

[*citations needed*] for the sum of all human knowledge

AAAS 2017 • February 18, 2017

Dario Taraborelli

@readermeter



What's the single, most important
ingredient of a Wikipedia article?

A Text

 Images

 Videos

 Table of contents

 Internal links

 External links

 Interlanguage links

 Infoboxes

 Maps

 Charts

 References

 Categories

A

Text



Images



Videos



Table of contents



Internal links



External links

A

Interlanguage links



Infoboxes



Maps



Charts



References



Categories

Wikipedia:Verifiability, not truth

From Wikipedia, the free encyclopedia



This [essay](#) contains the [advice](#) or opinions of one or more Wikipedia contributors on the [Verifiability](#) policies. Essays are *not* [Wikipedia policies](#) or [guidelines](#). Some essays represent widespread norms; others only represent minority viewpoints.



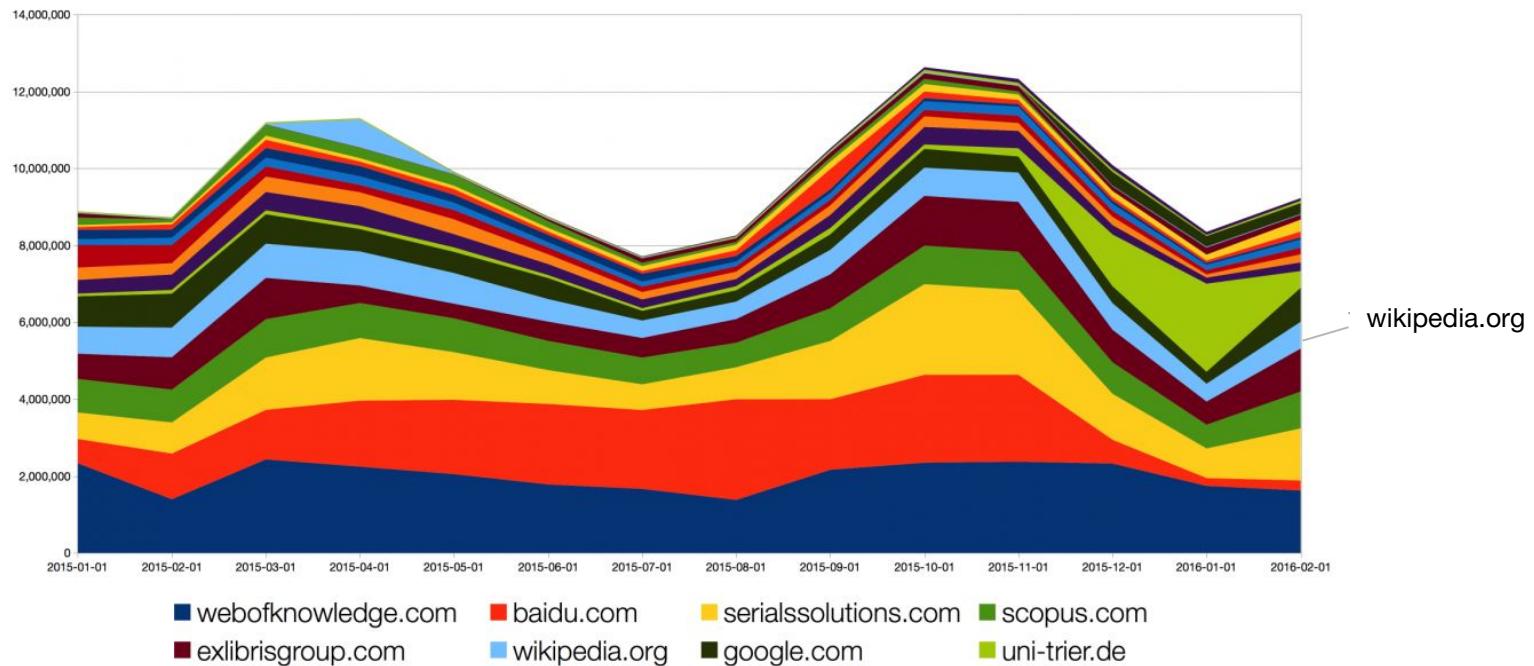
This page in a nutshell: Any material added to Wikipedia must have been published previously by a [reliable source](#). Editors may not add content solely because they believe it is true, nor delete content they believe to be untrue, unless they have verified beforehand with a [reliable source](#).

Shortcuts:
[WP:NOTTRUTH](#)
[WP:TRUTH](#)
[WP:VNT](#)

Wikipedia's core sourcing policy, [Wikipedia:Verifiability](#), used to define the threshold for inclusion in Wikipedia as "**verifiability, not truth**". "Verifiability" was used in this context to mean that material added to Wikipedia must have been published previously by a [reliable source](#). Editors may not add their own views to articles simply because they believe them to be correct, and may not remove sources' views from articles simply because they disagree with them.



Top referrals to the literature (DOI lookups)



<http://crosstech.crossref.org/2014/02/many-metrics-such-data-wow.html>

<http://blog.crossref.org/2016/05/https-and-wikipedia.html>

Preferred resource for knowledge acquisition

Results: The participants valued lecture notes (73.7%) and Wikipedia (74%) as their most important online sources for knowledge acquisition. Missing integration of e-learning into teaching was seen as the major pitfall (58.7%). The learning management system was mostly used for study information (68.3%), preparation of exams (63.3%) and lessons (54.5%). Clarity (98.3%), teaching-related contexts (92.5%) and easy use of e-learning offers (92.5%) were rated highest. Interactivity was most important in free-text comments (n = 123).

Back et al. (2016) *Int J Med Educ.* 7:267-273; doi: [10.5116/ijme.57a5.f0f5](https://doi.org/10.5116/ijme.57a5.f0f5)

Most visited resource during public health crises

Most used internet site in Liberia,
Sierra Leone and Guinea for Ebola
during 2014 outbreak

Greater than CNN, CDC and WHO

Heilman (2016) tinyurl.com/jfuyduv





average goat lifespan

Web

Shopping

Images

Videos

News

More ▾

Search tools

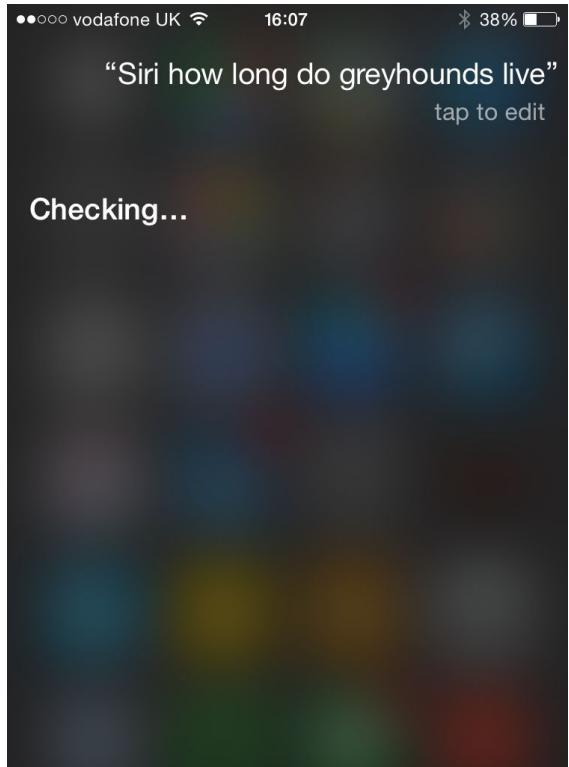
About 367,000 results (0.39 seconds)

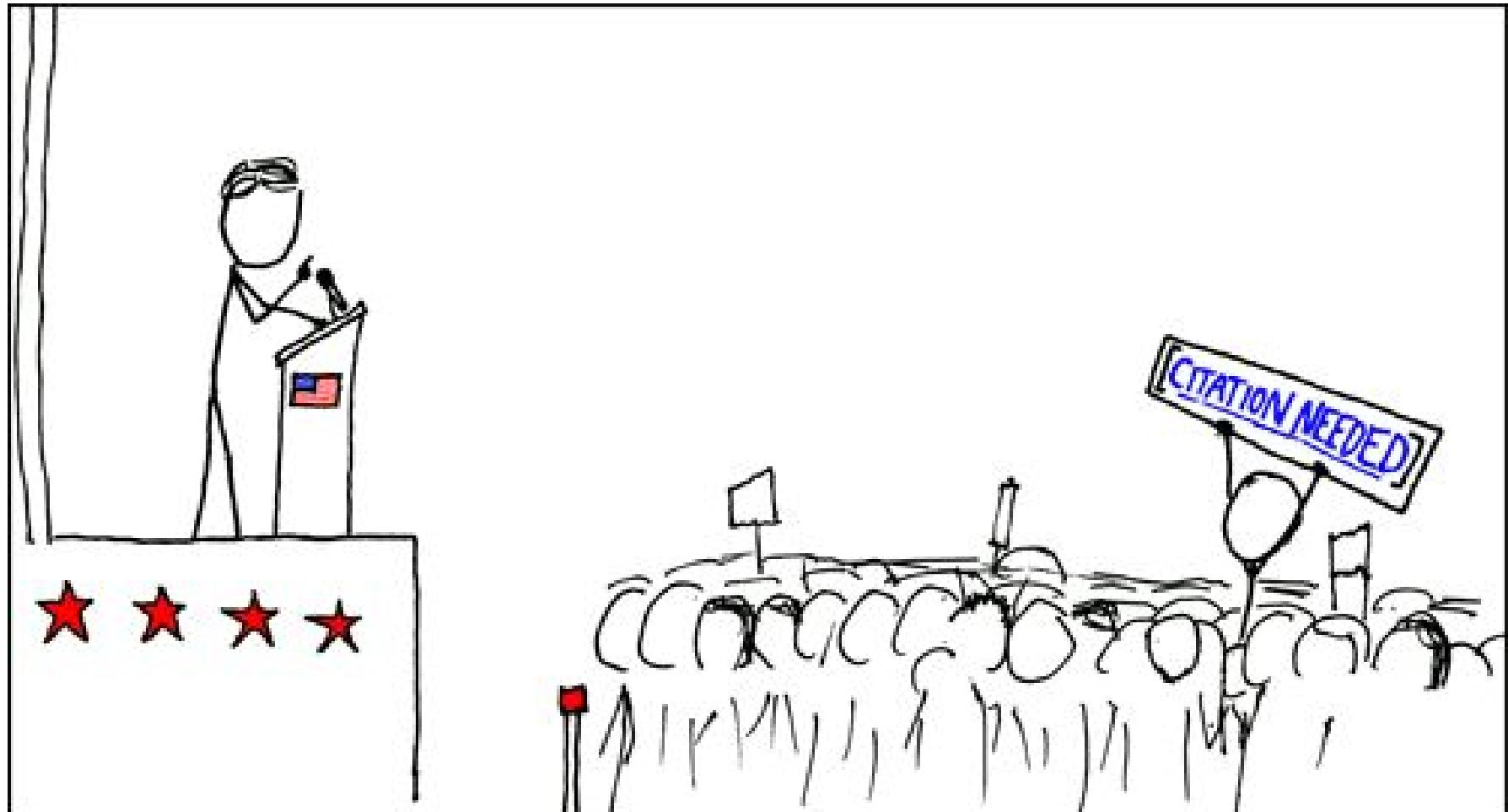
15 – 18 y

Goat, Lifespan



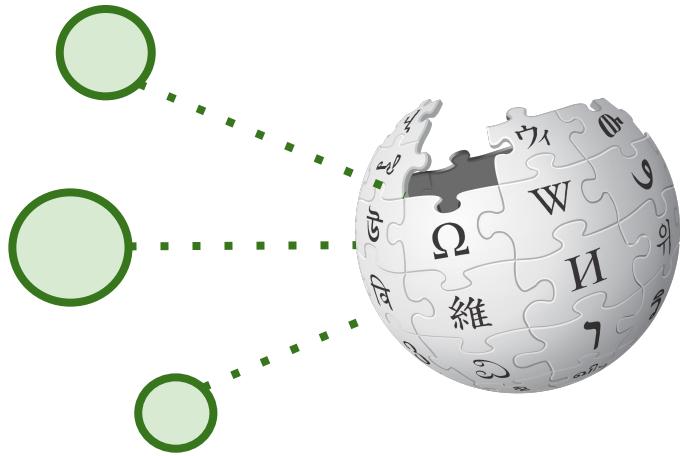
Feedback

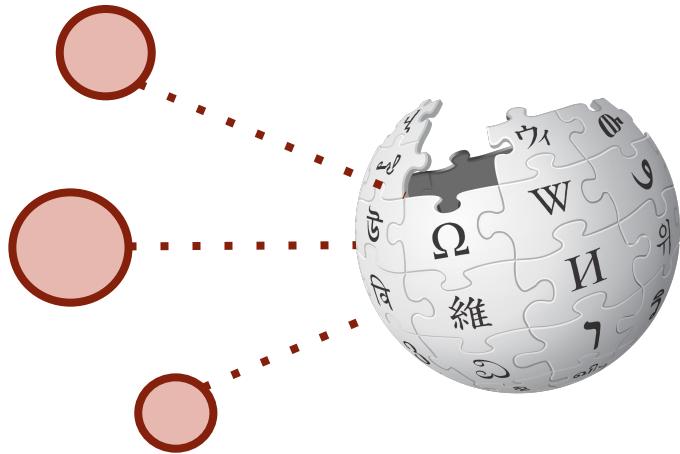




Randall Munroe, *Wikipedian protester* <http://tinyurl.com/p3rodlb> [CC BY]







References in Wikipedia

The molecular origins of insulin go at least as far back as the simplest unicellular [[eukaryotes]].<ref name='LeRoith'>{{cite journal | vauthors = LeRoith D, Shiloach J, Heffron R, Rubinovitz C, Tanenbaum R, Roth J | title = Insulin-related material in microbes: similarities and differences from mammalian insulins | journal = Can. J. Biochem. Cell Biol. | volume = 63 | issue = 8 | pages = 839–49 | year = 1985 | pmid = 3933801 | doi = 10.1139/o85-106 }}</ref> Apart from animals, insulin-like proteins are also known to exist in Fungi and Protista kingdoms.

A bibliographic database to
serve the sum of all knowledge?



WIKIDATA

Vision
Technology
Community
Scale
Licensing
Independence

Free knowledge base *that anyone can edit*

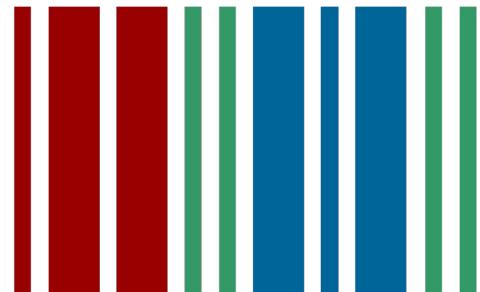
Launched in 2012

Integrated with Wikipedia and other sister projects

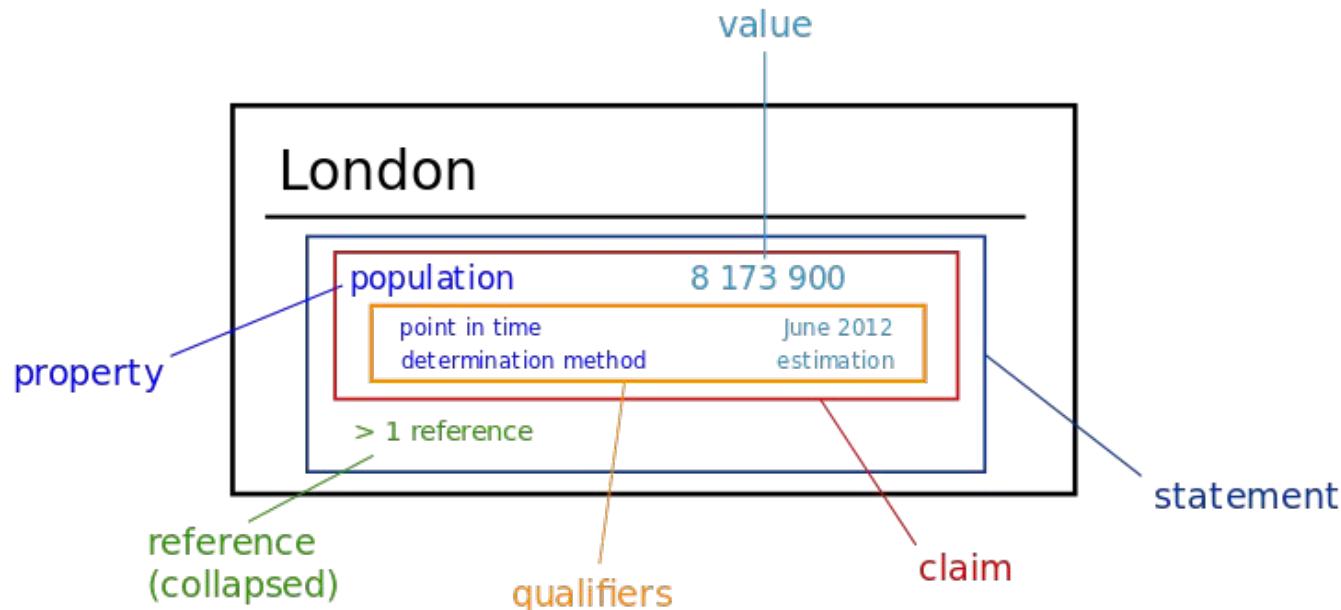
Statistics (February 2017)

Over 25M items

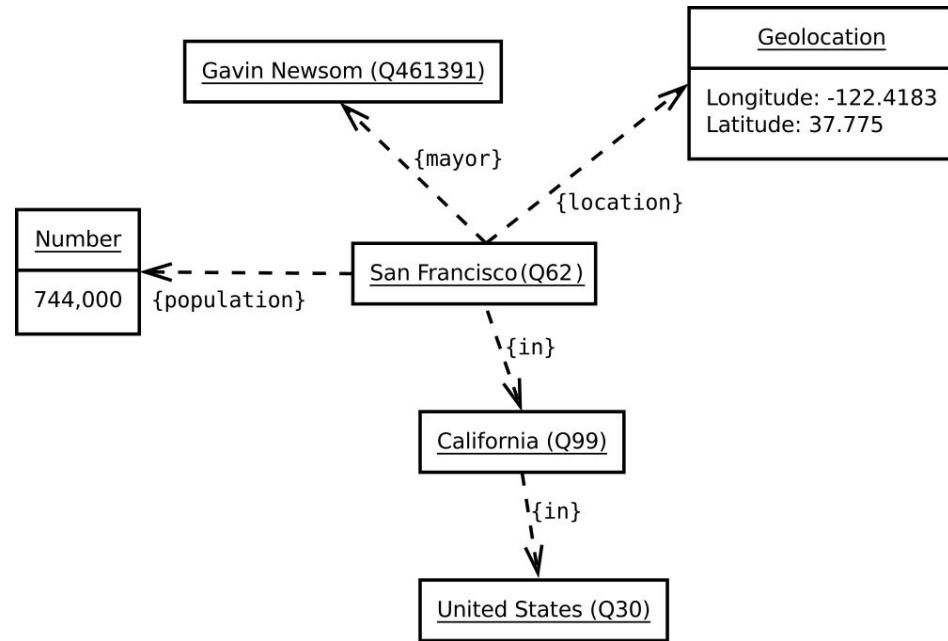
Over 130M statements



Wikidata's anatomy



Wikidata's anatomy



Scientific open data in Wikidata

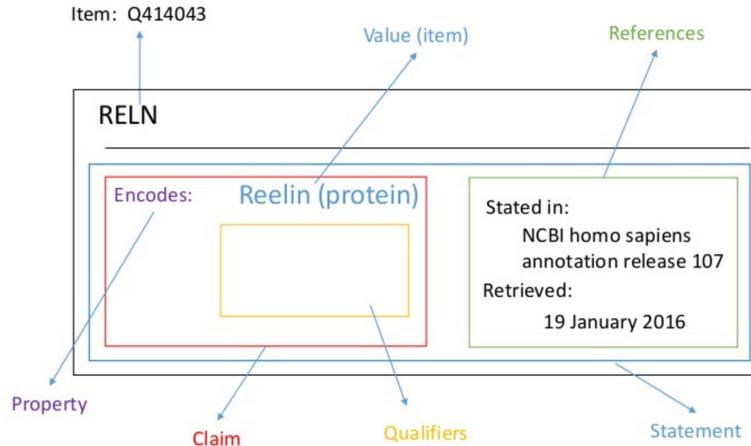


- All human, mouse genes and proteins (swissprot)
- All Gene Ontology terms
- All Human Disease Ontology terms
- All FDA approved drugs
- 109 reference microbial genomes

Mitraka et al (2015) *Semantic Web Applications for the Life Sciences*
Burgstaller-Muelbacher et al (2016) *Database*
Putman et al (2016) *Database*

Scientific open data in Wikidata

Linking the Reelin gene to a protein it encodes



Benjamin Good (2016) *Opportunities and challenges presented by Wikidata in the context of biocuration* <http://tinyurl.com/hk9qrmz>



Egon Willighagen
@egonwillighagen



Following

(in five years the verb "to wikidata" means look up a fact with literature provenance) "let me wikidata that for you"

RETWEETS

5

LIKES

8



9:25 AM - 8 Apr 2016

<https://twitter.com/egonwillighagen/status/718474906858582016>

Let's build a bibliographic database
in Wikidata to serve the sum of all
knowledge!

{ } wikicite



Alfred P. Sloan
FOUNDATION

{ } wikicite

Berlin, 25-26 May 2016



https://meta.wikimedia.org/wiki/WikiCite_2016

Rich bibliographic data models

The relative antibacterial activity of four penicillins (Q24685088)

scientific article

instance of: The relative antibacterial activity of four penicillins is a(n) [scientific article](#)

Statements	
page(s)	1695-6
volume	2
issue	5214
short author name	L P Garrod <small>series ordinal : 1</small>
title	The relative antibacterial activity of four penicillins [en]
published in	British Medical Journal (peer-reviewed medical journal)
publication date	1960-12-10
original language of work	English (West Germanic language originating in England)

<https://www.wikidata.org/wiki/Q24685088>



Active scholarly journals worldwide (2014):
34,000.

Broad coverage of scholarly journals

In English: 28,100.

In **@Wikidata**: 23,900.

Indexed in **@thomsonreuters** JCR: 10,900.

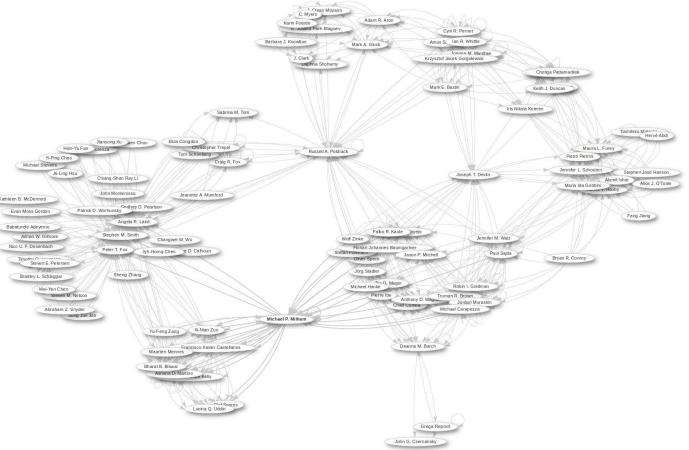
RETWEETS LIKES
16 **12**



11:36 PM - 26 Oct 2016

twitter.com/Wikicite/status/791529065198915585

2 million citation links between scholarly papers



The Zika corpus

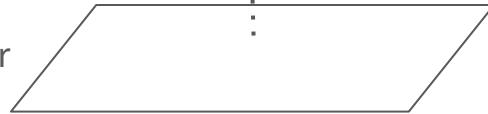
Encyclopedic layer



Expert annotation layer



Bibliographic metadata layer



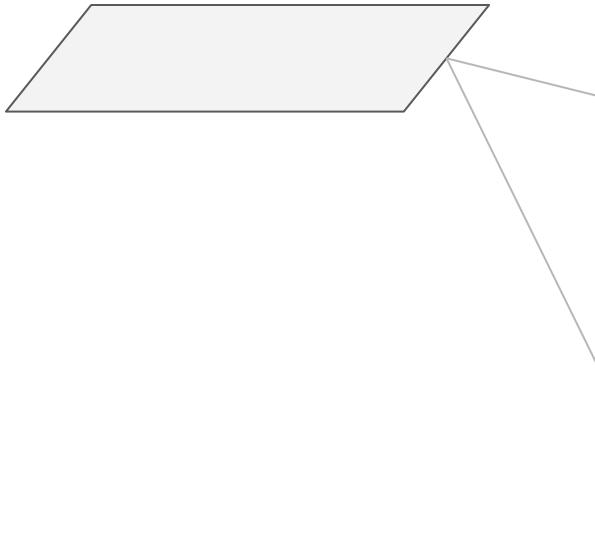
Open citation graph layer



meta.wikimedia.org/wiki/WikiCite_2016/Report#The_Zika_corpus

The Zika corpus

Encyclopedic layer

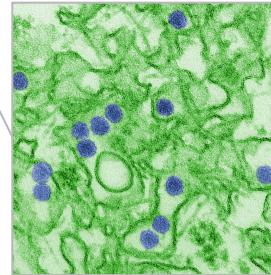


Zika virus

From Wikipedia, the free encyclopedia

This article is about the virus. For the disease, see [Zika fever](#). For the current outbreak, see [2015–16 Zika virus epidemic](#).

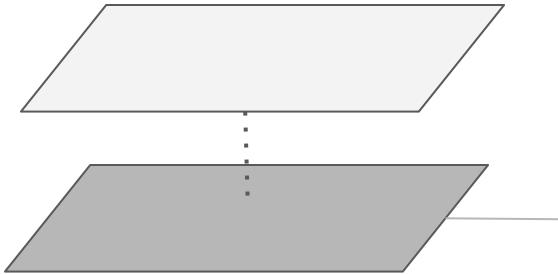
Zika virus (ZIKV) is a member of the *Flaviviridae* and the genus *Flavivirus*.^[3] It is spread by daytime-active *Aedes* mosquitoes, such as *A. aegypti* and *A. albopictus*.^[3] Its name comes from the [Zika Forest](#) of Uganda, where the virus was first isolated in 1947.^[4] Zika virus is related to the dengue, yellow fever, Japanese encephalitis, and West Nile viruses.^[4] Since the 1950s, it has been known to occur within a narrow equatorial belt from Africa to Asia. From 2007 to 2016, the virus spread eastward, across the Pacific Ocean to the Americas, leading to the 2015–16 Zika virus epidemic.



The Zika corpus

Encyclopedic layer

Expert annotation layer



pathogen transmission process

mosquito borne transmission (type of insect borne pathogen transmission)

Reference

stated Concurrent outbreaks of dengue, chikungunya and Zika virus infections – an unprecedented epidemic in wave of mosquito-borne viruses in the Pacific 2012–2014 (scientific article)

contact transmission (type of direct pathogen transmission)

placental transmission (type of congenital pathogen transmission process)

Reference

stated Zika virus damages the human placental barrier and presents marked fetal neurotropism (scientific in article)

sexual intercourse (insertion of a male's penis into a female's vagina for the purposes of sexual pleasure, reproduction, or both)

Reference

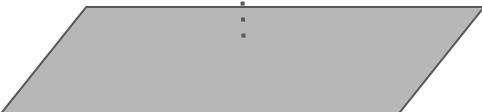
stated in Potential sexual transmission of Zika virus (scientific article)

The Zika corpus

Encyclopedic layer



Expert annotation layer



Bibliographic metadata layer



Potential sexual transmission of Zika virus (Q22330722)

scientific article

Instance of: Potential sexual transmission of Zika virus is a(n) scientific article

Statements	
	Own statements
page	359-61
volume	21
issue	2
cites	Zika virus. I. Isolations and serological specificity (1952 scientific article) Rapid spread of emerging Zika virus in the Pacific area (scientific article) Zika Virus, French Polynesia, South Pacific, 2013
published in	Emerging Infectious Diseases (medical journal)
title	Potential sexual transmission of Zika virus [en]
publication date	2015-02
original language of work	English (West Germanic language originating in England)
main subject	Zika virus (species of virus) Zika fever (infectious arboviral disease)
author	Didier Musso (researcher) series ordinal : 1 Van-Mai Cao-Lormeau (virologist) series ordinal : 6 Anita Tellesser (epidemiologist) series ordinal : 5

Identifiers

PMCID	4313657 ↗
PubMed ID	25625872 ↗
DOI	10.320...141363 ↗

The Zika corpus

Encyclopedic layer



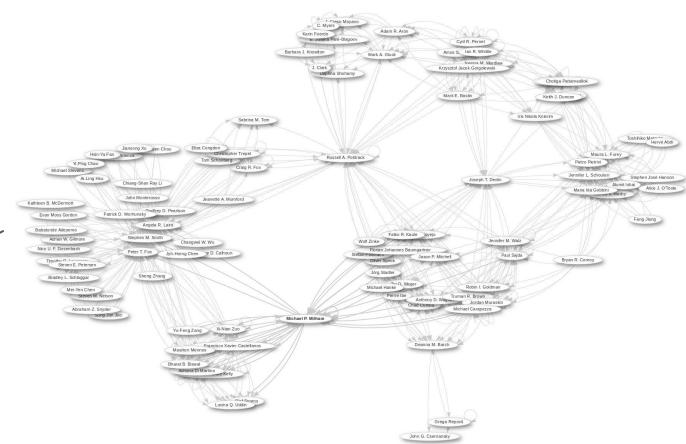
Expert annotation layer



Bibliographic metadata layer



Open citation graph layer



all statements citing a *New York Times* article

most popular journals cited by statements of any item that is a subclass of *economics*

all statements citing the works of *Joseph Stiglitz*

all statements citing journal articles by *physicists at Oxford University in the 1970s*

all statements citing a journal article that was *retracted*

Answer complex questions involving sources and statements via SPARQL queries

meta.wikimedia.org/wiki/WikiCite_2016/Report/Group_5

Zika virus • [Q202864](#)

TAXON

has natural reservoir • [P1605](#)

Aedes hensilli • [Q14573674](#)

TAXON

stated in • [P248](#)

Aedes hensilli as a potential vector of Chikungunya and Zika viruses • [Q22330738](#)

SCIENTIFIC ARTICLE

funded by • [P859](#)

Centers for Disease Control and Prevention • [Q583725](#)

GOVERNMENT AGENCY

published in • [P1433](#)

PLOS Neglected Tropical Diseases • [Q3359737](#)

SCIENTIFIC JOURNAL

publisher • [P123](#)

Public Library of Science • [Q233358](#)

PUBLISHER

Represent and analyze links between statements, authors, scientific outlets and funders



The background image is a wide-angle aerial photograph of Vienna, Austria, during sunset. The city is illuminated by the warm light of the setting sun, which casts a golden glow over the buildings and the Danube River. The river curves through the city, with several bridges spanning it. In the foreground, there's a green hillside with vineyard terraces. The sky is a mix of orange, yellow, and blue.

{ } wikicite

Vienna, May 23-25 2017

applications open through **February 27**

https://meta.wikimedia.org/wiki/WikiCite_2017

Thank you

Acknowledgments

Daniel Mietchen, Jonathan Dugan, Lydia Pintscher, Cameron Neylon, James Hare, James Heilman, Magnus Manske, Egon Willighagen, the Gene Wiki team (especially Andrew Su, Andra Waagmeester, Tim Putman, Benjamin Good), the ContentMine team, the University of Chicago Knowledge Lab; all WikiCite 2016 participants and Wikidata Source Metadata project contributors; the WikiCite initiative funders, in particular: Carly Strasser, Josh Greenberg, Greg Boustead, Ginny Hendricks.

Image credits

Thomas Fabian, Vienna flickr.com/photos/126875359@N03/28210887335 [CC BY SA]