbers (ii)

Description

Count

works to topics

12,216,602

works to people

19,576,747

people to education

1,486,667

people to employer

1,077,272

WikiCite curation workflows

1. topic tagging
 - e.g., "all the papers about SARS-COV-2"
2. author disambiguation
 - e.g., "all papers by a given author"
3. ontology development
 - e.g., "all papers about math education", incl. subtopics
4. subject affiliation
 - e.g., "all papers from people at a given institution"

Topic tagging

14 results in 33983 ms [Code](#) [Download](#) [Link](#)

Ngram	Count	Length	CountTimesLength	N	ExamplePub	ExamplePubTitle
bologna	198	7	1386	1	Q24793288	Problemata Varia Anatomica: MS 1165 The University of Bologna
bologna italy	37	13	481	2	Q56454995	Abstracts of the 23rd Congress of the European Sleep Research Society, 13-16 September 2016, Bologna, Italy
province of bologna	15	19	285	3	Q56991438	Persistence and distribution of pesticide residues in fresh agricultural food consumed in the province of Bologna
city of bologna	14	15	210	3	Q60331681	Exploring the production capacity of rooftop gardens (RTGs) in urban agriculture: the potential impact on food and nutrition security, biodiversity and other ecosystem services in the city of Bologna
university of bologna	9	21	189	3	Q24793288	Problemata Varia Anatomica: MS 1165 The University of Bologna
bologna experience	9	18	162	2	Q56989567	Comprehensive surgical intestinal rescue and transplantation program in adult patients; Bologna experience
pediatric clinic of bologna	6	27	162	4	Q73072978	[Tuberculosis in infants in the case records of the Pediatric Clinic of Bologna]
clinica pediatrica di bologna	4	29	116	4	Q73592317	[Considerations on 143 cases of thymus hyperplasia observed in the last 12 years at Clinica Pediatrica di Bologna]
clinic of bologna	6	17	102	3	Q73072978	[Tuberculosis in infants in the case records of the Pediatric Clinic of Bologna]
bologna-type sausage	5	20	100	2	Q56992383	Effect of pre-emulsified fish oil - as source of PUFA n-3 - on microstructure and sensory properties of mortadella, a Spanish bologna-type sausage
bologna sausage	6	15	90	2	Q56327117	The effect of sodium reduction and the use of herbs and spices on the quality and safety of bologna sausage
population of bologna	4	21	84	3	Q64385949	[Diffusion of complement-fixing antibodies against adenovirus in the population of Bologna.]
pediatrica di bologna	4	21	84	3	Q73592317	[Considerations on 143 cases of thymus hyperplasia observed in the last 12 years at Clinica Pediatrica di Bologna]
bologna criteria	5	16	80	2	Q56927427	The effect of the pooling method on the live birth rate in poor ovarian responders according to the Bologna criteria

Source:
<https://w.wiki/bfh>

Most popular topic tags, August 2020

1. schizophrenia
2. teenager
3. asthma
4. patient
5. tuberculosis
6. obesity
7. India
8. endothelium
9. chemotherapy
10. colorectal cancer
11. nanoparticle
12. statistics

Zika virus infection

features -

1. model for disaster response
2. relatively complete for a pathogen
3. data model enriched as epidemic progresses
4. triggered tool development
5. re-used for technical tests and education

See also -

- ⬡ [WikiProject Zika Corpus](#)
- ⬡ [COVID-19](#)
- ⬡ [WikiProject Humanitarian metadata](#)

Recently published works on the topic

Show entries

Date	Work	Topics
2020-01-06	Neurodevelopmental Abnormalities Associated With In Utero Zika Virus Infection in Infants and Children—The Unfolding Story	Zika virus // neurodevelopmental disorder Zika virus infection
2019-12-13	Measurement error, microcephaly prevalence and implications for Zika: an analysis of Uruguay perinatal data	Uruguay // Zika virus

Showing 1 to 10 of 500 entries

Topics

Co-occurring topics

Show entries

Count	Topic
1084	Zika virus
457	congenital Zika virus infection
343	Zika fever
228	human pregnancy
172	birth defect
136	travel-associated Zika virus infection
124	microcephaly
121	travel-associated infection

Zika virus infection

001



Zika virus infection

Locations of authors of papers on the topic



Author disambiguation

strategy

1. use ORCID if/ when possible
2. anyone can disambiguate as they see fit
3. various mechanisms for quality control
4. tool support available
5. institutions can disambiguate their people
6. avoiding thinking about how big the challenge is

Author name:

☐ Fuzzy match

☐ Wikibase search

☐ Specify name strings

[Look for author](#)

Limit:

Additional SPARQL filters separated by semicolons (eg. for papers on Zika virus, enter wdt:P921 wd:Q202864):

☐ Filter potential authors as well?

Potential publications

83 publications found

[Click here to create clusters based on exact author strings rather than rougher matches.](#)

Group #1

Matched potential author: **Rosa Prato** - author of 9 items

[Check all](#) | [Uncheck all](#)

Title	Authors (identified)	Published In	Identifier(s)	Topic	Published Date	Match?
A Cluster of Cryptic Plasmodium falciparum Malaria in African Migrants in Southern Italy, October 2017	[1]Gaetano Brindicci, [2]Carmen Rita Santoro, [3]Daniela Loconsole, [4]Domenico Martinelli, [5] <input checked="" type="checkbox"/> Rosa Prato , [6]Gaetano Lonerio, [7]Pietro Loperfido, [8]Giovanni Battista Buccoliero, [9]Maria Chironna	Vector-Borne and Zoonotic Diseases [missing]	DOI: 10.1089/VBZ.2019.2607 [ORCID] PubMed: 32833595 [ORCID]	human migration [missing]; Southern Italy [missing]; African people [missing]; Plasmodium falciparum malaria [missing]	2020-08-20	Rosa Prato (Q61150100; 9 items)
Post-marketing surveillance study of the DTaP2-IPV-HB-Hib (Hexyon) vaccine administered in preterm infants in the Apulia region, Italy, in 2017	[1]Domenico Martinelli, [2]Francesca Fortunato, [3]Giulia Del Matto, [4]Giuseppina Iannelli, [5] <input checked="" type="checkbox"/> Rosa Prato	Vaccine [missing]	DOI: 10.1016/J.VACCINE.2020.06.033 [ORCID] PubMed: 32586764 [ORCID]	Italy [missing]; Apulia [missing]; hexavalent vaccine [missing]; preterm infant [missing]	2020-06-22	Rosa Prato (Q61150100; 9 items)

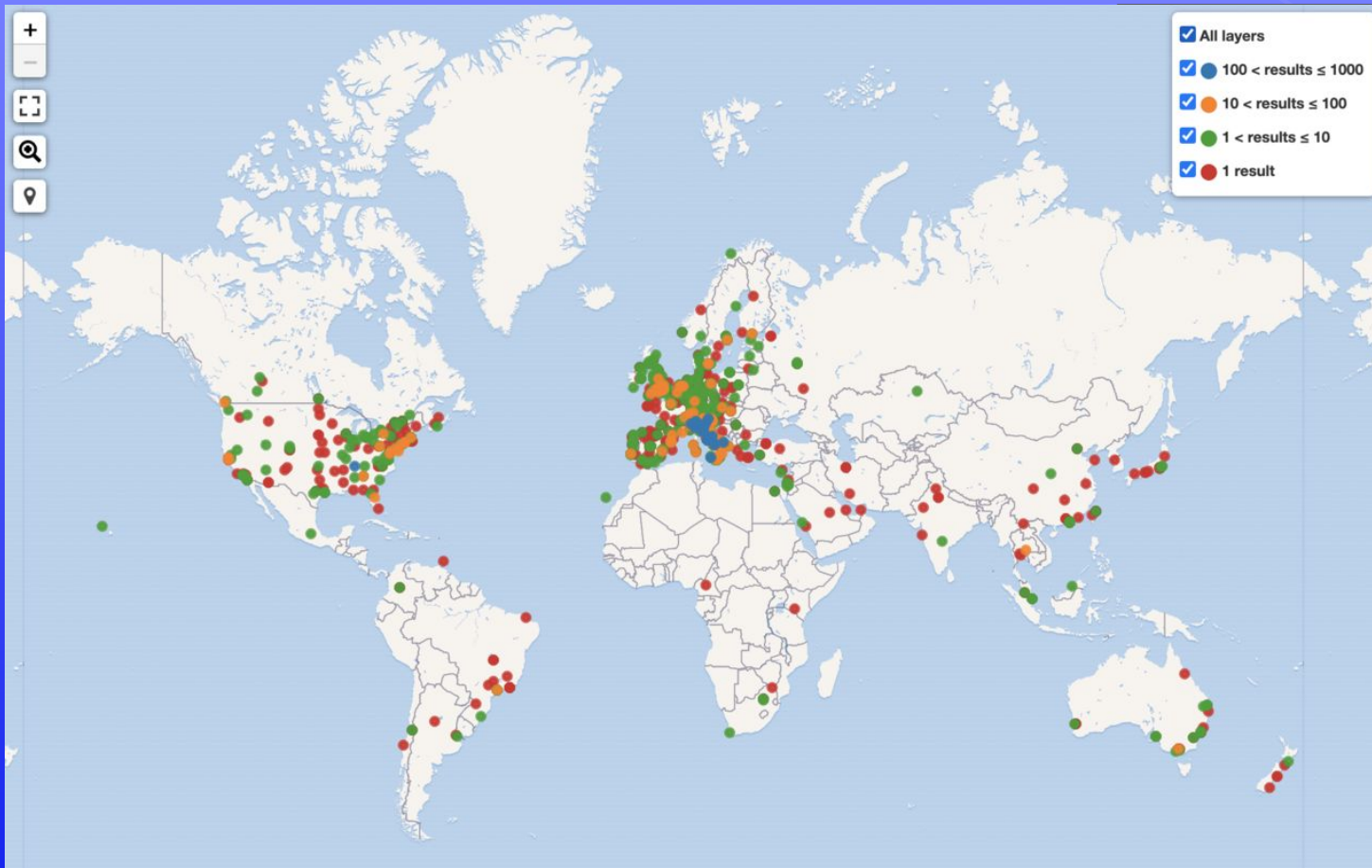
Author disambiguation

Source:
<https://w.wiki/bg6>

Subject affiliation

With affiliations, we can search by university or organization and get other info

1. faculty and their papers
2. research sites and clinical trials
3. events, such as conferences and location



Affiliation tagging

Source:
<https://w.wiki/bgC>

Vanderbilt University

Features -

1. model for how any university can benefit
2. loaded all faculty into Wikidata
3. Working on linking them with their works

see also -

1. [Librarians @ Vandy](#)
2. [Vanderbot](#)
3. [Indiana](#)
4. [Canary Islands](#)
5. [Netherlands](#)

Clinical trials

Medical conditions

Show entries

Count	Medical condition	Ex
9884	diabetes mellitu	The
7194	syndrome	Clin
5350	carcinoma	The
4884	leukemia	Bor
4677	lymphoma	Rac

Interventions

Show entries

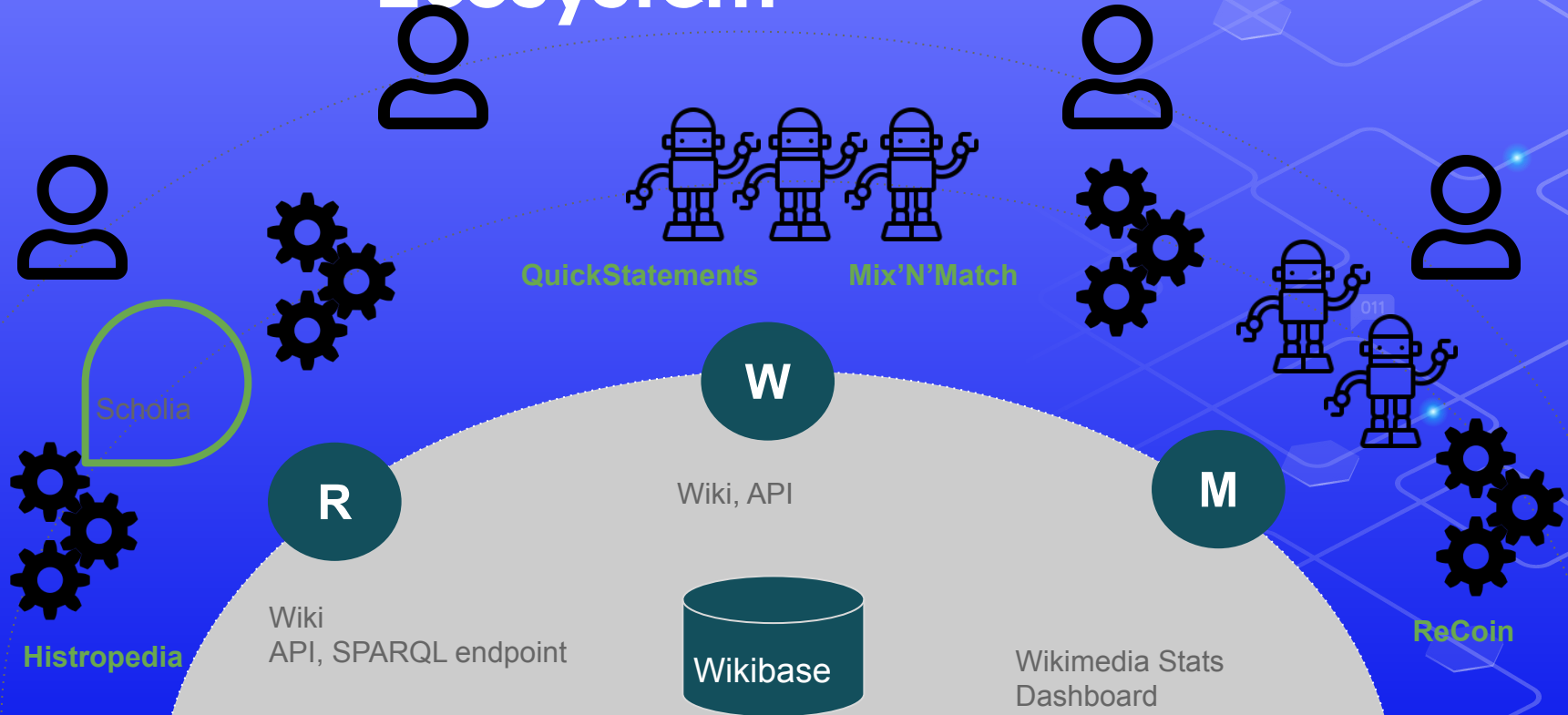
Count	Intervention
5048	cyclophosphamide
3137	Anesthetics, Combined
3099	paclitaxel

Tools in the Wikidata Ecosystem

Bot icon [by Robot](#) by Creatica Creative Agency from the Noun Project (CC BY 3.0)

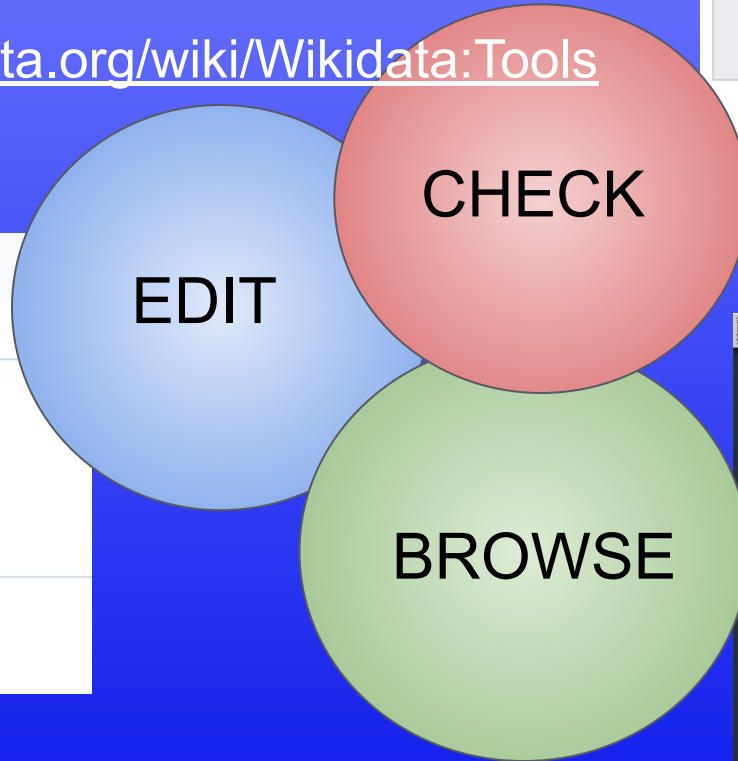
User icon [by Font Awesome](#) (CC BY 4.0)

Tools icon [by Verdy n\(CC0\)](#)



Overview of Wikidata tools

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

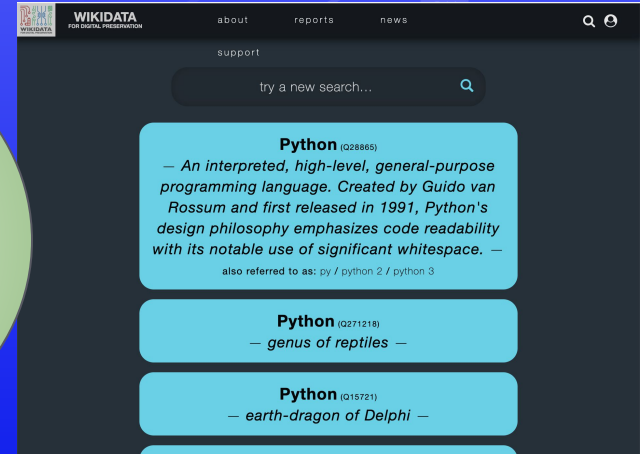
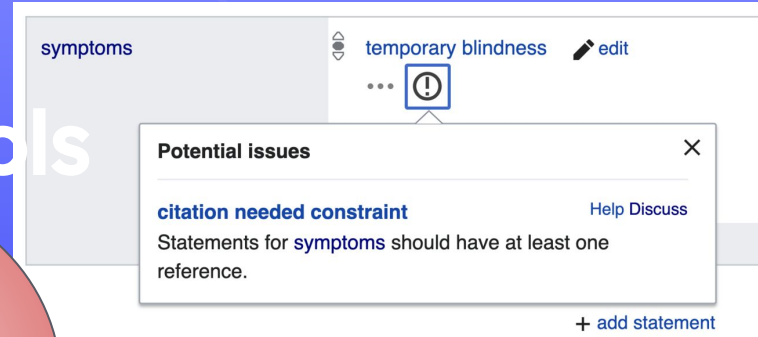


Source MetaData

Source ID:

10.1371/journal.pone.0201796

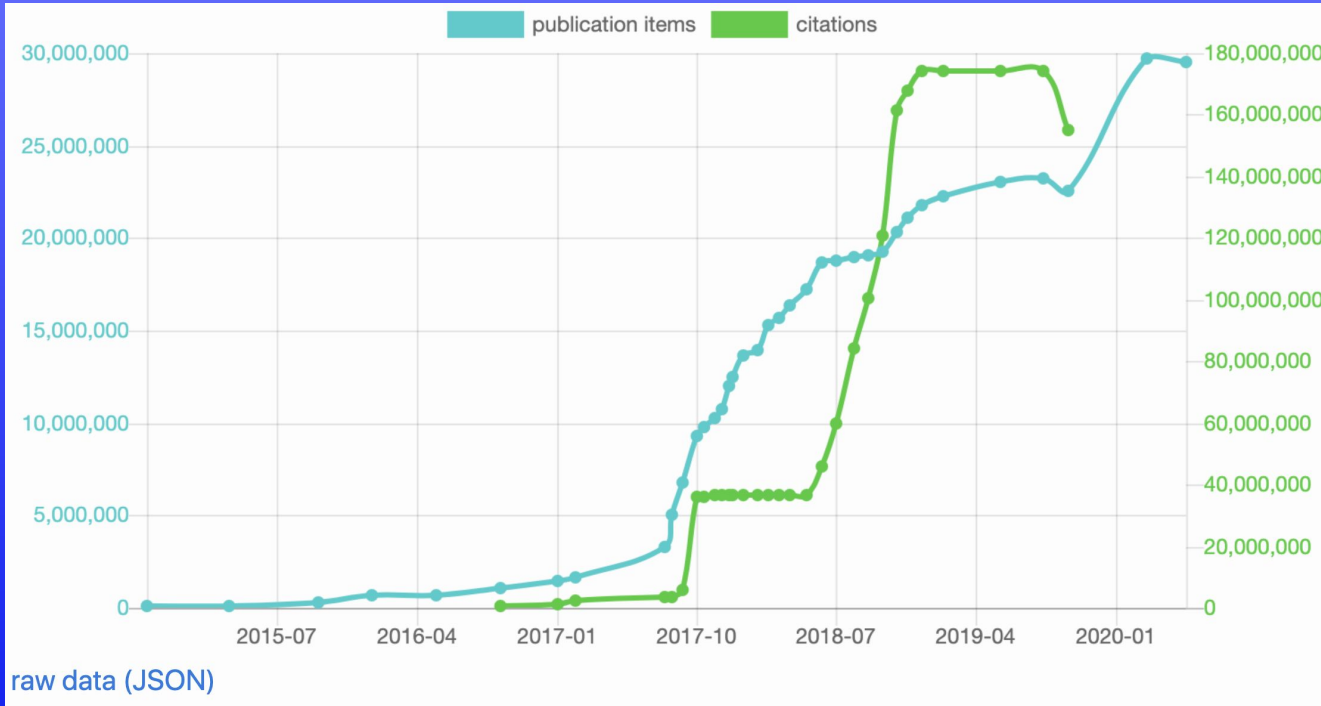
Check source





<https://commons.wikimedia.org/wiki/Category:Scholia>

What about open citations?



Source:
<http://wikicite.org/statistics.html>

Bots

- ⬡ VanderBot
 - Curates Vanderbilt data
- ⬡ LargeDataSetBot
 - Curates publication items
- ⬡ RefB
 - Adds references to Wikidata statements
- ⬡ OpenCitations_Bot (pending)
 - To work on the open citation graph
- ⬡ ...

local anesthesia (Q1455662)

subclass of
by MazinX and RefB (WikiCred)

anesthesia

2 references

stated in The feasibility of local anesthesia for the surgical treatment of umbilical hernia: a systematic review of the literature.

stated in Comparative study between local anesthesia and general anesthesia in the treatment of primary spontaneous pneumothorax

WikiCite events - history



2018 conference - Berkeley



2017 conference - Vienna

 wikicite
Berlin 25-26 May 2016



2016 conference - Berlin

Events - this week's WikiCite 2020 virtualconference

<https://www.wikidata.org/wiki/Q100741900>

- not very detailed yet
- same problem with WikiCite 2016

Overview of WikiCite 2020 events

All sessions apply the Wikimedia Foundation [Friendly space policy](#)

		 UTC	Hosts	Topic & Link to details	YouTube
Monday 26	English	10:00 - 11:00 	Jakob Voß & Eva Seidlmayer 	State of WikiCite	 
	English	13:00 - 16:45 	Simon Cobb  & Jason Evans 	Author items	 
	Deutsch	14:00 - 17:00 	Rabea Müller & Konrad Förstner 	Hands-on: Wikidata-Einführung (Library Carpentry style)	 
	English	17:00 - 20:00 	Jan Ainali  & Daniel Eriksson 	Citations in Swedish Parliamentary documents	 
	English	22:00 - 24:00 	Thomas Shafee & Alex Lum 	Research output items	 
Tuesday 27	English & Deutsch	10:00 - 12:00 	Eva Seidlmayer & Jakob Voß 	Wissensorganisation und Bibliographische Daten	 
	English	13:00 - 15:00 	Amir E. Aharoni & Eran Rosenthal 	The frontend of WikiCite	 
	Français	15:00 - 19:00 	Houcemeddine Turki & Mohamed Ben Aouicha 	WikiCite et scientométrie ouverte	 
Wednesday 28	Bahasa Indonesia	09:00 - 10:30 	Raisha Abdillah, Biyanto Rebin & Hillun Vilayl Napis 	Proyek Wikicite Indonesia	 
	English	09:00 - 10:40 	Leyla Garcia, Arindam Halder, Muhammad Elhoussary 	Gene Wiki	 
	Deutsch	12:00 - 13:00 	Jakob Voß & Eva Seidlmayer 	Methoden und Beispiele für Wikidata-Datenimporte	 
	English	12:00 - 13:30 	EricaAzzellini & Sturm 	OpenVirus and Scholia interfaces	 
	English	13:00 - 16:00 	Merrilee Proffitt  & Houcemeddine Turki 	Advancing Librarianship and Scholarly Communications	 
	Português	17:00 - 20:00 	EricaAzzellini & Sturm 	Estruturação de projetos no Wikidata	 

Source: https://meta.wikimedia.org/wiki/WikiCite/2020_Virtual_conference

Other activities

WikiCite awards 23 grants & eScholarships to improve open citations

October 22, 2020 by [Liam Wyatt](#)



A pre-1941 lontar of Bali. Now in the Netherlands Tropenmuseum collection.

Source:
<https://w.wiki/ipi>

WikiCite coordination

Coordination

Project manager: [Liam Wyatt](#) (WMF contractor - [Position description](#), from November 2019)

Contact on lwatt-ctr@wikimedia.org. Reporting to [Amanda Bittaker](#) (WMF program manager)

Former project managers: [Dario Taraborelli](#) (WMF head of Research), [Elizabeth Seiver](#) (WMF contractor)

Steering committee: To contact the committee email wikicite_wikimedia.org

- [Dario Taraborelli](#)
- [Merrilee Proffitt](#)
- [Phoebe Ayers](#)
- [Alex Stinson](#)
- [Daniel Mietchen](#)

Prior to 2019, the Wikicite grant was project managed directly, within the Research team of the Wikimedia Foundation. A steering committee formed from the volunteer event management team from the 2018 Wikicite conference has served an oversight role for the new role of a part-time contractor from the community.

The steering committee is **seeking new members**. People who have a professional expertise in library metadata, especially outside of the context of Europe/North America, are particularly encouraged to apply. Please contact the project manager if you are interested.

Know someone
who would be a
great addition to
the team? Perhaps
even you?

Let us know!

Source: <https://meta.wikimedia.org/wiki/WikiCite/administration>

Daniel Mietchen - 26 October 2020 - DOI: [10.5281/zenodo.4121846](https://doi.org/10.5281/zenodo.4121846) - Q&A: <https://w.wiki/iGK>

Thanks!

Providers of open infrastructure

Presentation template by SlidesCarnival, under CC BY 4.0

Wikimedia/ Wikidata/ WikiCite community

Scholia team: Finn Nielsen, Egon Willighagen, Lane Rasberry

Alfred P. Sloan Foundation

Organizers of WikiCite events

YOU

Your turn!

Any questions or comments?

You can find me at:

- ⬡ [User:Daniel Mietchen](#)
- ⬡ [@EvoMRI](#)
- ⬡ dm7gn@virginia.edu



Backchannels for questions & comments

Please put questions and comments into the Etherpad at

<https://w.wiki/iGK>

For issues requiring a deeper discussion, consider joining the
[WikiCite group on Telegram](#).