

A hands-on introduction to Wikidata and WikiCite



+ { } wikicite

Dario Taraborelli
Wikimedia Foundation
@readermeter

Daniel Mietchen
University of Virginia
@evomri

FORCE 2018 • Montreal, 12 October 2018



WIKIDATA

Infoboxes

Marie Skłodowska Curie



c. 1920

Born	Maria Salomea Skłodowska 7 November 1867 Warsaw, Kingdom of Poland , then part of Russian Empire ^[1]
Died	4 July 1934 (aged 66) Passy, Haute-Savoie, France
Cause of death	Aplastic anemia from exposure to radiation
Residence	Poland, France
Citizenship	Poland (by birth) France (by marriage)
Alma mater	University of Paris ESPCI ^[2]

Infoboxes

Marie Skłodowska Curie



c. 1920

Born	Maria Salomea Skłodowska 7 November 1867 Warsaw, Kingdom of Poland , then part of Russian Empire ^[1]
Died	4 July 1934 (aged 66) Passy, Haute-Savoie , France
Cause of death	Aplastic anemia from exposure to radiation
Residence	Poland, France
Citizenship	Poland (by birth) France (by marriage)
Alma mater	University of Paris ESPCI ^[2]

Born	Maria Salomea Skłodowska 7 November 1867 Warsaw, Kingdom of Poland , then part of Russian Empire ^[1]
Died	4 July 1934 (aged 66) Passy, Haute-Savoie , France
Cause of death	Aplastic anemia from exposure to radiation
Residence	Poland, France
Citizenship	Poland (by birth) France (by marriage)
Alma mater	University of Paris ESPCI ^[2]

Infoboxes

	居里照，约1920年摄
出生	玛丽亚·萨洛美娅·斯克沃多夫斯卡 Maria Salomea Skłodowska 1867年11月7日 俄罗斯帝国波兰王国华沙 ^{[1]}
逝世	1934年7月4日（66歲） 法国帕西
死因	再生不良性贫血
公民权	波兰（出生） 法国（成婚）
母校	巴黎大学 巴黎高等物理化工学院
知名于	放射性 钋 镭
配偶	皮埃尔·居里（1895年成婚）
儿女	伊雷娜·约里奥-居里 兹兹·居里

معلومات شخصية	
الاسم عند الولادة	Maria Salomea Skłodowska
الميلاد	7 نوفمبر 1867 وارسو، فستولا لاند، روسيا (في بولندا حاليًا)
الوفاة	4 يوليو 1934 (66 سنة) باسي، باريس، فرنسا
سبب الوفاة	فقر الدم اللاتنسجي ^{[1]}
مكان الدفن	مقبرة العظام
الإقامة	باريس
مواطنة	روسية (بموجب تبعية بولندا) للالإمبراطورية الروسية
الجنسية	بولندية، ثم فرنسية فيما بعد
العرق	بولنديون
الديانة	ملحدة
عضوة في	الأكاديمية الملكية السويدية للعلوم، والأكاديمية الروسية للعلوم، والأكاديمية الوطنية للطب
الزوج	بيار كوري (1895-1906)
أبناء	إيرين جوليو-كوري، وإيف كوري

জন্ম	নভেম্বর ৭ ১৮৬৭ ওয়ারশ, কংগ্রেস পোল্যান্ড
মৃত্যু	জুলাই ৪, ১৯৩৪ (৬৬ বছর) Sancellemoz, ফ্রান্স
জাতীয়তা	পোলীয়, ফরাসি
কর্মক্ষেত্র	পদার্থবিজ্ঞান এবং রসায়ন
প্রতিষ্ঠান	সরবোন
প্রাক্তন ছাত্র	সরবোন এবং ইএসপিসিআই
পিএইচডি উপদেষ্টা	অঁরি বেকেরেল
পিএইচডি ছাত্রেরা	অঁদ্রে-লুই দ্যবিয়েঁ Marguerite Catherine Perey
পরিচিতির কারণ	তেজস্ক্রিয়তা
উল্লেখযোগ্য পুরস্কার	 পদার্থবিজ্ঞানে নোবেল পুরস্কার (১৯০৩) রসায়নে নোবেল পুরস্কার (১৯১১)



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

Tools

- What links here
- Related changes
- Special pages
- Permanent link
- Page information
- Concept URI
- Cite this page

Item Discussion

Read

View history

Search Wikidata

Marie Curie (Q7186)

French-Polish physicist and chemist

Maria Salomea Skłodowska | Maria Skłodowska-Curie | M. Curie

In more languages

Statements

instance of



human

edit

1 reference

+ add value

part of



Pierre and Marie Curie

edit

1 reference

+ add value

Wikidata is to data what Wikipedia is to text

- All data CC0
- Anybody can contribute
- Covers all domains of knowledge
- Fully version controlled and collaborative
- Integrated with the semantic web via RDF / open APIs
- High performance query engine (SPARQL)
- Stable. Not tied to short-term funding cycles
- Actively developed, full stack is open source
- Active community. Fastest growing Wikimedia project



50 million
entities

550M statements • 760M edits

[as of October 2018]

Multilingual labels and descriptions

Language	Label	Description	Also known as 
Spanish	Marie Curie	química y física polaca naturalizada francesa	Maria Skłodowska-Curie Maria Salomea Skłodowska Marie Skłodowska-Curie Madame Curie Mme Curie Marie Curie-Skłodowska Madame Pierre Curie M. Curie
Traditional Chinese	瑪麗·居里	物理學家，化學家	居里夫人
Chinese	玛丽·居里	物理学家，化学家	居里夫人
Afrikaans	Marie Curie	Fraans-Poolse fisikus en chemikus	
Amharic	ማሪ ኩሪ	No description defined	
Aragonese	Marie Curie	No description defined	
Arabic		ماري كوري	عالمة كيمياء بولندية
Egyptian Arabic		مارى كورى	No description defined
Assamese	মেডাম কুৰী	No description defined	
Asturian	Marie Curie	química y física polaco-francesa	Maria Salomea Skłodowska Marie Skłodowska-Curie

Statements and references

Statements

instance of

 human  edit

▼ 1 reference

stated in

data.bnf.fr

retrieved

10 October 2015

reference URL

<http://data.bnf.fr/ark:/12148/cb121447141>

+ add reference

+ add value

Identifiers

Identifiers

Mathematics Genealogy

Project ID



177944

▼ 0 references

VIAF ID



76353174

▼ 0 references

Types of content in Wikidata

- People
 - Places
 - Taxa
 - Buildings
 - Organisations
 - Artworks
 - Events
 - Astronomical Bodies
 - ...
- Chemicals
 - Processes
 - Theorems
 - Concepts
 - Creative works
 - Journals
 - Publishers
 - Meta-items
 - ...

Sample of current biomedical content 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*](#)

Data about species (endemic moths of NZ)



 commons:Asaphodes aegrota male.jpg
Q Asaphodes aegrota



 commons:Asaphodes aegrota female.jpg
Edit this query: s aegrota



 commons:Graphania scutata female.jpg
Q Graphania scutata



 commons:Graphania rubescens male.jpg
Q Graphania rubescens



 commons:Ichneutica notata female.jpg
Q Ichneutica notata

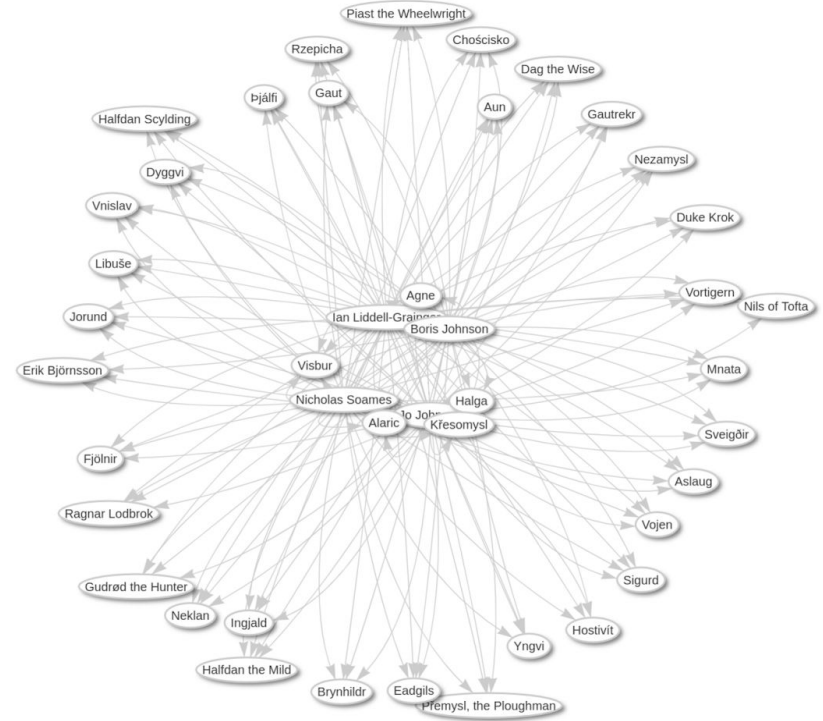


 commons:Ichneutica notata male.jpg
Q Ichneutica notata

655 Results in 58 ms

[[SPARQL query](#)]

Members of the British
Parliament with
ancestors allegedly
identified as mythical



[\[SPARQL query\]](#)

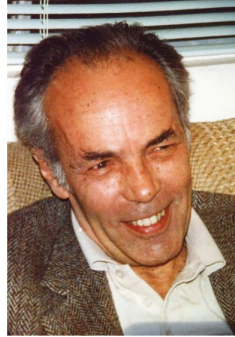
Biologists with Canadian citizenship



Severn Cullis-Suzuki
[commons:Severn Cullis-Suzuki.jpg](#)



Boucar Diouf
[commons:Boucar Diouf \(Salon du livr...](#)



Guy Courtois
[commons:Dr Guy Courtois.jpg](#)



Vladimir Hachinski
[commons:V hachinski.jpg](#)



Jack Szostak
[commons:Nobel Prize 2009...](#)



Anne Vallée
[commons:Anne Vallée.jpg](#)



Suzanne Morrow
[commons:Suzanne Morrow 194...](#)

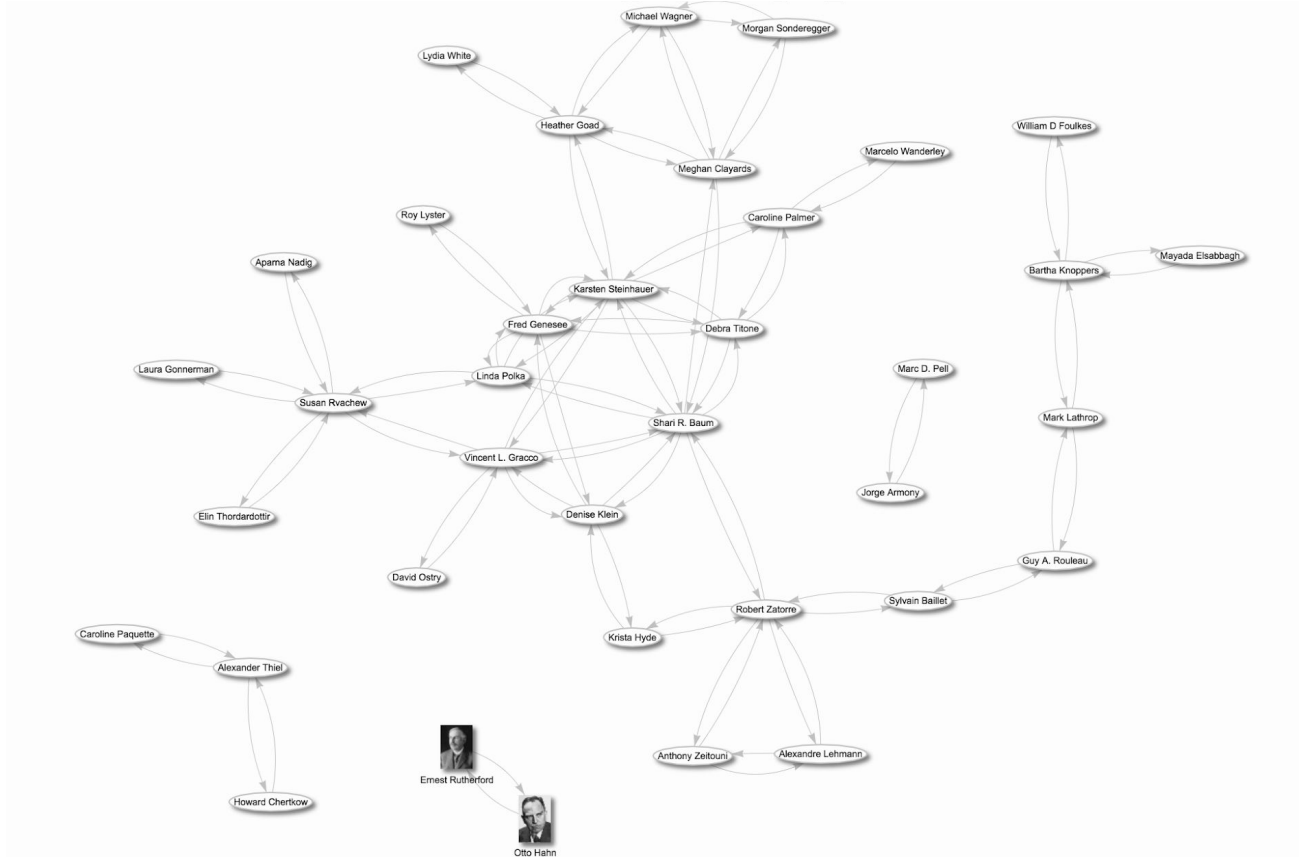


Sidney Altman
[commons:Sidney Altman crop.jpg](#)



Roberta Bondar
[commons:Roberta Bondar2.jpg](#)

Co-author graph of McGill-affiliated authors



[[SPARQL query](#)]

Award recipients affiliated with McGill

Count	Award	Recipients
56	Officer of the Order of Canada	Louis Dudek, Margaret Lock, Carl Goldenberg, Carl Goresky, Daniel Ling, David McCurdy Baird, Mark Wainberg, Henry Mintzberg, John Jonas, Eric Kierans, Don Johnston, Alan B. Gold, Alice Vibert Douglas, Balfour Mount, Bartha Knoppers, Clarke Fraser, Michel Chrétien, Nahum Sonenberg, Irwin Cotler, Bruce Trigger, Richard Goldbloom, Roy Heenan, Saul Hayes, Stanley Hartt, Stephen Toope, Thomas Cameron, Thomas Chang, Paul Davenport, Oliver Jones, Antoine Hakim, George Karpati, Heinz Lehmann, Jean Davignon, Desmond Morton, Donald Baxter, Rémi Quirion, Samuel O. Freedman, Theodore Sourkes, Thomas J. Hudson, Ronald Melzack, Warren Allmand, Krešimir Krnjević, Leo Yaffe, Albert Aguayo, Arthur Vineberg, Barry Posner, Gregory Baum, Marian Engel, Frederick Andermann, George Gray Bell, Heather Munroe-Blum, John Dossetor, Jonathan Larmonth Meakins, Michael Kelway Oliver, Moyra Allen, Myer Horowitz
24	Fellow of the Royal Society	Brenda Milner, Ernest Rutherford, William Osler, Donald O. Hebb, Charles Philippe Leblond, Arnold Burgen, Wilder Penfield, Charles Scriver, Michel Chrétien, Nahum Sonenberg, Otto Maass, Frank Dawson Adams, Edgar William Richard Steacie, Boris Babkin, John William Dawson, V. C. Wynne-Edwards, John George Adami, Allan Hay, Archibald Macallum, Arthur Stewart Eve, Frederick Soddy, Juda Hirsch Quastel, Henry Bovey, John Stuart Foster
19	Canadian Medical Hall of Fame	Mark Wainberg, Brenda Milner, William Osler, John McCrae, Donald O. Hebb, Charles Philippe Leblond, Wilder Penfield, Clarke Fraser, Charles Scriver, Michel Chrétien, Phil Gold, Richard Goldbloom, John Browne, Antoine Hakim, Heinz Lehmann, Ronald Melzack, Maude Abbott, Albert Aguayo, Henry Friesen
16	Companion of the Order of Canada	Kenneth Hare, Brenda Milner, Hugh MacLennan, Charles Philippe Leblond, Wilder Penfield, Charles Scriver, Paul-André Crépeau, Phil Gold, Richard Cruess, F. R. Scott, Victor Goldbloom, Henry Friesen, Raymond Klibansky, Charles Taylor, Juda Hirsch Quastel, Gerald Le Dain
16	Officer of the National Order of Quebec	Guy A. Rouleau, Luc Vinet, Margaret Lock, François Tavenas, Mark Wainberg, Brenda Milner, Henry Mintzberg, Michel Chrétien, Paul-André Crépeau, Irwin Cotler, Bruce Trigger, Richard Cruess, Ronald Melzack, Jean-Marie Dufour, Frederick Andermann, Heather Munroe-Blum

Types of content in Wikidata

- People
- Places
- Taxa
- Buildings
- Organisations
- Artworks
- Events
- Astronomical Bodies

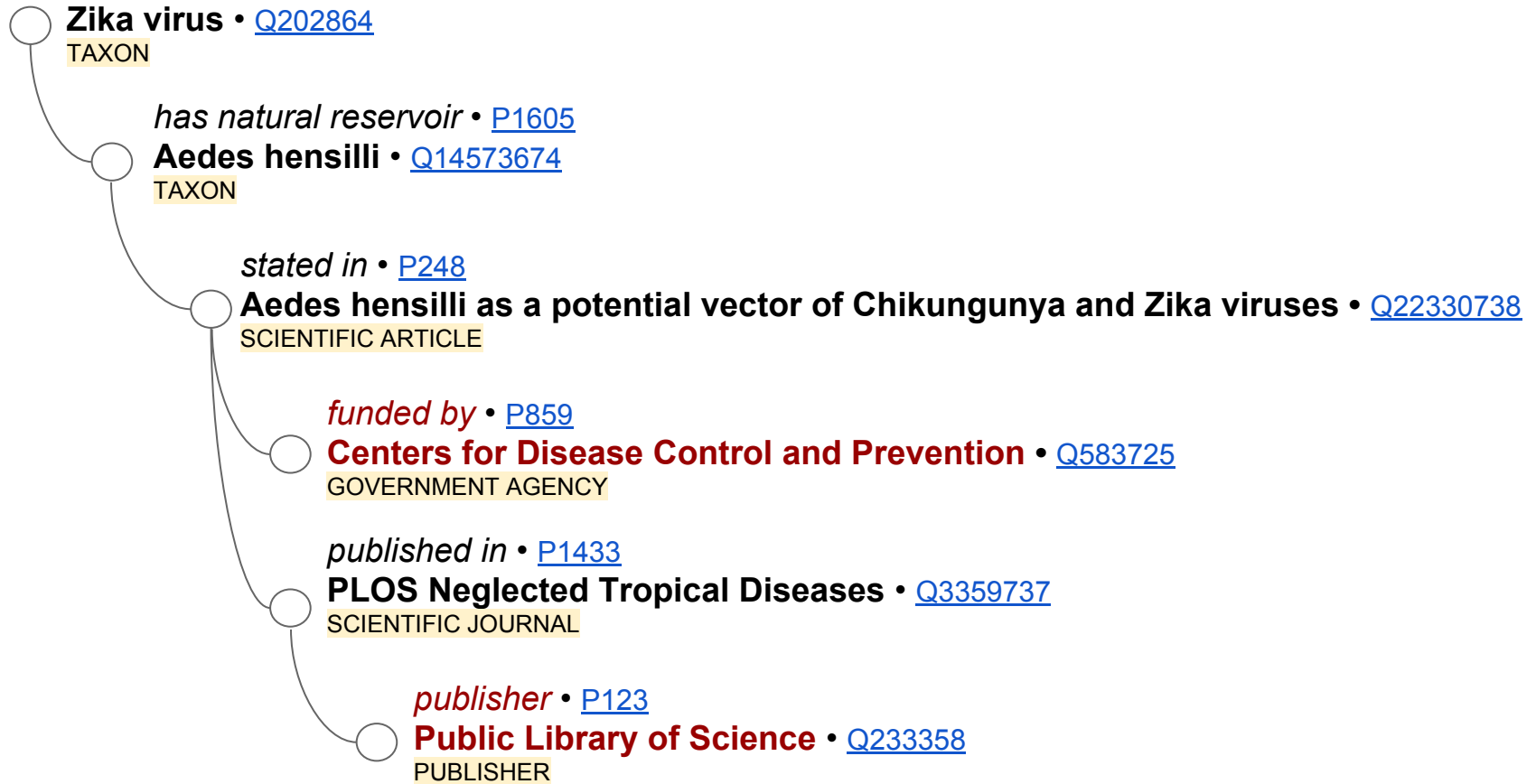
...

- Chemicals
- Processes
- Theorems
- Concepts
- **Creative Works**
- **Journals**
- **Publishers**
- Meta-items

...



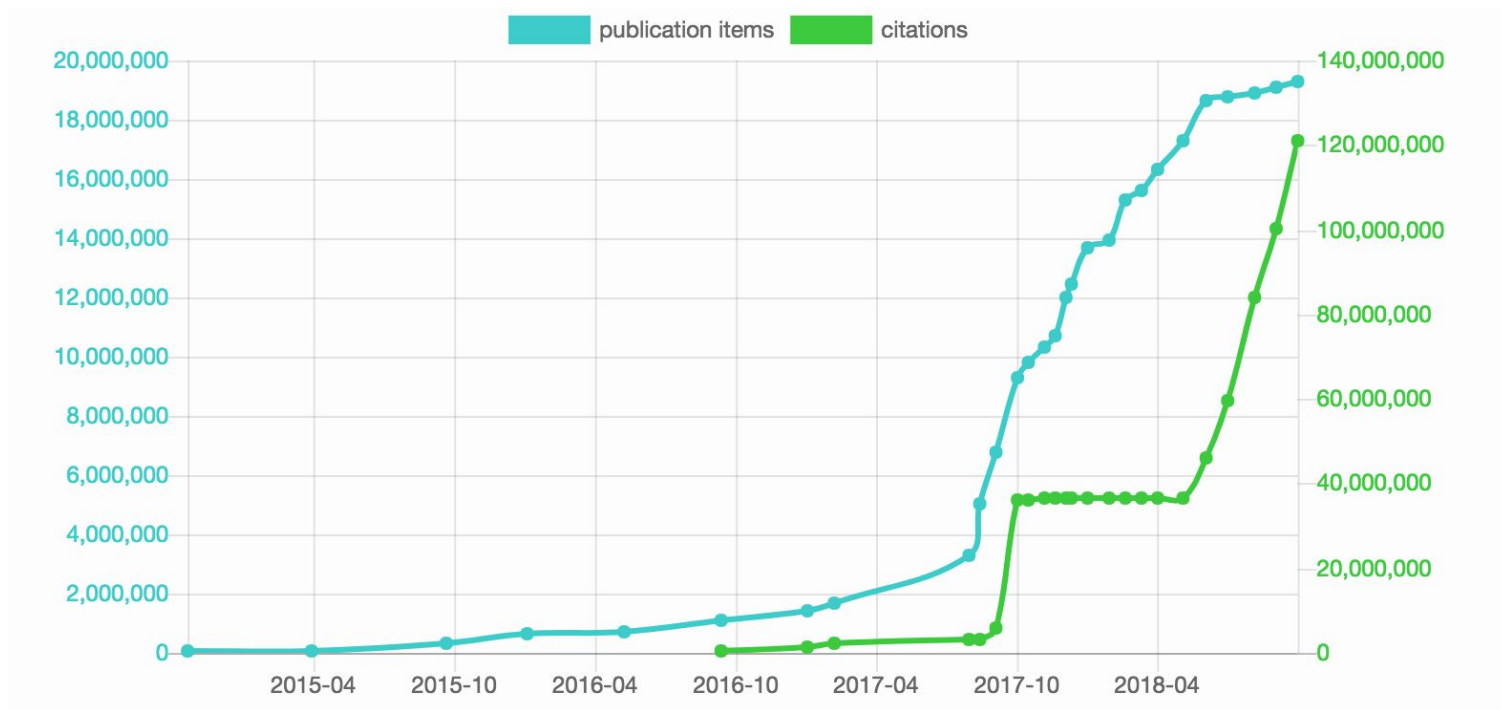
A structured bibliographic database of linked data
to serve as a foundation of free knowledge



DATA PROVENANCE OF STATEMENTS BY OUTLET, PUBLISHER OR FUNDER

A growing database of sources stored as Wikidata items

19 million items, 38% of all Wikidata



Key bibliographic properties

(% growth since [October 2017](#))

author (P50) 860K → [3.1M](#) (+260%)

author name string (P2093) 40.2M → [81.9M](#) (+103%)

ISBN-13 (P212) *often used for book editions* 30K → [40K](#) (+33%)

DOI (P356) *often used for journal articles* 6.8M → [14M](#) (+105%)

main subject (P921) 400K → [3.6M](#) (+800%)

All scholarly articles cited in Wikipedia across languages

All scholarly sources cited on Wikipedia using a **DOI**, a **PubMed**, or **PubMed Central** identifier now available as Wikidata items

5. [^] [Penicillin at the Free Dictionary](#)
6. [^] Garrod, L. P. (1960). "Relative Antibacterial Activity of Three Penicillins". *British Medical Journal*. **1** (5172): 527–29. doi:10.1136/bmj.1.5172.527.
7. [^] Garrod, L. P. (1960). "The Relative Antibacterial Activity of Four Penicillins". *British Medical Journal*. **2** (5214): 1695–6. doi:10.1136/bmj.2.5214.1695. PMC 2098302. PMID 13703756.
8. [^] "Penicillin G and Penicillin V". *livertox.nih.gov*. Retrieved 2016-09-25.
9. [^] "Penicillin (Benzylpenicillin, Penicillin G, Bicillin C-R/L-A, Pfizerpen, Wycellin)". *The Antimicrobial Index*. Knowledgebase. Retrieved 4 March 2014.
10. [^] "Penicillin G sodium salt Susceptibility and Resistance Data" (PDF). TOKU-E. Retrieved 4 March 2014.
11. [^] ^a ^b ^c Rossi S, ed. (2006). *Australian Medicines Handbook*. Adelaide: Australian Medicines Handbook. ISBN 0-9757919-2-3.

Broad coverage of scholarly journals



Web of Science currently includes 32,840 scholarly journals –@habib #force2017



3:30 AM - 25 Oct 2017



There are now 42,650 items about scholarly journals in @wikidata, up +80% since last year.

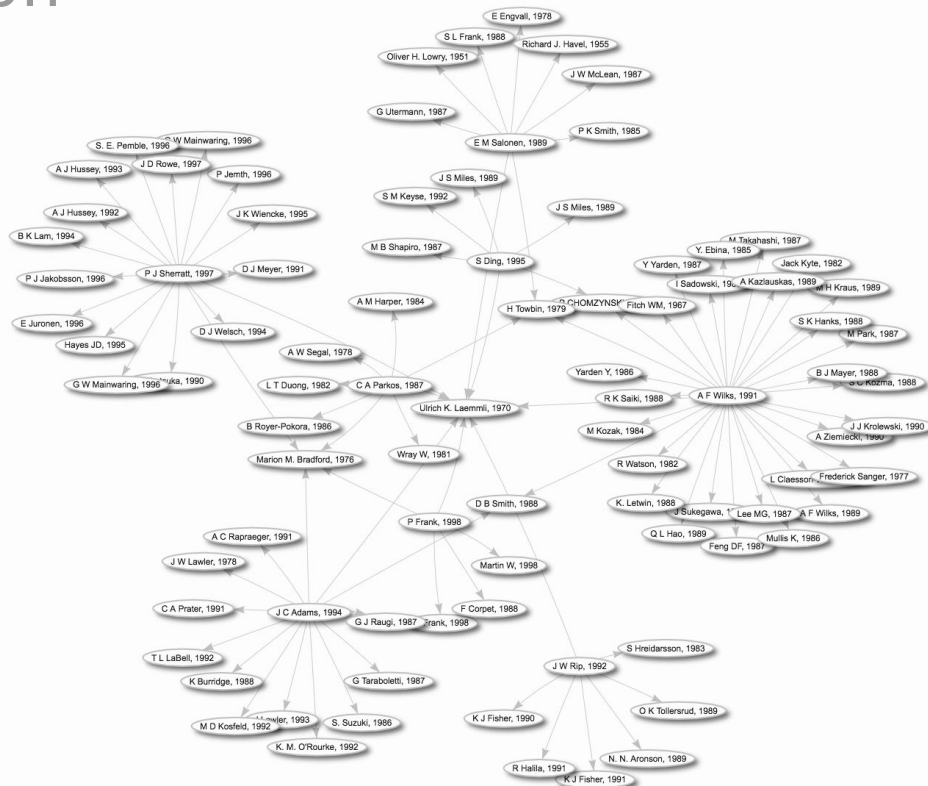
<https://twitter.com/Wikicite/status/923134697680629760>

<https://twitter.com/Wikicite/status/921520882392088576>

The Wikidata citation graph

130 million citation links

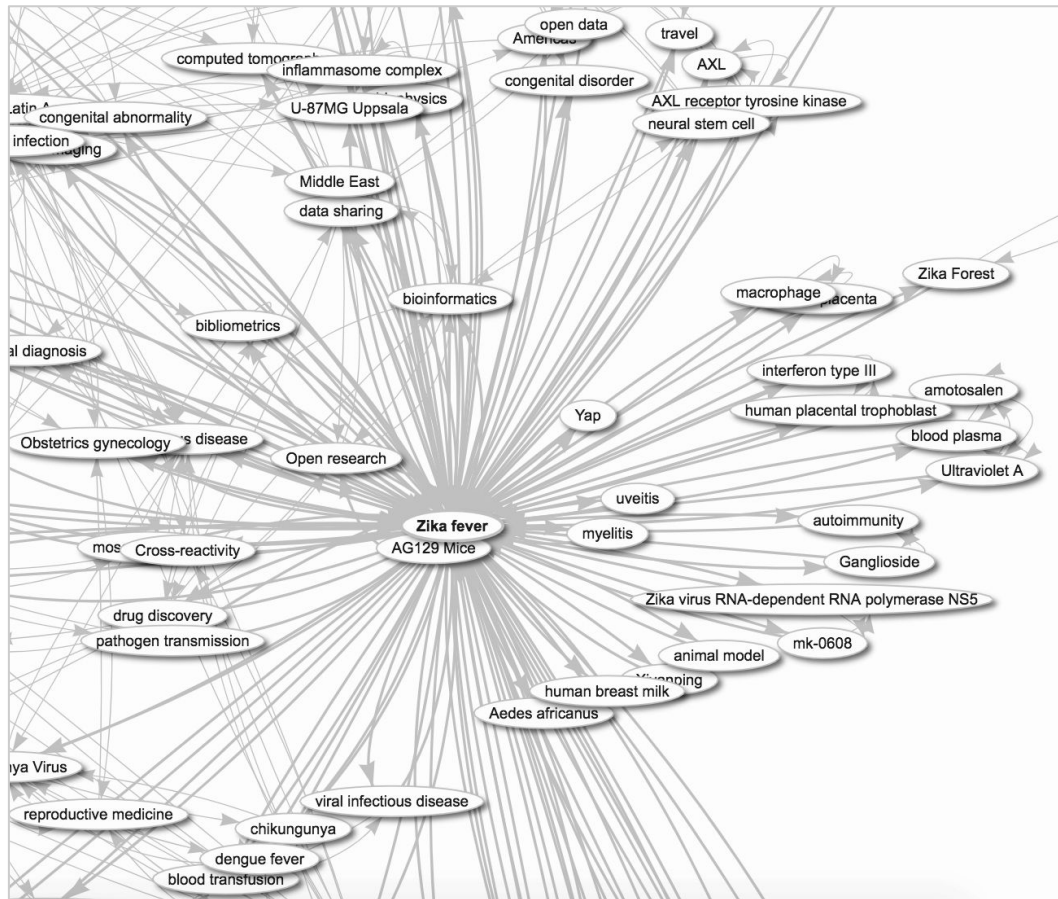
using the *cites* (P2860)
property



PARTIAL CITATION GRAPH FOR **ULRICH K. LAEMMLI (1970)** • <http://tinyurl.com/y7acpqz4>

The Zika corpus

the complete, **annotated**
scholarly publication
corpus on Zika virus
(Q202864)



Linking a scholarly paper to the assertions it supports

The Alzheimer's disease-associated amyloid beta-protein is an antimicrobial peptide ([Q21090025](#))

DOI: [10.1371/JOURNAL.PONE.0009505](#).

Supports the following statement(s)

Statements in Wikidata supported by references to this work. Only a maximum of around 2000 statements are shown.

Search:

Item	Property	Value
Cathelicidin antimicrobial peptide	biological process	antimicrobial humoral immune response mediated by antimicrobial peptide
Cathelicidin antimicrobial peptide	biological process	innate immune response
Amyloid beta precursor protein	biological process	antifungal humoral response
Amyloid beta precursor protein	biological process	antibacterial humoral response
Amyloid beta precursor protein	biological process	defense response to Gram-negative bacterium

Using annotations for provenance

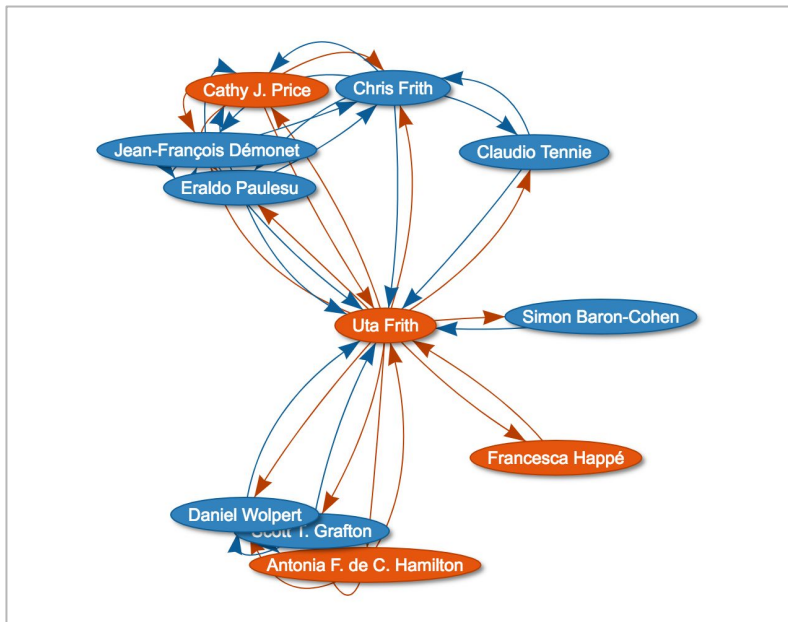
decreased expression in	schizophrenia ...
determination method	Immunohistochemistry ...
1 reference	
stated in	Reduction in Reelin immunoreactivity in hippocampus of subjects with schizophrenia, bipolar disorder and major depression.
retrieved	24 July 2017
reference URL	https://via.hypothes.is/https://www.ncbi.nlm.nih.gov/pubmed/11126396#annotations:HVGXGMnfUEetV2sj-_VpSQ

Example: [Q13561329](#) (Reelin)

Paper: [Reduction in Reelin immunoreactivity in hippocampus of subjects with schizophrenia, bipolar disorder and major depression](#)

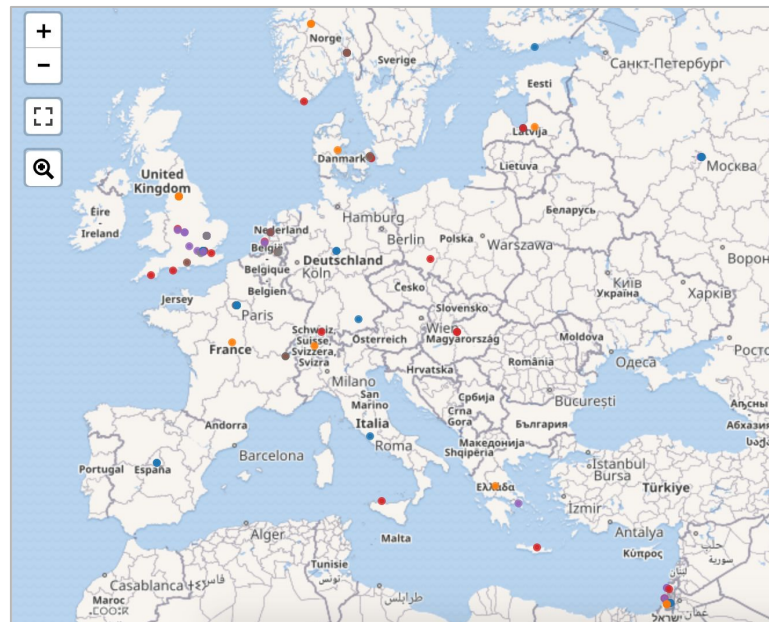
Exploring Wikidata contents: Scholia

an open source, Wikidata-driven **scholarly profile application**



CO-AUTHOR GRAPH FOR **UTA FRITH**

<https://tools.wmflabs.org/scholia/author/Q8219>



LOCATION OF **TURING AWARD** RECIPIENTS

<https://tools.wmflabs.org/scholia/award/Q185667>



DEMO

Demo

1. Create a Wikidata item for an author
2. Add a paper to Wikidata via its DOI
3. Annotate a paper with topic information
4. Add an author to a list of conference participants
5. Query Wikidata

1. Create a Wikidata item for an author

Create a new item for an author

<https://www.wikidata.org/wiki/Special:NewItem>

Add a label and description

E.g. *Kristen Ratan*

Add basic statements and identifiers

E.g. class, occupation, gender, ORCID, affiliation

2. Add a paper to Wikidata via its DOI

Pick a DOI

<https://doi.org/10.3233/ISU-180010>

Run DOI ingesting tool SourceMD on it

https://tools.wmflabs.org/sourcemd/index_old.php

(newer version: <https://tools.wmflabs.org/sourcemd/>)

3. Annotate a paper with topic information

Open the Wikidata item about the paper you want to annotate

Check if a “main subject” (P921) statement exists or add one

4. Add an author to a list of conference participants

Type “FORCE2018” into the search box on Wikidata

<https://www.wikidata.org/wiki/Q56579271>

Check the “participant” section

<https://www.wikidata.org/wiki/Q56579271#P710>

Click “add statement”

Type “participant” (or P710) in the property field. Type the name of the participant in the value field

(the participant needs to have an item as an author on Wikidata)

5. Query Wikidata

Check out a Scholia profile, e.g.

Canada

<https://tools.wmflabs.org/scholia/topic/Q16>

Zika virus

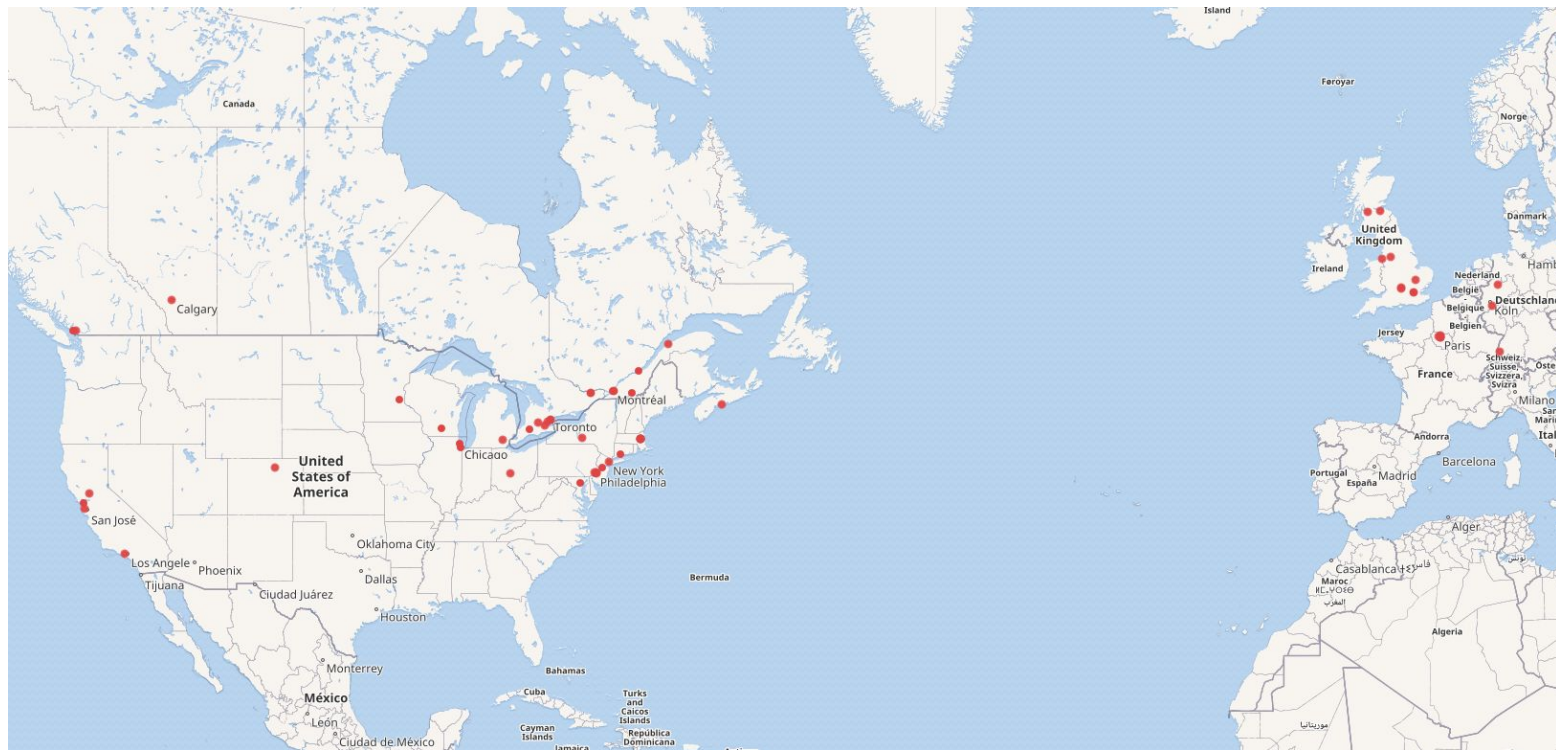
<https://tools.wmflabs.org/scholia/topic/Q202864>

Default: FORCE2018 page

<https://tools.wmflabs.org/scholia/event/Q56579271>

Edit any of the queries, e.g. change “Canada” to “Botswana”

Institutions where Canadians got their PhD



Other tasks

Link authors to institutions, journals, awards, conferences

Link all of these within Wikidata and to external identifiers

More at

https://www.wikidata.org/wiki/Wikidata:Wikidata_curricula/Activities

Thank you

D.Taraborelli, D. Mietchen (2018) A hands-on introduction to Wikidata and WikiCite. *FORCE 2018* doi.org/10.6084/m9.figshare.7201469 [CC BY]



Slides mashed up with contributions by
Andy Mabbett and Andra Waagmeester

growth by Fabio Rinaldi [CC BY] from the Noun Project
Toolkit and tools by Todd Quackenbush [CC0] from Unsplash

Appendix

Linking papers with authors without ORCID

SourceMD has a semi-automated tool for author disambiguation:

https://tools.wmflabs.org/sourcemd/new_resolve_authors.php?name=Jane+Miller&doit=Look+for+author

Adding a recent paper to Wikidata via its DOI

Check for recently published research (see [FORCE2018 workshop](#))

E.g. [via Twitter](#)

Pick a DOI

Backup 1 (from [this tweet](#)): 10.1111/geoj.12279

Backup 2 (already done): [Q57267424](#)

Run DOI ingesting tool SourceMD on it

https://tools.wmflabs.org/sourcemd/index_old.php

(newer version: <https://tools.wmflabs.org/sourcemd/>)

Syncing Wikidata items about people with their ORCID

SourceMD (aka ORCIDator)

Can handle batches of ORCIDs

<https://tools.wmflabs.org/sourcemd/> (requires a Wikidata account)

0000-0002-9989-319X

0000-0001-5413-9102

Knowledge Integrity

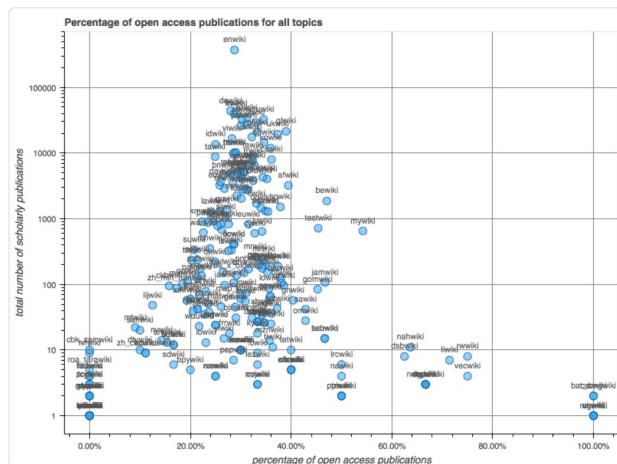
Strengthening the citation infrastructure for Wikimedia projects



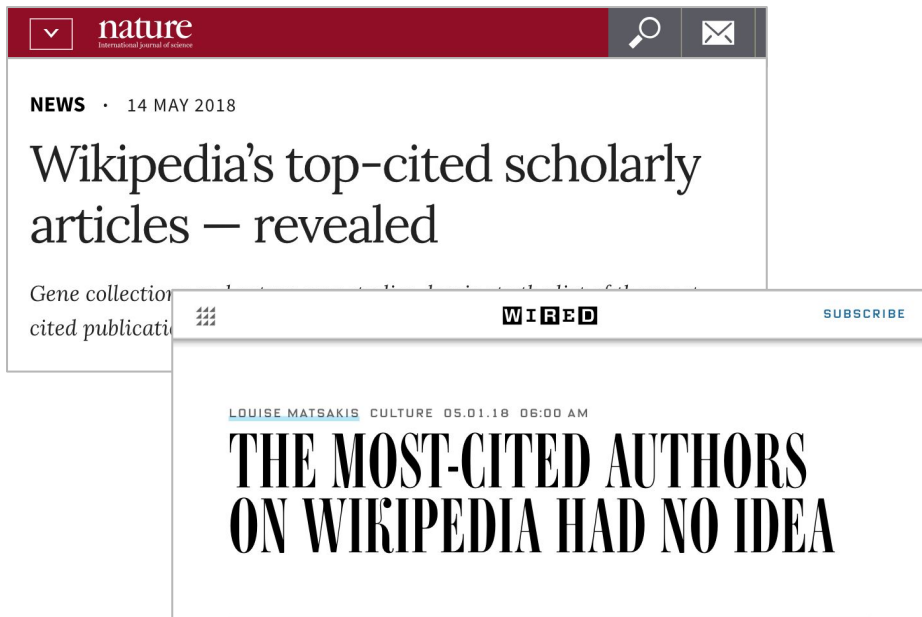
WikiResearch @WikiResearch · Aug 20

How many scientific articles cited in Wikipedia can you read without hitting a paywall? What if the only language you speak is English, Khmer, or Malagasy? We answer this question with new data obtained from @Wikipedia and @unpaywall:

wikimediafoundation.org/2018/08/20/how... #openaccess



Wikipedia, Wikimedia, Unpaywall and 3 others



<http://bit.ly/KnowledgeIntegrity>

WikiCite roadmap

3 distinct objectives, 4 possible federation options

A database of all sources
cited in Wikimedia projects
as linked data

A platform to curate and
annotate custom bibliographic
corpora

A platform to build the
“bibliographic commons”

bit.ly/WikiCiteRoadmap