

Wikidata for university profiling

A decorative graphic consisting of multiple parallel, wavy lines of small blue dots, creating a sense of motion and data flow across the background.

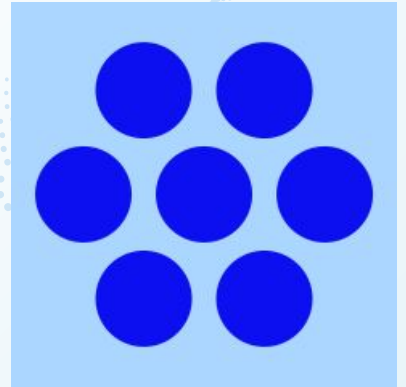
Lane Raspberry

user:blueraspberry

Wikimedian in Residence

School of Data Science

University of Virginia



University librarians and communication staff!

Wikidata makes the world aware of research and publications at your university

Share research metadata in the Wikipedia platform for promotion!

Wikipedia promotion opportunity

Share this data in Wikidata:

- list of faculty
- list of research publications
- awards, grants, and funding received
- major research projects

Discussion outline

1. How should people understand universities?
2. the data: WikiCite and more
3. Solution: scholarly profiles
4. Social and ethical issues



How can the Wikipedia
platform describe
universities?



WIKIPEDIA
The Free Encyclopedia

[Main page](#)

[Contents](#)

[Current events](#)

[Random article](#)

[About Wikipedia](#)

[Contact us](#)

[Donate](#)

[Contribute](#)

[Help](#)

[Learn to edit](#)

[Community portal](#)

[Recent changes](#)

[Upload file](#)

Article

[Talk](#)

University of Virginia

From Wikipedia, the free encyclopedia

"UVA" redirects here. For other uses, see [Uva](#).

The **University of Virginia** (**U.Va.** or **UVA**) is a [public research university](#) in [Charlottesville, Virginia](#), for [Jefferson](#). It is the [flagship university](#) of Virginia and home to the [Academical Village](#), a [UNESCO World Heritage Site](#) known for [its historic foundations](#), [student-run honor code](#), and [secret societies](#).

The original governing Board of Visitors included Jefferson, [James Madison](#), and [James Monroe](#). Monroe was the President of the United States at the time of its foundation, and earlier presidents Jefferson and Madison were [rectors](#). Jefferson conceived and designed the [original courses](#) of study and [original architecture](#). UVA is a research-driven [Association of American Universities](#) member for 117 years, and the journal *Science* recently ranked it as one of the top ten global scientific breakthroughs in a single year (2015).^[10]

The University of Virginia has graduated 55 [Rhodes Scholars](#),^[11] eighth-most in the United States, and many companies (such as [Reddit](#), [CNET](#), and [Space Adventures](#)) which together produce more than \$1 billion in revenue and have created 2.3 million jobs.^[12] UVA offers 121 majors across the eight undergraduate schools.

Wikipedia articles

- encyclopedia article
- limited to what people have written in text
- data maybe in infobox, or maybe no data



Histoire et statut

Fondation	1819
Type	Université publique
Nom officiel	<i>The Rector and Visitors of the University of Virginia</i>
Régime linguistique	Anglais américain
Fondateur	Thomas Jefferson
Président	Teresa A. Sullivan
Recteur	James B. Murray Jr. (d) ✎
Membre de	AAU, Universitas 21
Site web	www.virginia.edu 🌐

Chiffres-clés

Étudiants	21 000 ✎
Effectif	15 216 (2019) ✎
Enseignants	2 102
Budget	4,45 milliards \$

Data can tell stories with lists

For one university...

- Faculty
- Research publications
- Awards
- Research projects
- Medical research
- funding and grants

For multiple universities...

- co-author networks
- trends in receiving awards and funding
- trends in research management
- demographic trends

Stories to tell

- What research papers did my university publish on COVID last year?
- What unusual research speciality does a given university have?
- How diverse are the faculty in a university department?
- Show a map of which universities research a certain topic

WikiCite as a foundational collection

{ } wikicite

Smith, Denise A. (18 February 2020). ["Situating Wikipedia as a health information resource in various contexts: A scoping review"](#). *PLOS ONE*.

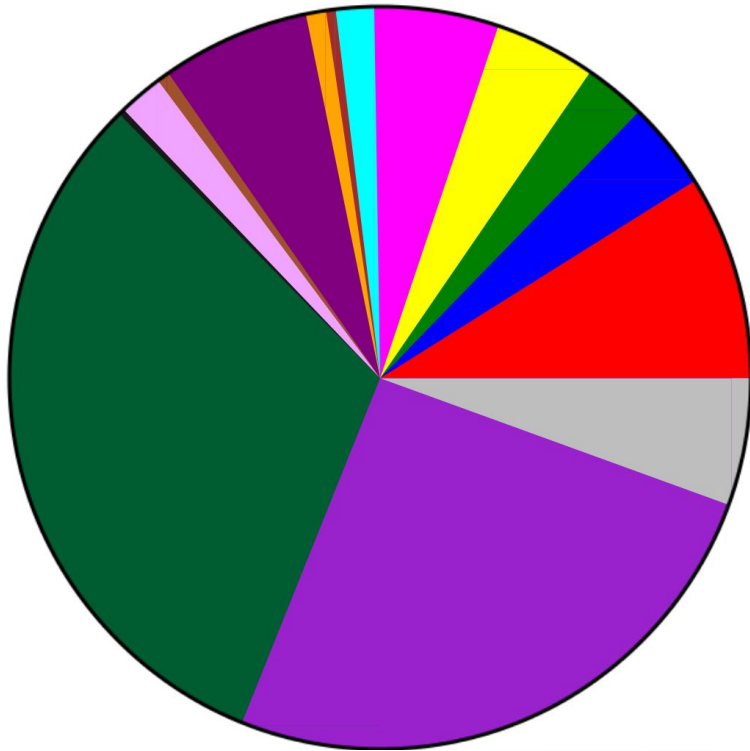
- collection of citation metadata
- currently we have 37 million out of 200 million
- imported from free external datasets
- single most popular Wikidata project?
- connects topics, people, places, and more

Citation metadata

Smith, Denise A. (18 February 2020). ["Situating Wikipedia as a health information resource in various contexts: A scoping review"](#). *PLOS ONE*.

- person
 - has institutional affiliations
 - demographic characteristics
 - publications are accomplishments
- paper / journal
 - has topics

Wikidata:statistics, of 71 million items...



WikiCite

miscellaneous

???

people

taxon

territory

structure

event

Scholia is a
scholarly profiling
solution



WIKIPEDIA

Other scholarly profiling tools



Google
Scholar



SEMANTIC SCHOLAR



Scopus®



WEB OF SCIENCE™



Microsoft Academic

Wiki tools must be like this

- transparently developed
- credit contributors
- FAIR and open
- high community engagement
- highly interdisciplinary
- identify and address biases
- innovate with ethical issues
- multilingual

Quick critiques: all non-free, plus...

Google
Scholar

- no data export button

SEMANTIC SCHOLAR

- users cannot innovate in

Scopus[®]

- Elsevier / Cost of Knowledge campaign

WEB OF SCIENCE™

- US\$100,000+ year?

Microsoft Academic

- steals wiki's hamster

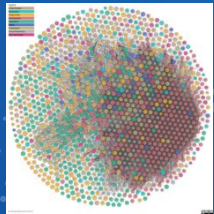


Hamsters are rodents (order Rodentia) belonging to the subfamily Cricetinae, which contains 19 species classified in

Scholia is a wiki tool for scholarly profiling



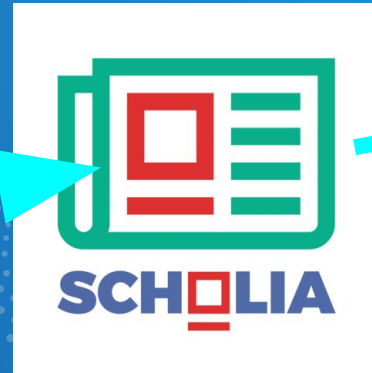
WIKIPEDIA



LOD cloud



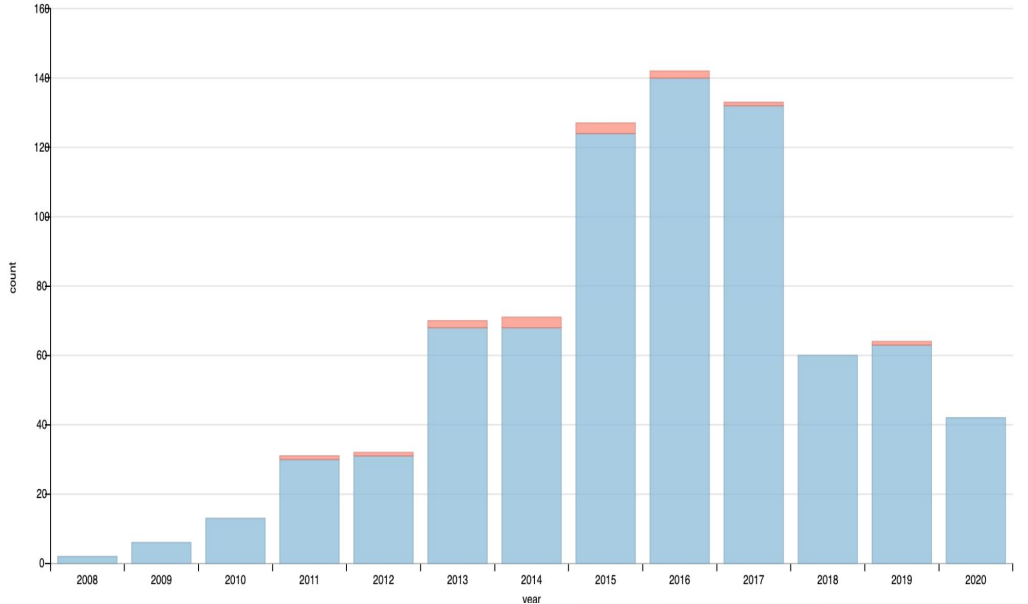
other data



profile

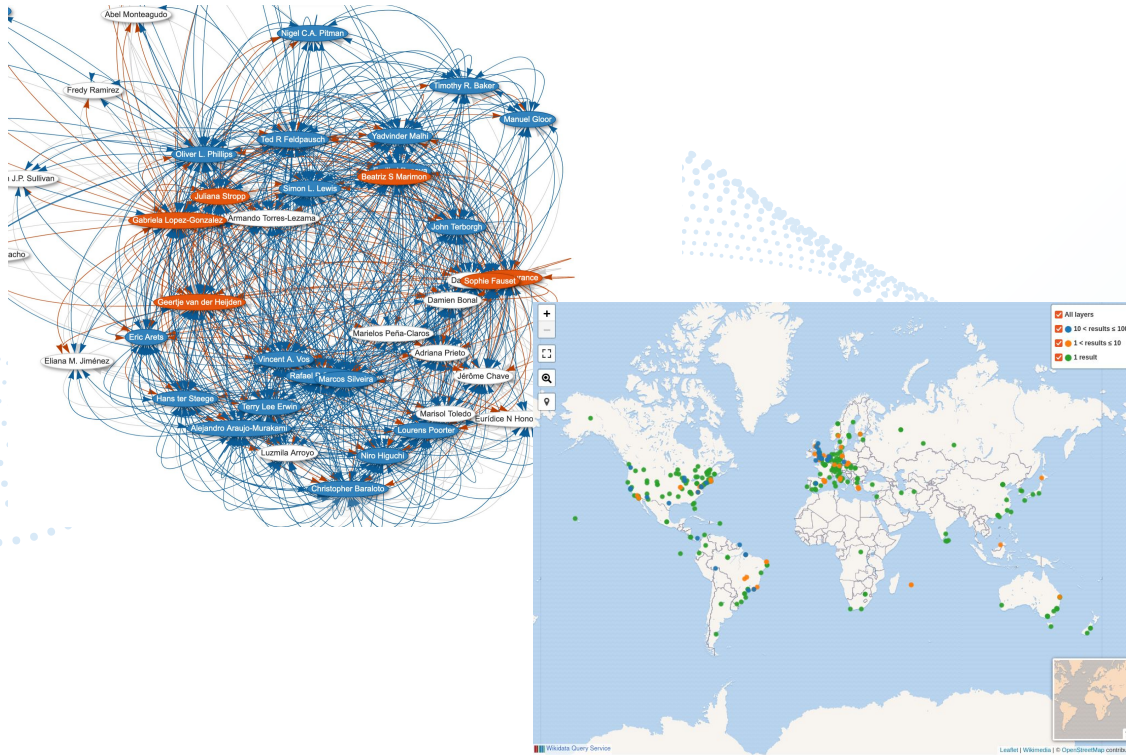
Author's publication history

Citations by year



- Based on FAIR & open data about
 - Author
 - Papers
 - Citations
 - Publication dates
- ~ complete only for pilot use cases

Co-author network



- Based on FAIR & open data about
 - Author
 - Papers
 - Co-authors
 - Genders
 - Affiliations
 - Geolocations
- ~ complete only for pilot use cases

<https://scholia.toolforge.org/author/Q56556745#coauthors-header>

<https://scholia.toolforge.org/author/Q56556745#coauthor-map-header>

Observations with Scholia



- paper available online
- paper has DOI
- paper has copyright license
- author has ORCID
- author has demographic data



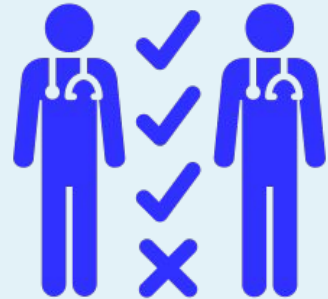
- paper metadata not available
- copyright license non-standard
- paper fails to mention funder
- paper / project / PI not associated with grant catalog entry



Beyond WikiCite

- awards, grants, and funding received
- major research projects
- demographics of researchers

WikiProject Clinical Trials imported ClinicalTrials.gov to Wikidata



Sample clinical trial

NIH U.S. National Library of Medicine

ClinicalTrials.gov

Find Studies ▾

About Studies ▾

Submit Studies ▾

Resources ▾

Ab

[Home](#) > [Search Results](#) > Study Record Detail

A Study of Induction and Maintenance Treatment of Advanced Non-squamous Non-Small Cell Lung Cancer



The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Read our [disclaimer](#) for details.

ClinicalTrials.gov Identifier: NCT00789373

Recruitment Status ⓘ : Completed

First Posted ⓘ : November 11, 2008

Results First Posted ⓘ : July 18, 2011

Last Update Posted ⓘ : December 14, 2018

Sponsor:

Eli Lilly and Company

Information provided by (Responsible Party):

Eli Lilly and Company

ClinicalTrials.gov



CLINICAL TRIALS TRANSFORMATION INITIATIVE

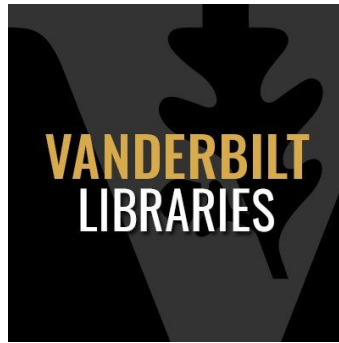
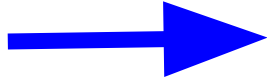


end time	P582	Point in time	end time: time an item ceases to exist or a statement stops being valid	Use of Dalfampridin in Primary Lateral Sclerosis <end time> January 2020	-
sponsor	P859	Item	sponsor, Tabula gratulatoria and research sponsor: organization or individual that sponsors this item	Opioid Prescribing After Cesarean Delivery <sponsor> Vanderbilt University Medical Center	-
main subject	P921	Item	topic, matter, topic of a work and subject: primary topic of a work (see also P180: depicts)	Use of Dalfampridin in Primary Lateral Sclerosis <main subject> motor neuron disease	-
number of participants	P1132	Quantity	number of participants: number of participants of an event, e.g. people or groups of people that take part in the event (NO units)	Opioid Prescribing After Cesarean Delivery <number of participants> 160	-
title	P1476	Monolingual text	original title and title: published name of a work, such as a newspaper article, a literary work, piece of music, a website, or a performance work	PARAMOUNT trial <title> A Study of Induction and Maintenance Treatment of Advanced Non-squamous Non-Small Cell Lung Cancer	-
short name	P1813	Monolingual text	abbreviation: short name of a place, organisation, person, Wikidata property, etc.	PARAMOUNT trial <short name> PARAMOUNT trial	-
medical condition treated	P2175	Item	disease that this pharmaceutical drug, procedure, or therapy is used to treat	Use of Dalfampridin in Primary Lateral Sclerosis <medical condition treated> motor neuron disease	di u:

Thanks Sheri Tibbs



VANDERBILT
UNIVERSITY



Vanderbot



- library profiles all university researchers
- library transfers profiles to Wikidata
- cross-reference with clinical trials

Thanks

Steven James Baskauf

Clifford Anderson

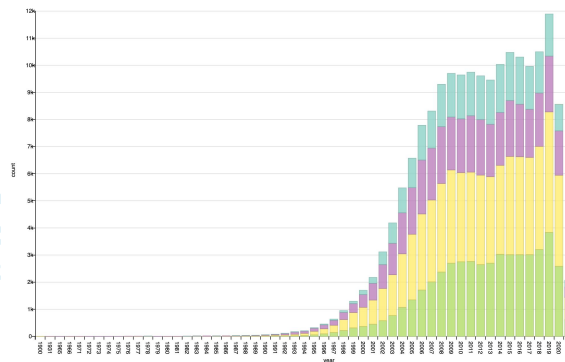
Philip Walker

- papers from PIs
- trials at Vanderbilt



- no data added
- data visualized!

Phases



Medical conditions

Show 10 entries

Count	Medical condition	Example trial
10,093	diabetes	Atherosclerosis Risk in Communities
7,433	syndrome	Serum Metabolomics Study of Polycystic Ovary Syndrome
5,613	carcinoma	Randomized Phase 1/2 Open-Label Trial of PR104 and Sorafenib in Patients with Metastatic Hepatocellular Carcinoma (HCC)

Interventions

Show 10 entries

Count	Intervention
5,180	cyclophosphamide
3,187	paclitaxel
3,182	Anesthetics, Combined
2,625	cisplatin
2,323	carboplatin

Thanks

Daniel Mietchen

Finn Årup Nielsen

Egon Willighagen



বাংলা



উইকিপিডিয়া
একটি মুক্ত বিশ্বকোষ

- translation of English terms into Bangla
- For Wikidata, also migrates into Wikipedia
- রোগীভিত্তিক পরীক্ষণ for everyone

Thanks

Tanvir Rahman

Social and ethical issues



Recurring issues

Global issues

- Who do we include in discussions?
- Rich always benefit first

Big but topical issues

- reporting demographics
- addressing data gaps
- reporting problematic observations



Do not use Wikipedia for promotion



Wikidata properties for a researcher



WIKIPEDIA



expected

ORCID

occupation / field of work

educated at

affiliation / employer

country of citizenship

languages spoken

perhaps surprising

personal pronoun

sex or gender

place of birth

ethnic group

religion

sexual orientation

Demographic data: ethical issues

Good things

- supports diversity, equity, and inclusion
- credits / showcases underserved communities
- promotes collaboration, humility, transparency

Bad or odd things

- Is gender / ethnicity private?
- cannot come from institutions
- structured data can have weird consequences

Sources of demographic data



WIKIPEDIA



Women in Red



ART + FEMINISM



WIKIDATA LGBT



AFRO FREE CULTURE CROWDSOURCING WIKIMEDIA

Sources of demographic data



AcademiaNet 

MULHERES NA CIÊNCIA

500
QS

Cell

COMMENTARY | VOLUME 184, ISSUE 3, P561-565, FEBRUARY 04, 2021

Fund Black scientists

WOMEN IN DATA SCIENCE
STANFORD UNIVERSITY



Vanderbilt University

Gender distribution

Count of the number of researchers wrt. gender.

Show entries

Count	Gender
2951	male
2080	female

Thanks



Thanks, WikiCite and Scholia teams

- Wikimedia community
- [Scholia team](#)
- Alfred P. Sloan Foundation ([G-2019-11458](#))

Further reading

- [Wikidata:Scholia](#)
- [Scholia about Scholia](#)
- [Scholia on GitHub](#)

Thanks, WikiProject Clinical Trials

- Duke University, Sheri Tibbs
- ClinWiki
 - William Hoos
 - Amy Westermann
- Vanderbilt University
 - Steven James Baskauf
 - Clifford Anderson
 - Philip Walker
- PCORI, Jeffrey Keefer
- Scholia
 - Daniel Mietchen
 - Finn Årup Nielsen
 - Egon Willighagen
- Wikimedia Bangladesh, Tanvir Rahman
- University of Virginia, Cherrie Kwok
- Wellcome Trust
- Presentation template by SlidesCarnival



Any questions?

Lane Raspberry

user:blueraspberry

rasberry@virginia.edu

