State of WikiCite in 2020

Daniel Mietchen - 9 September 2020 - DOI: 10.5281/zenodo.4019954 - Q&A: https://w.wiki/be4

be4 we start: let's open a backchannel for questions and comments

Please put questions and comments into the Etherpad at

https://w.wiki/be4

Structure of the talk

- Hello!
- Overview of WikiCite
- WikiCite since the 2018 Workshop on Open Citations
  - Content
  - Infrastructure, tools and workflows
  - Events
- Getting involved
- Q & A

Hello!

- I am
  - Daniel Mietchen
    - User:Daniel Mietchen
  - thankful for the organization of the workshop
  - 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

Overview of WikiCite

- Community with the mission
  - “to develop open citations and linked bibliographic data to serve free knowledge” using Wikimedia platforms, especially Wikidata
- Homebase
  - on (Wikimedia) Meta: WikiCite
  - on Wikidata: WikiProject Source Metadata
    - Oversees about ⅓ of Wikidata
- Series of events
  - 2016 Berlin, 2017 Vienna, 2018 Berkeley, 2020 decentral

WikiCite reports

Annual reports

2019-2020

2018-2019

2017

2016

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

Image by Alex Lehner, CC BY 2.0

Wikidata in WikiCite numbers (i)

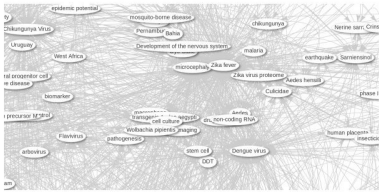
Description	Count
Triples	11,406,670,125
Citations	174,342,519
Scholarly articles	35,979,005
Items with a DOI	25,755,859
ORCIDs	1,576,769

Wikidata in WikiCite numbers (ii)

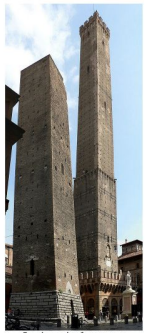
Description	Count
works to topics	11,872,474
works to people	19,091,720
people to education	1,416,674
people to employer	1,065,921

A GUIDED TOUR THROUGH CITATION NETWORKS AROUND PUBLIC HEALTH EMERGENCIES

WATCH THE VIDEO



SCHOLIA



Starting point for our tour

Contents

- 1 Abstract
- 2 Tour plan
- 3 Some impressions from the tour
- 4 Bonus pictures
- 5 About
- 6 Notes
- 7 Formats
- 8 See also

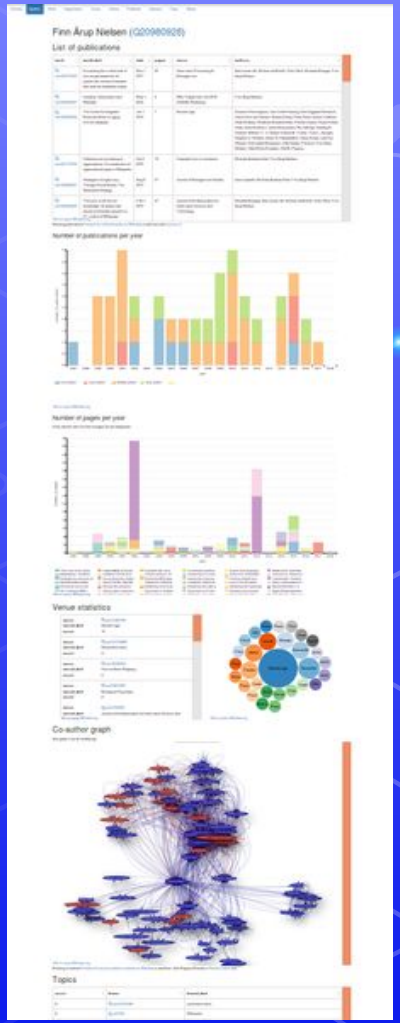
What about open citations?

My 2018 WOOC talk was centred on citation networks around public health emergencies, specifically

- Zika virus pandemic
- Chikungunya outbreaks in Italy

One way to do this 2020 talk is to change those queries to “SARS-CoV-2”, “COVID-19” or “COVID-19 pandemic”, e.g. via the Scholia profile for the SARS-CoV-2 virus, https://scholia.toolforge.org/topic/Q82069695

Source: https://www.wikidata.org/wiki/WD:WoOC2018



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

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	wd:Q24793288	Problemata Varia Anatomica: MS 1165 The University of Bologna
bologna italy	37	13	481	2	wd: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	wd:Q56991438	Persistence and distribution of pesticide residues in fresh agricultural food consumed in the province of Bologna
city of bologna	14	15	210	3	wd: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	wd:Q24793288	Problemata Varia Anatomica: MS 1165 The University of Bologna
bologna experience	9	18	162	2	wd:Q56989567	Comprehensive surgical intestinal rescue and transplantation program in adult patients: Bologna experience
pediatric clinic of bologna	6	27	162	4	wd:Q73072978	[Tuberculosis in infants in the case records of the Pediatric Clinic of Bologna]
clinica pediatrica di bologna	4	29	116	4	wd: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	wd:Q73072978	[Tuberculosis in infants in the case records of the Pediatric Clinic of Bologna]
bologna-type sausage	5	20	100	2	wd: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	wd:Q58327117	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	wd:Q64385949	[Diffusion of complement-fixing antibodies against adenovirus in the population of Bologna.]
pediatrica di bologna	4	21	84	3	wd: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	wd: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 10 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 // ne virus infection
2019-12-13	Measurement error, microcephaly prevalence and implications for Zika: an analysis of Uruguay perinatal data	Uruguay // Zika

Showing 1 to 10 of 500 entries

Topics

Co-occurring topics

Show 10 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

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. universities can disambiguate their people
6. avoiding thinking about how big the challenge is

Author disambiguation

Author name: Rosa Prato ☐ Fuzzy match ☐ Wikibase search ☐ Specify name strings **Look for author** Limit: 500

Additional SPARQL filters separated by semicolons (eg. for papers on Zika virus, enter wdt:P921 wd:Q202864): wdt:PXXX wd:QYYYYY; wdt:PXX2 wd:QYY2 ☐ 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]☑**Rosa Prato**, [6]Gaetano Lonero, [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]☑**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)

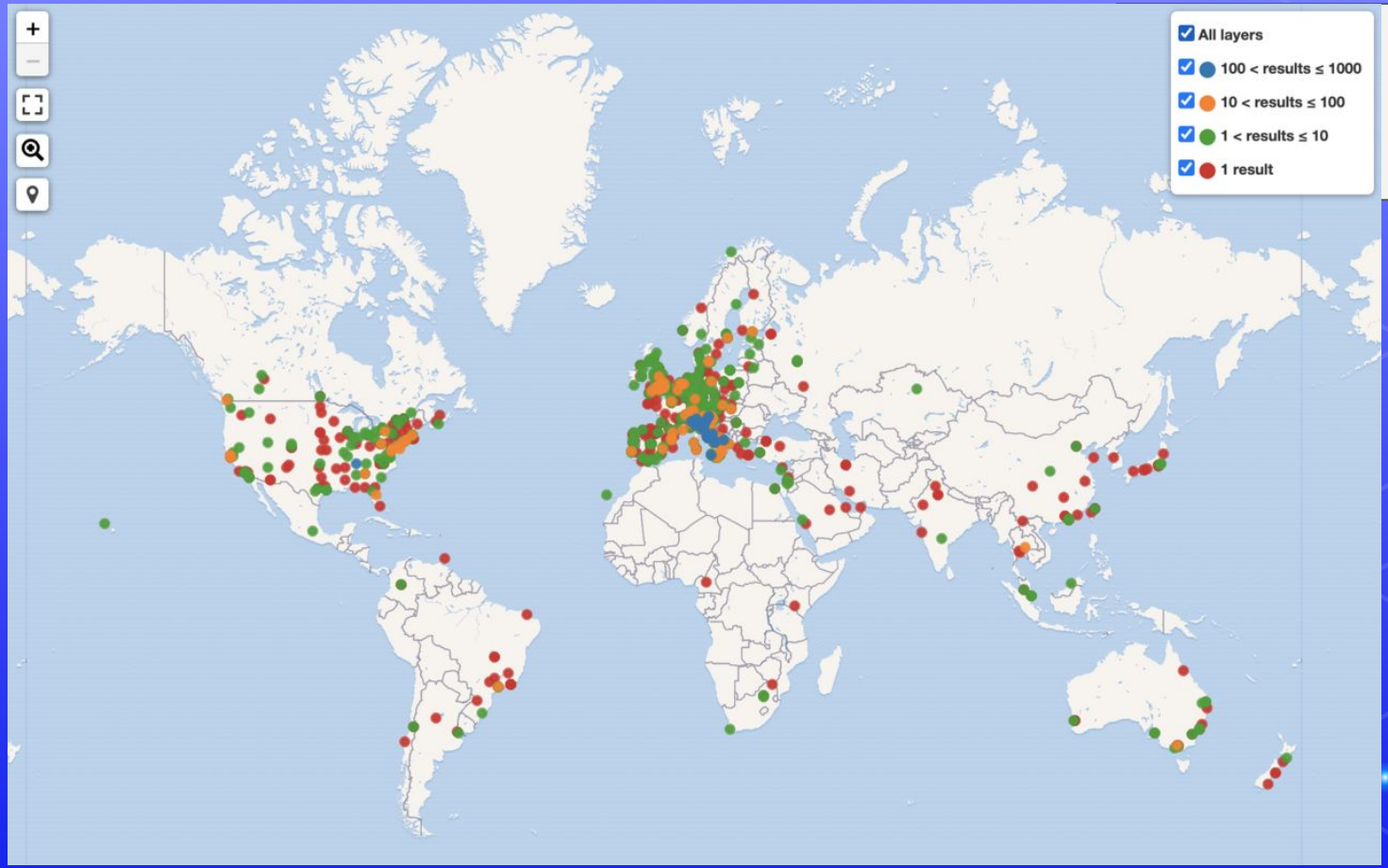
Source: https://w.wiki/bg6

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 10 entries

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

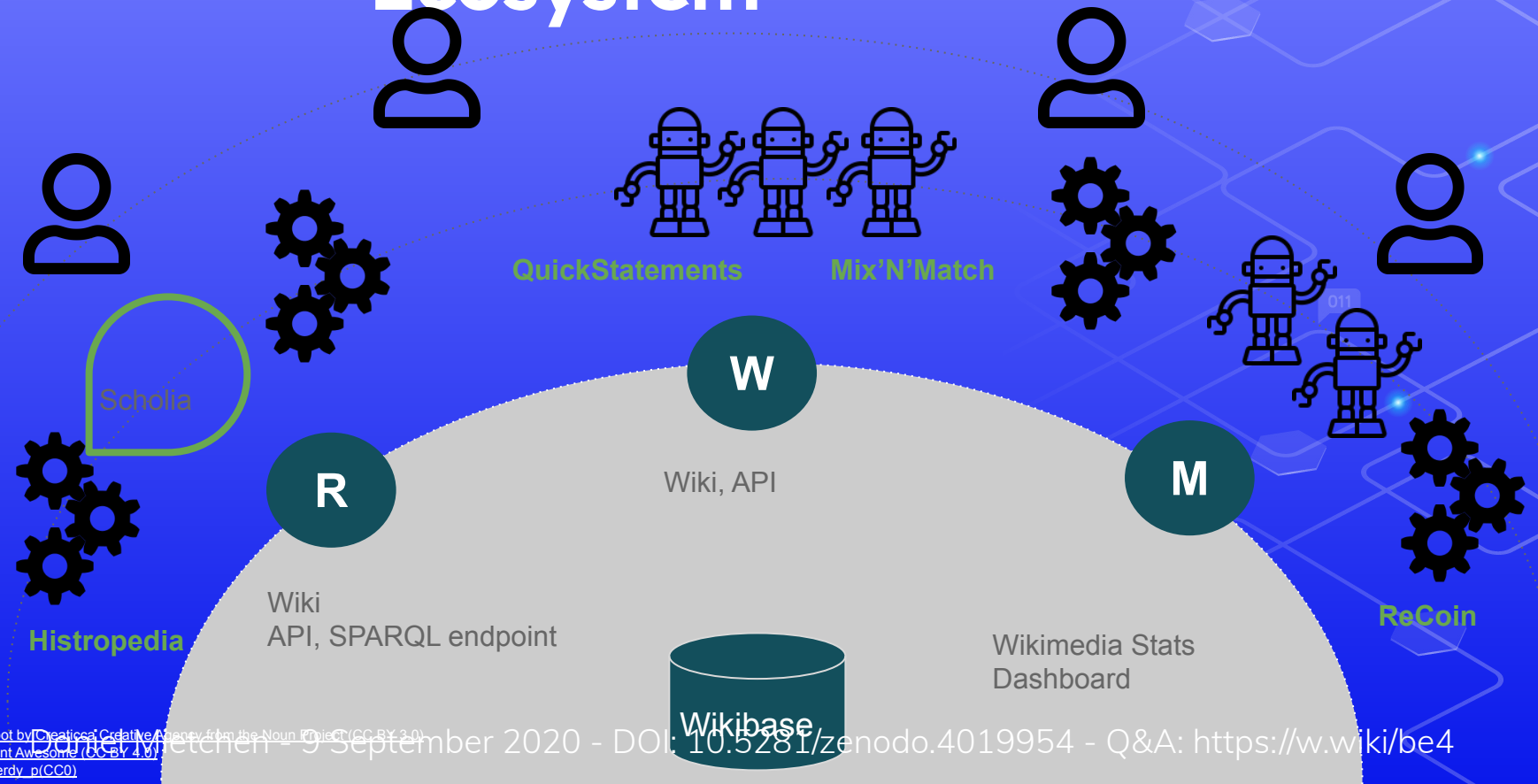
Interventions

Show 10 entries

Count	Intervention
5048	cyclophosphamide
3137	Anesthetics, Combined
3099	paclitaxel

DOI: 10.5281/zenodo.4019954 - Q&A: https://w.wiki/be4

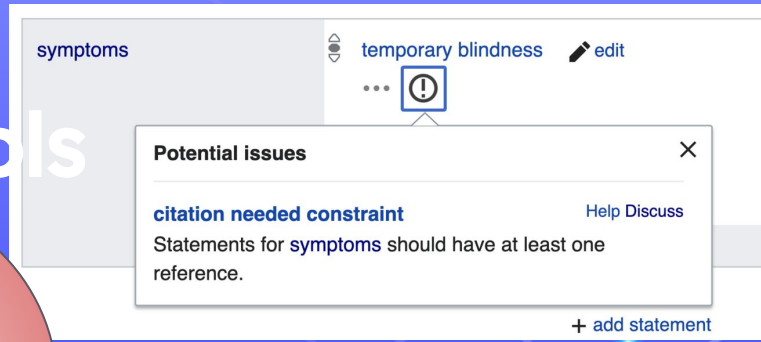
Tools in the Wikidata Ecosystem

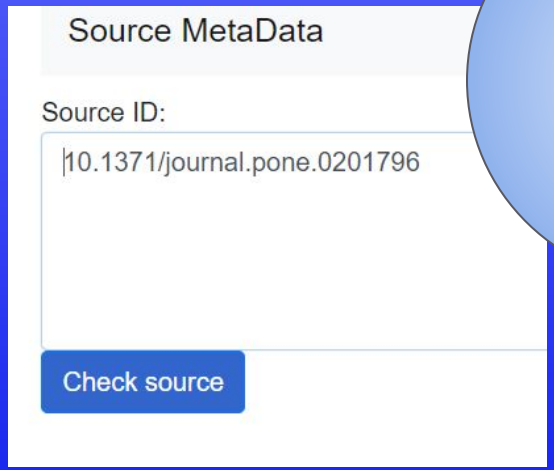


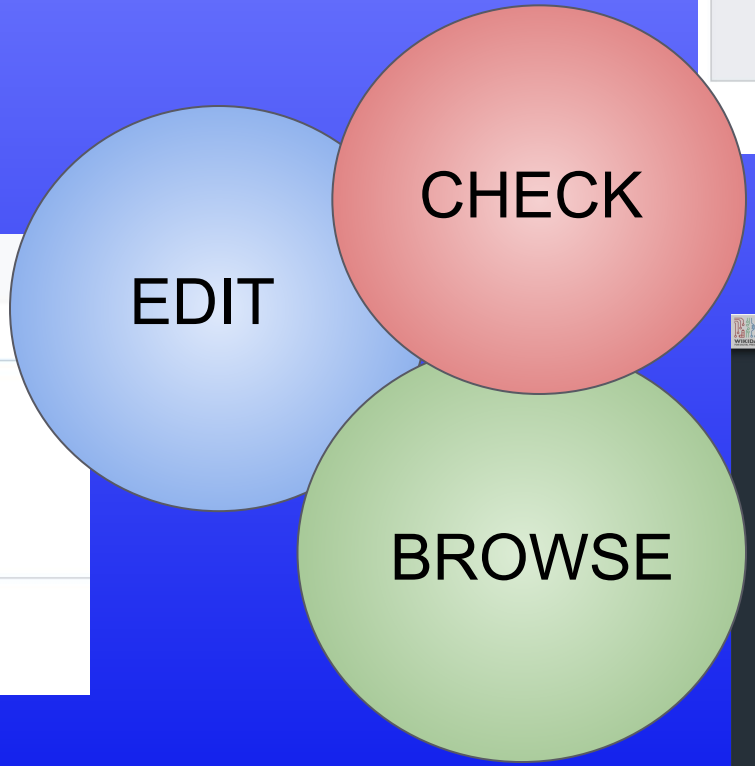
Bot icon by Robot by Creaticca Creative Agency from the Noun Project (CC BY 3.0)
User icon by Font Awesome (CC BY 4.0)
Tools icon by Verdy_p(CC0)

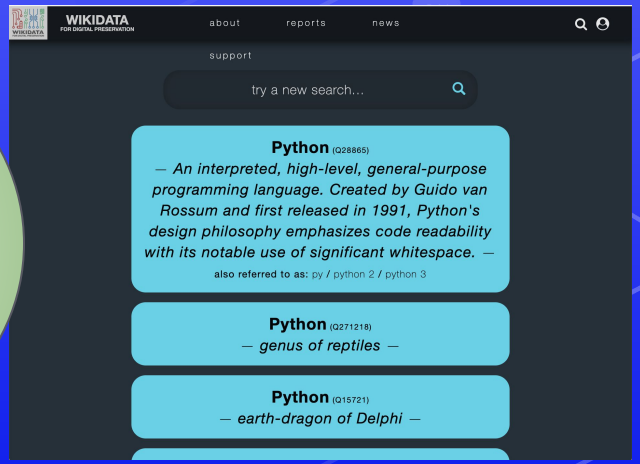
Daniel Mietchen - 9 September 2020 - DOI: 10.5281/zenodo.4019954 - Q&A: https://w.wiki/be4

Overview of Wikidata tools



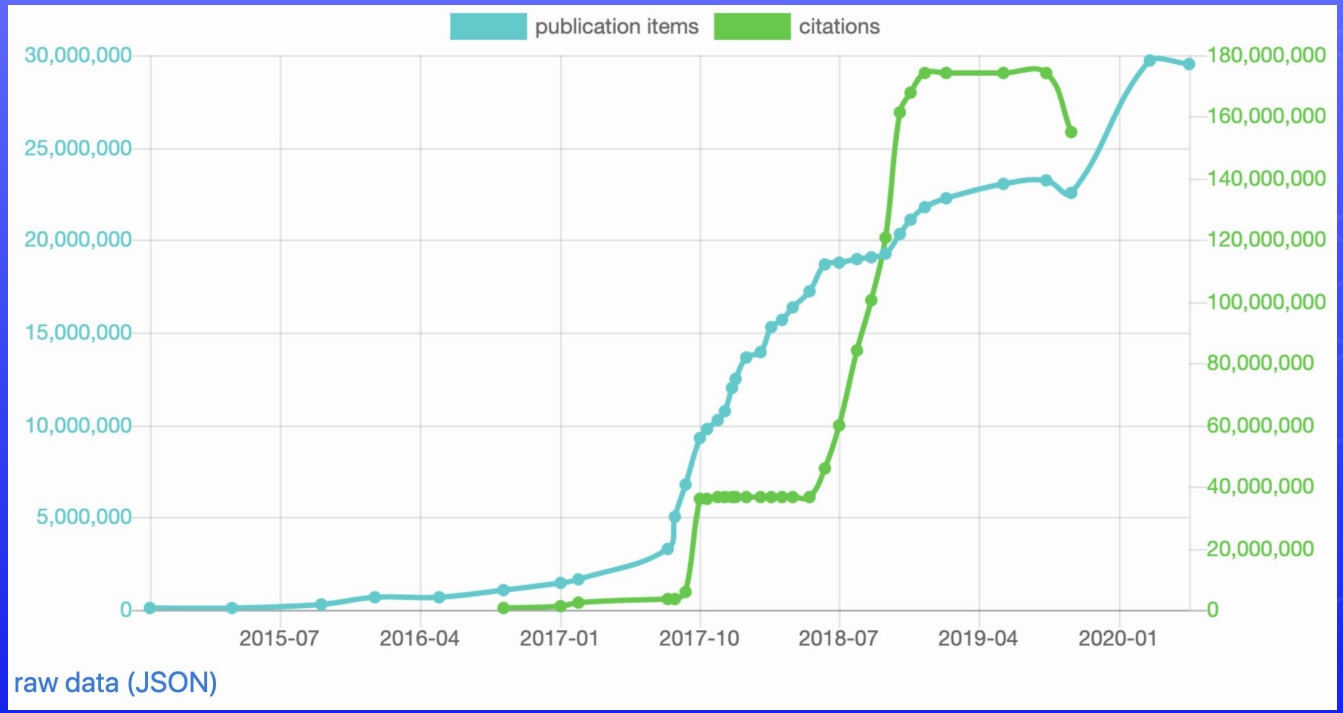






https://www.wikidata.org/wiki/Wikidata:Tools

What about open citations?



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

Bots

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

WikiCite events - history

2018 conference - Berkeley

2017 conference - Vienna

{} wikicite

Berlin, 25-26 May 2016

2016 conference - Berlin

Events - today's workshop

Workshop On Open Citations And Open Scholarly Metadata 2020 (Q98918134)

People

Show 10 entries Search:

Works	Person	Person description	Roles	Example work
91	Silvio Peroni	Italian computer scientist	speaker, organizer, participant	Setting our bibliographic references free: towards open citation data
90	Cameron Neylon	Australian biophysicist, technologist and open access advocate	speaker, participant	Panton Principles
89	Daniel Mietchen	German biophysicist and open science advocate, Wikimedian	speaker, participant	Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity information
64	David Shotton	British biologist and bioinformatician	speaker, organizer, participant	Adventures in semantic publishing: exemplar semantic enhancements of a research article
50	Ludo Waltman	researcher in scientometrics	speaker, organizer, participant	Exploring the relationship between the engineering and physical sciences and the health and life sciences by advanced bibliometric methods

https://scholia.toolforge.org/event/Q98918134

Upcoming activities - your turn!

WikiCite grants and e-scholarships

The steering committee with the financial support of the Alfred P. Sloan Foundation and the Wikimedia Foundation, are currently offering two categories of grants for projects which support the goals of WikiCite.

For both categories of grants the **application deadline** is 1 October 2020, and all projects must be **completed** by 1 May 2021.

Source: https://w.wiki/bgD

WikiCite grants

Project & Event grants [edit]

WikiCite grants between $2,000 and $10,000 (USD equivalent) are available to individuals, groups, and organisations with a project that supports the goals of Wikicite.

All the details, the eligibility criteria, and the application form are available on the **WikiCite project & event grants** homepage.

Individuals, groups, and organizations may apply, and projects may be of any nature. This includes technical (e.g. software, tools), event (online, or in-person), resources (training materials, documentation), or other forms not mentioned – as long as it supports the goals of WikiCite.

Note: Applications for *in-person events are eligible*. The WMF will release a risk assessment calculator for in-person events and a checklist of precautions in early September.[1] This protocol will apply to all Wikimedia Foundation supported events. Applicants may publish a draft of their proposals now, and edit it before the submission deadline to include a copy of the results from the risk assessment protocol once it is published.

Source: https://w.wiki/bgD

WikiCite scholarships

e-Scholarships [edit]

The e-scholarship program is a new kind of grant in Wikimedia, created in response to an era of COVID-19 quarantines, and the 2030 Movement strategy goals.

An e-scholarship provides a per-diem equivalent allowance for 1-5 people to stay at their home(s) and work for 2-4 days on a project supporting the mission of WikiCite. e-scholarship recipients' projects can be the kinds of things they might have previously undertaken at an in-person hackathon, unconference, or research trip.

All the details, eligibility criteria, program principles, and the application form are available on the **WikiCite e-scholarships** homepage.

Funding will:

- be provided in advance;
- be calculated at the WMF per-diem rate for the city where the e-scholarship recipient lives;
- and – as it is a living allowance – *not* require recipients to submit expense reports.

"Remote group" applications are encouraged, as are projects which focus on content or communities which are historically underrepresented in Wikimedia projects. Building a bot, fixing a tool, wrangling a dataset, writing complete documentation... all are valid e-scholarship projects. A confirmation letter (in advance) and/or participation certificate (afterwards) can also be provided.

Source: https://w.wiki/bgD

WikiCite scholarships

e-Scholarships [edit]

The e-scholarship program is a new kind of grant in Wikimedia, created in response to an era of COVID-19 quarantines, and the 2030 Movement strategy goals.

An e-scholarship provides a per-diem equivalent allowance for 1-5 people to stay at their home(s) and work for 2-4 days on a project supporting the mission of WikiCite. e-scholarship recipients' projects can be the kinds of things they might have previously undertaken at an in-person hackathon, unconference, or research trip.

All the details, eligibility criteria, program principles, and the application form are available on the **WikiCite e-scholarships** homepage.

Funding will:

- be provided in advance;
- be calculated at the WMF per-diem rate for the city where the e-scholarship recipient lives;
- and – as it is a living allowance – *not* require recipients to submit expense reports.

"Remote group" applications are encouraged, as are projects which focus on content or communities which are historically underrepresented in Wikimedia projects. Building a bot, fixing a tool, wrangling a dataset, writing complete documentation... all are valid e-scholarship projects. A confirmation letter (in advance) and/or participation certificate (afterwards) can also be provided.

Source: https://w.wiki/bgD

WikiCite coordination

Coordination

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

Contact on *lwyatt-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
- Elizabeth Seiver
- 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

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

The organizers of the workshop

YOU

Daniel Mietchen - 9 September 2020 - DOI: 10.5281/zenodo.4019954 - Q&A: https://w.wiki/be4

Your turn!

Any questions or comments?

You can find me at:

- User:Daniel Mietchen
- @EvoMRI
- dm7gn@virginia.edu

be4 we start: a backchannel for questions and comments

Please put questions and comments into the Etherpad at

https://w.wiki/be4