

Some fun with WikiCite in Wikidata



Miriam Redi
Wikimedia Foundation

Antonin Delpeuch
University of Oxford

<https://phabricator.wikimedia.org/T208018>

Tell us about you!

Your expertise, your interests, what do you expect
from this workshop?



WIKIDATA

You are familiar with Wikipedia 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

Behind them..

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](#)

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



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

What do we care about in WikiCite? Among the types of items in Wikidata, we look at bibliographic properties.

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

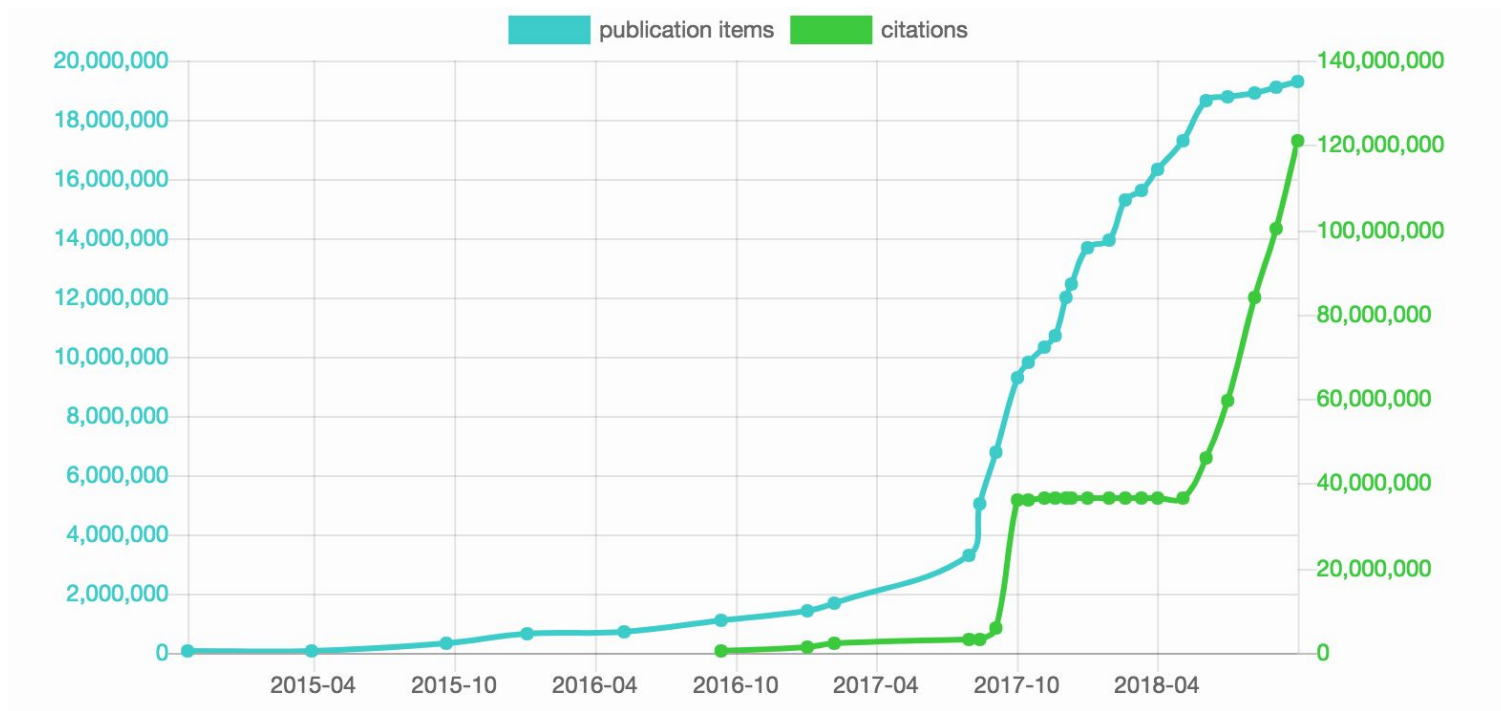
...

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

...

A growing database of sources stored as Wikidata items

19 million items, 38% of all Wikidata



Key Wikicite 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%)

More interesting properties you can play with!

Scholarly articles	Data	author (P50)  · author name string (P2093)  · editor (P98)  · genre (P136)  · title (P1476)  · subtitle (P1680)  · publication date (P577)  · published in (P1433)  · volume (P478)  · issue (P433)  · page(s) (P304)  · number of pages (P1104)  · license (P275)  · copyright owner (P3931)  · full work available at (P953)  · reference URL (P854)  · language of work or name (P407)  · main subject (P921)  · Commons category (P373)  · sponsor (P859)  · external data available at (P1325)  · reviewed by (P4032)  · corrigendum / erratum (P2507)  · is retracted by (P5824)  · describes a project that uses (P4510) 
	Identifiers	article ID (P2322)  · DOI (P356)  · PubMed ID (P698)  · PMCID (P932)  · arXiv ID (P818)  · ADS bibcode (P819)  · ZooBank publication ID (P2007)  · JSTOR article ID (P888)  · Social Science Research Network ID (P893)  · NIOSHTIC-2 ID (P2880)  · Dialnet article (P1610)  · CiNii article ID (P2409)  · OpenCitations bibliographic resource ID (P3181)  · ACM Digital Library citation ID (P3332)  · · Publons Publication ID (P3431)  · Google Scholar paper ID (P4028)  · URN-NBN (P4109)  · GONIAT paper ID (P5214)  · BioStor work ID (P5315)  · Altmetric ID (P5530) 
Books (more)	Data	author (P50)  · translator (P655)  · editor (P98)  · illustrator (P110)  · publisher (P123)  · title (P1476)  · subtitle (P1680)  · license (P275)  · full work available at (P953)  · language of work or name (P407)  · main subject (P921)  · Commons category (P373)  · copyright owner (P3931) 
	Identifiers	ISBN-10 (P957)  · ISBN-13 (P212)  · DOI (P356)  · CiNii book ID (P1739)  · HathiTrust ID (P1844) 
Authors (more)	Data	name in native language (P1559)  · affiliation (P1416)  · e-mail (P968)  · field of work (P101)  · notable work (P800)  · Commons category (P373)  · author citation (zoology) (P835) 
	Identifiers	ORCID iD (P496)  · VIAF ID (P214)  · ZooBank author ID (P2006)  · Scopus Author ID (P1153)  · Google Scholar author ID (P1960)  · ResearchGate person ID (P2038)  · ISNI (P213)  · Dialnet author ID (P1607)  · Persée author ID (P2732)  · Unz Review author identifier (P2734)  · ACM Digital Library author ID (P864)  · DBLP ID (P2456)  · Twitter username (P2002)  · SlideShare username (P4016)  · CiNii author ID (books) (P271)  · CiNii author ID (articles) (P4787)  · GONIAT author ID (P5211) 

Why is this important?

Open knowledge

Wikipedia Verifiability

Fact checking beyond Wikipedia

IDENTIFIERS??

Numbers that allow to uniquely identify and describe each publication:

PMID/PMC: medicine

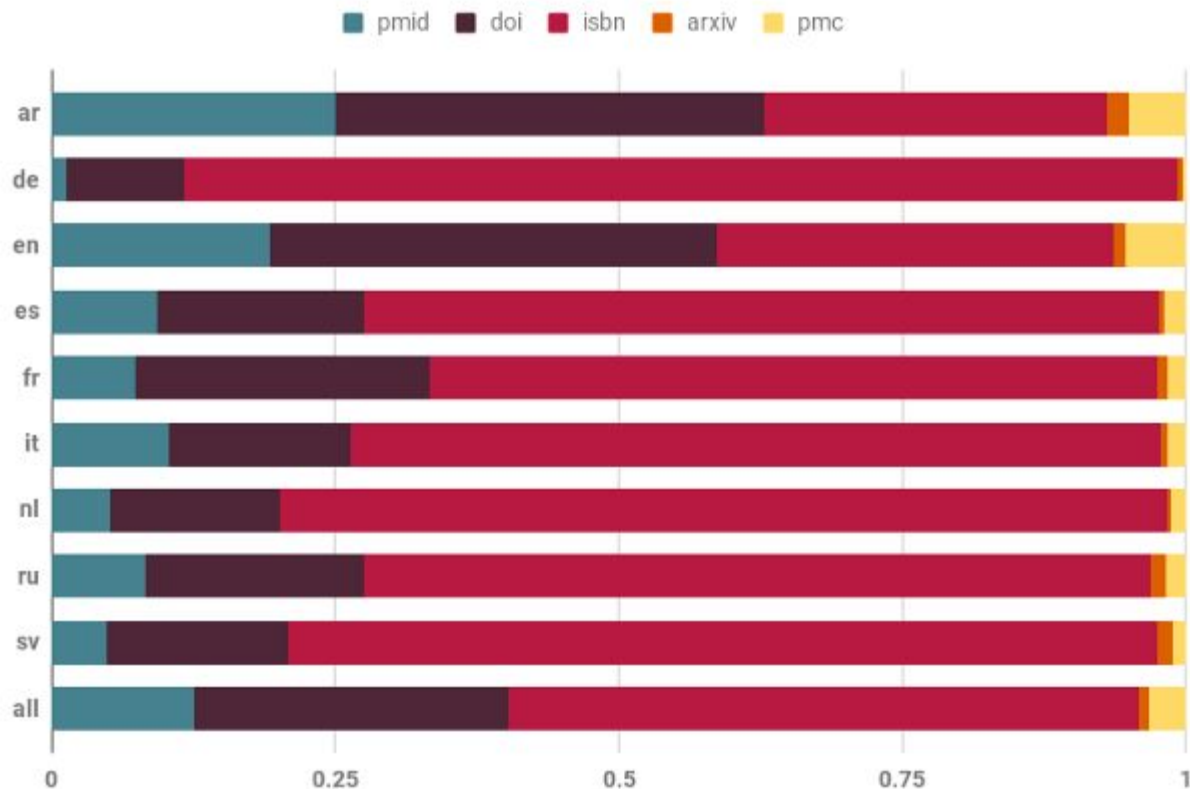
DOI: scientific

ISBN: books

ARXIV: free

publications

Distribution of citation types per language (distinct publications)



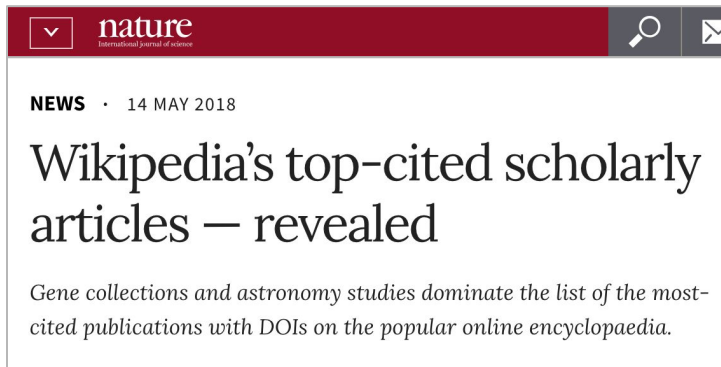
All scholarly articles cited in Wikipedia across languages

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.

We have extracted all references with ID in Wikipedia..

And categorized them by topic, accessibility and “popularity”

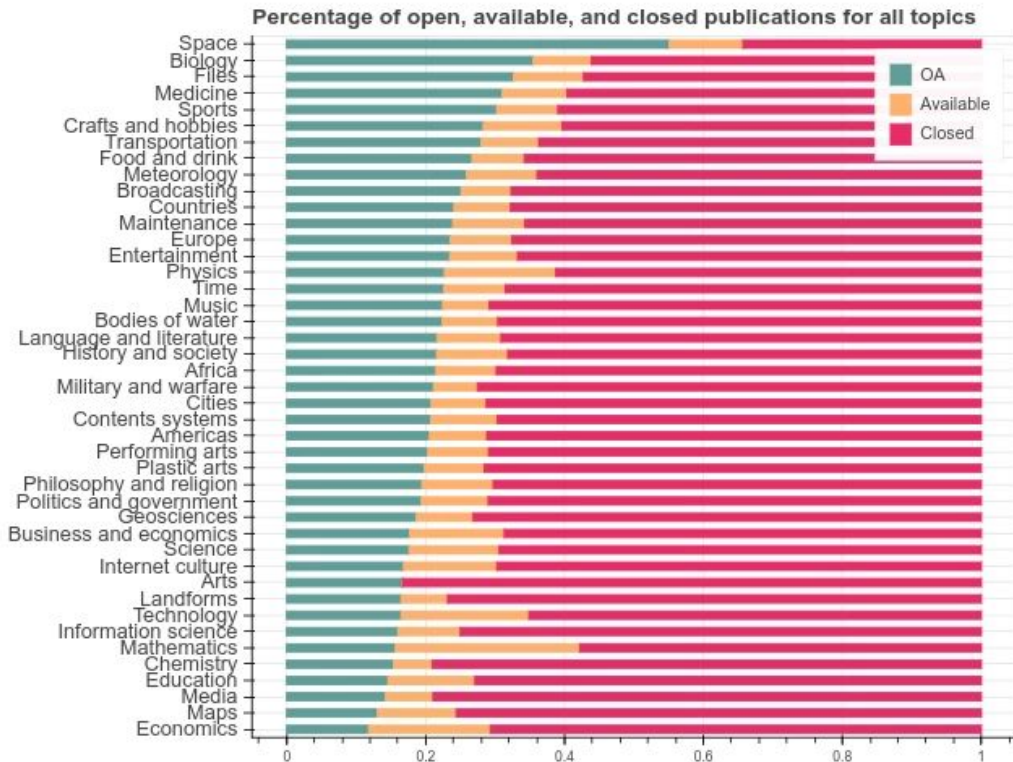


Popular refs:

<https://docs.google.com/spreadsheets/d/1uIA4YmTLIsnlzhpFFS3-4xJm-vpdv0m74S1UC4XHkPc/edit?usp=sharing>

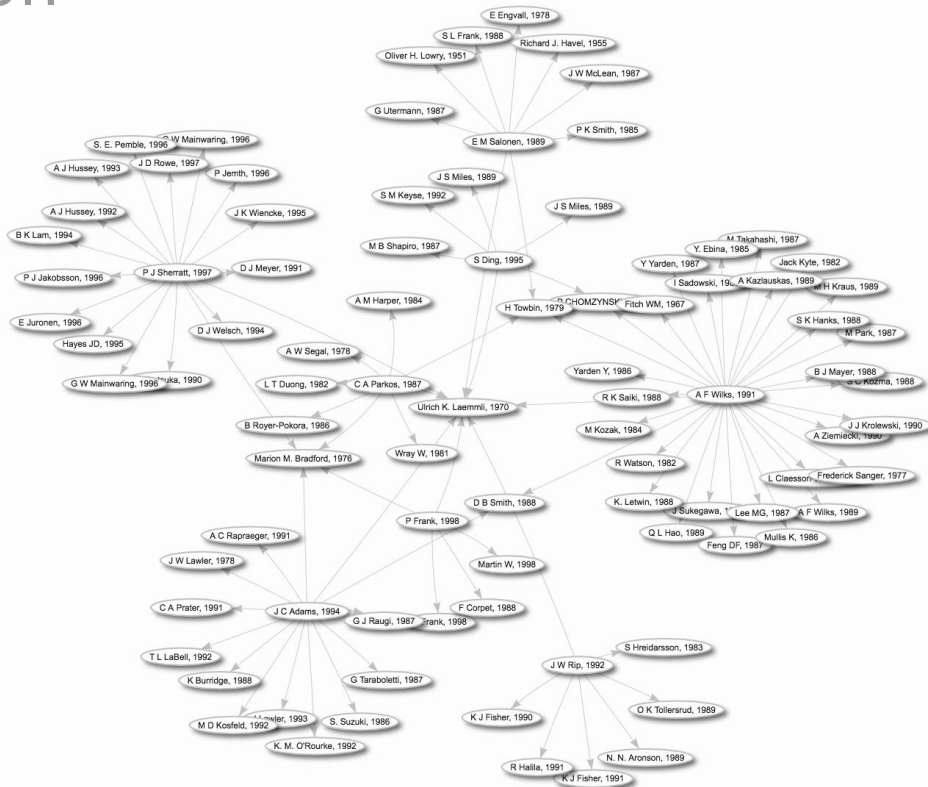
Refs by topic and accessibility:

https://nbviewer.jupyter.org/github/mirrys/citation-accessibility-viz/blob/master/visualize_WP_publication_accessibility.ipynb



130 million citation links

using the *cites* (P2860)
property



PARTIAL CITATION GRAPH FOR **ULRICH K. LAEMMLI (1970)** • <http://tinyurl.com/y7acpqzd>

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

Wikicite Applications: Inventaire

inventaire



Keep an inventory of your books

and of what can be done with them: *"this book I can lend, this book I can give, this one is for sell..."*



Share it with your friends and communities

Always dreamed of a collaborative library for your coworking / association / book club? Here it is!



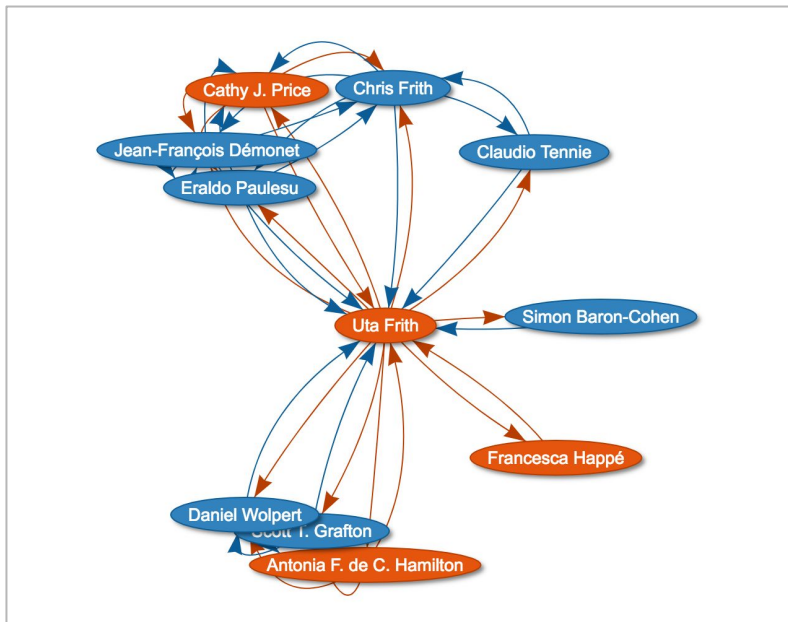
Remember the books you lend or borrow

The books you lend might actually come back now!

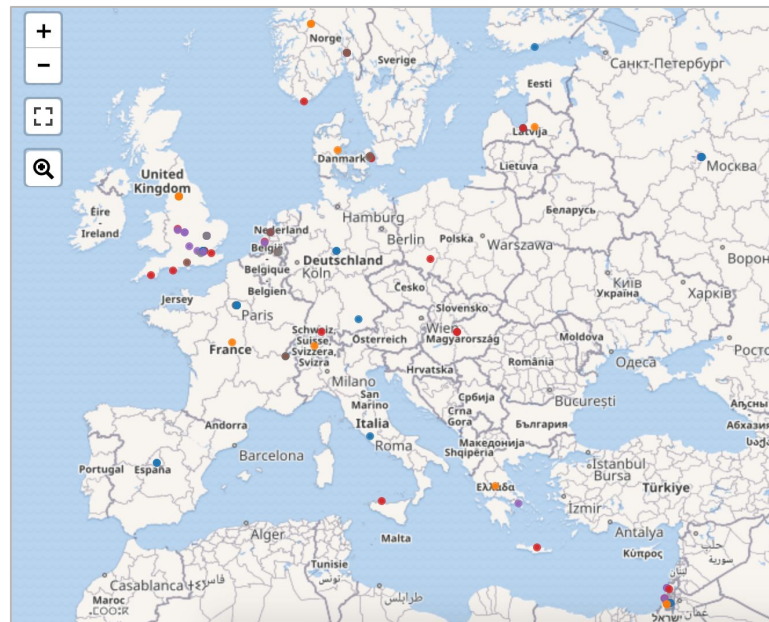
inventaire.io

Exploring Wikidata contents: Scholia

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



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



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

What can you do?

1. Create a Wikidata item for an author
2. Add a paper to Wikidata via its DOI
3. Annotate a paper/book with topic information
4. Find gaps about your favorite author/researcher through Scholia, then fill it (1-3)

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. Use Scholia to find gaps about your favorite author or topic!

Fei-Fei Li

<https://tools.wmflabs.org/scholia/author/Q18686107>

Machine Learning

<https://tools.wmflabs.org/scholia/topic/Q2539>

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

Contribute even more data!

Import your own dataset of journals, researchers, institutions...



OpenRefine tutorials:

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



Ongoing imports:

[https://www.wikidata.org/wiki/Wikidata:WikiProject Source MetaData#Current activities](https://www.wikidata.org/wiki/Wikidata:WikiProject_Source_MetaData#Current_activities)



Open data bootcamp in Brussels by @PACKEDvzw

From a table to Wikidata edits...



researcher	orcid	organization_name	organization_co	organization_id	role	start_date	end_date
Herbert J. Gans Choose new match	0000-0001-6567-5859						
Anna Frebel Choose new match	0000-0002-2135						
Dieter Korn Choose new match	0000-0003-1104						
Wilhelm Boland Choose new match	0000-0001-6784						
Andreas Küttel Choose new match	0000-0003-0235						
Fred Karlsson Choose new match	0000-0001-7536						
Axel Hochkirch Choose new match	0000-0002-4475						
Bernad Batinic Choose new match	0000-0002-8873						
Sándor Csányi Choose new match	0000-0001-7757						
Attila Grandpierre Choose new match	0000-0002-1505						

researcher orcid organization_name organization_country organization_id start_date end_date

researcher remove

Terms
no labels, descriptions or aliases added

Statements

employer

Francis Schaffner (Q30105720)

employer (P108)



University of Zurich (Q206702)

start time
(P580)

January 2007

► 1 references

Jacco Wallinga (Q30105749)

employer (P108)



Leiden University Medical Center
(Q2406409)

start time
(P580)

September 2015

► 1 references



Netherlands National Institute for Public
Health and the Environment (Q1055186)

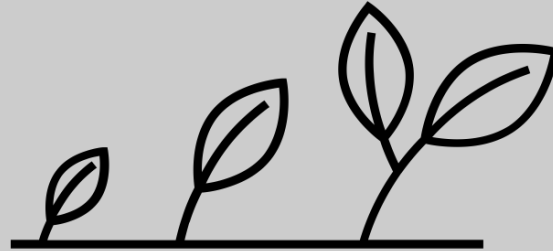
start time
(P580)

June 1997

► 1 references

Thank you

Antonin&Miriam



Slides mashed up with contributions by
Andy Mabbett, Andra Waagmeester and Dario Taraborelli

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

Appendix