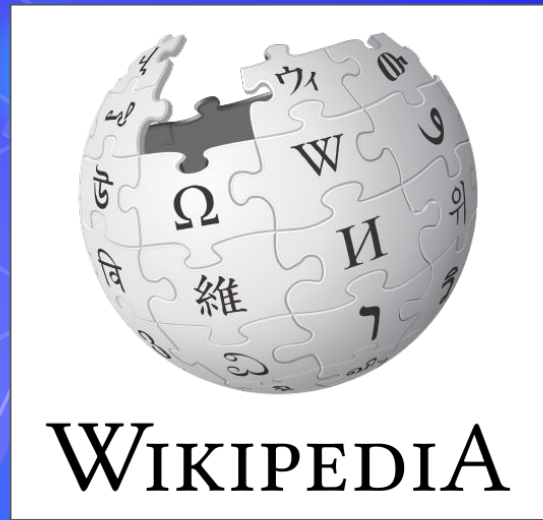# Presenting Scholia - a scholarly profiling tool

**Lane Rasberry**

**Wikimedian in Residence**

**Daniel Mietchen**

**Researcher on Open Science**

**School of Data Science**

**University of Virginia**

2

# **Agenda**

1. overview of Scholia
2. why Scholia matters
3. data and curation process
4. showcase
5. opening your data = contributing

Author  Work ▾  Organization ▾  Location ▾  Event ▾  Project ▾  Award  Topic ▾  Tools ▾

Scholia is a service that creates visual scholarly profiles for topic, people, organizations, s
information in Wikidata.

Scholia relies on Wikidata, and Wikidata contains only a limited albeit growing subset of
Read more about the limitations in the FAQ.

# Search

Search for a scientist, topic, publication, organization, award, event, etc.
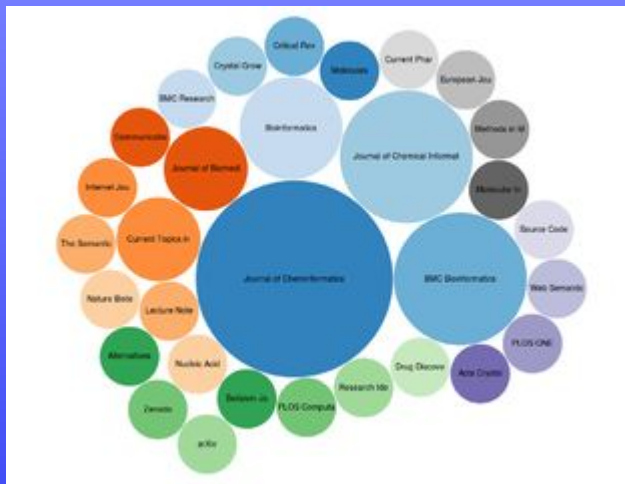
# Examples

| Profiles | Combinations |
|---|---|

5

# Scholia profiles academic literature

Parts:

1. Wikipedia     - prose encyclopedia
2. Wikidata     - structured data complement
3. content     - FAIR and open structured data
4. Scholia     - a front end for Wikidata
5. WikiCite     - project for curating citation data
6. WikiProject     - subject matter community group

# Scholia profiles academic literature

Common profiles:

1. topic
2. person
3. institution

More unusual profiles:

1. events
2. academic trees
3. pathways

Quickly get a report of all academic literature by, about, or from any of these, and share your annotations thereof.

# 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** RSS

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 // Zik... |

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 |

10

# Zika virus infection



co-occurring topic network

# Zika virus infection



## Publications per year

- **first papers in 1952**
- **around 2016: illustration of an event signal in academic publishing**

# Zika virus infection

locations discussed in papers

# Zika virus infection

## Locations of authors of papers on the topic



**Legend:**
- ☑ All layers
- ☑ 🔵 100 < results ≤ 1000
- ☑ 🟠 10 < results ≤ 100
- ☑ 🟢 1 < results ≤ 10
- ☑ 🔴 1 result

# Scholia is in Wikimedia platform

1. uses and frees open data
2. free and accessible on the web
3. develops with Wiki platform
4. anyone editing Wiki contributes
5. strong ethical foundation
6. radically inclusionary and diverse

# **Scholia provides a useful service**

1. orientation to academic literature
2. for researchers and layman audiences
3. largest free and open FAIR option
4. native multilingual support
5. integration with Wikipedia?

# Compare to Google Scholar

# Compare to Google Scholar

1. no export button, no open data
2. non-participatory; no editing beyond own record, no topic tagging for special needs
3. Can't use its API at scale
4. Not allowed to openly license a screenshot
5. 😛

# Compare to Elsevier

# Compare to Scopus/Elsevier

1. export button with weird terms of use
2. if you curate data then it becomes Elsevier's product to sell back to academia
3. unsustainable subscription fee to support <u>30% profit margin</u>
4. 🙂

# Wikipedia's secret strength!

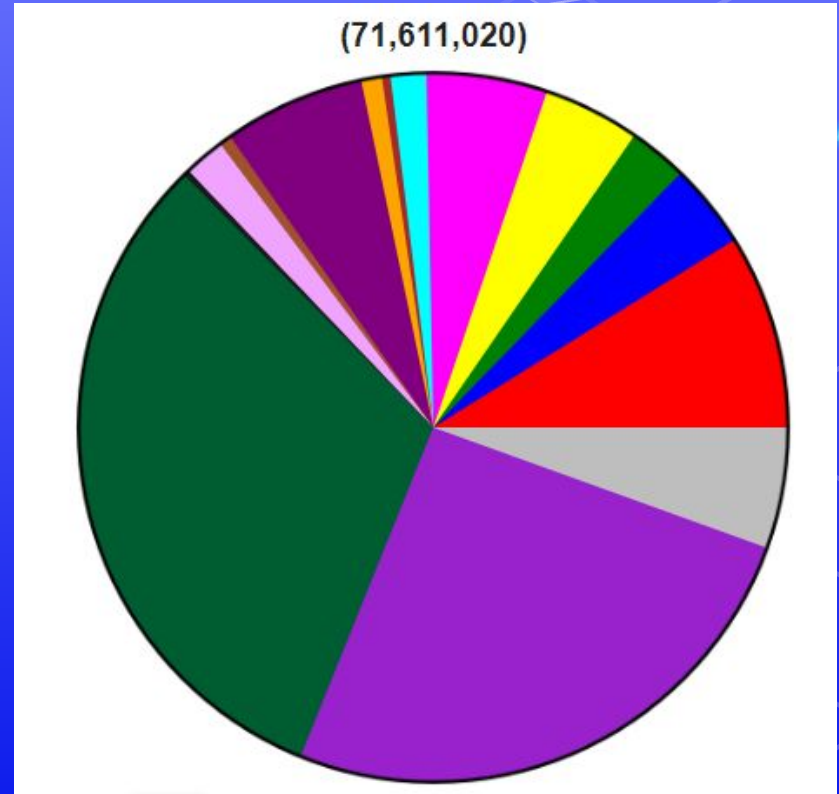Q: How can Wikipedia compete with multibillion-dollar corporate megapowers?

A: 1. Wiki designs for the end user; others do not
2. We do not have the design restriction of only offering products which competitors cannot copy.

# Data and curation

# Overview of Wikidata in general



publications
miscellaneous
???
people
taxon
territory
structure
event

(71,611,020)

# Wikidata numbers

| 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 |

# Overview of WikiCite - links

| Description | Count |
| --- | --- |
| people to education | 1,409,781 |
| people to employer | 1,059,597 |
| people to awards | 1,014,003 |

# Curation workflows

1. topic tagging
   - e.g., "all the papers about diabetes"
2. author disambiguation
   - e.g., "all papers by an author"
3. ontology development
   - e.g., "all papers about math education", get subtopics
4. subject affiliation
   - e.g., "all papers from people at a university"

# 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

# 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

# Group #3

Check all | Uncheck all

| Title | Authors (identified) |
|---|---|
| Revisiting the didactic triangle: from the particular to the general | [1]Simon Goodchild, [2]☐ **Bharath Sriraman** [Full author list] |
| Mediated action in teachers' discussions about mathematics tasks | [1]Claire Vaugelade Berg, [2]Anne Berit Fuglestad, [3]Simon Goodchild, [4]☐ **Bharath Sriraman** [Full author list] |

# Ontology development

Happens in WikiProjects by subject matter interest examples

1.  WikiProject Clinical Trials
2.  WikiProject Universities
3.  WikiProject Lighthouses

# 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

# 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

# Vanderbilt

## Employees and affiliated

Past and present employees, affiliated, and members

Show 10 ▼ entries

| Works | Wikis | Researcher | Researcher description |
| --- | --- | --- | --- |
| 2068 | 3 | Carlo La Vecchia | Italian epidemiologist |
| 657 | 4 | Bert W. O'Malley | American endocrinologist |
| 623 | 0 | Xiao-Ou Shu | physician |

# Vanderbilt

**Wikidata collects lists of awards**
- associates awards to people
- people to institutions

## Awards

Show [10 ▼] entries

| Count | Award | Recipients |
|-------|-------|-----------|
| 25 | John Simon Guggenheim Memorial Foundation Fellowship | Colin Dayan, Oscar Touster, Jo... Blackbourn, James Cullen Mar... Bartels, Mark Jarman, Dean S.... Langdon Brown Gilkey, Nicho... |
| 11 | Fellow of the American Academy of Arts and Sciences | Colin Dayan, Garret A. FitzGer... Blackbourn, Brigid Hogan, Lar... |
| 10 | Fellow of the American Mathematical Society | Akram Aldroubi, Mark Sapir, C... Ralph McKenzie, Mark Ellingha... |
| 9 | AAAS Fellow | Joseph Lee Rodgers, Alan R. B... Roden, Sandra J. Rosenthal, Vi... |
| 6 | IEEE Fellow | Padma Raghavan, Janos Sztipa... Fleetwood |
| 6 | Berlin Prize | Gregg Horowitz, Joy Haslam C... |

# Vanderbilt

## Gender distribution

Count of the number of researchers wrt. gender.

Show [ 10 ▼ ] entries

| Count | Gender |
|-------|--------|
| 2951 | male |
| 2080 | female |

# Clinical trials

Features -

1. opens clinical trials data to web
2. enables matching trials to people, orgs, sponsors, regions, etc
3. combines registries
4. language translation

see also -

1. AACT from Duke
2. ClinWiki
3. Faculty from Vanderbilt
4. WikiProject Clinical Trials

# Clinical trials

## Medical conditions

Show [ 10 ▼ ] entries

| Count | Medical condition | Ex |
|-------|-------------------|-----|
| 9884 | diabetes mellitu | The |
| 7194 | syndrome | Clin |
| 5350 | carcinoma | The |
| 4884 | leukemia | Bor |
| 4677 | lymphoma | Rac Pat |

## Interventions

Show [ 10 ▼ ] entries

| Count | Intervention |
|-------|--------------|
| 5048 | cyclophosphamide |
| 3137 | Anesthetics, Combined |
| 3099 | paclitaxel |

# Clinical trials – data model

expected and nice to have

- ⬡ [ClinicalTrials.gov Identifier](#)
- ⬡ [medical condition treated](#)
- ⬡ [research subject recruitment status](#)
- ⬡ [research intervention](#)

Wikidata disambiguates these

- ⬡ [research site](#)
- ⬡ [principal investigator](#)

# **Scholia is free and open source**

- ⬡ Code on GitHub
- ⬡ on-wiki documentation @ WikiProject Scholia
- ⬡ Active development
- ⬡ Issues and pull requests welcome

# Vision of the future:

This data will be accessible to anyone using

- ⬡   web search
- ⬡   virtual assistants
- ⬡   any tool which uses machine actionable content

Sets the precedent that this data is open and accessible to humans

# **Vision of the future:**

Wikipedia...

- ⬡ biographies link to publication bibliography
- ⬡ school articles link to all publications
  - · images and media on Commons
- ⬡ topics give a profile

Openly accessible in every language

Continuously updated

# Uta Frith

From Wikipedia, the free encyclopedia

⭕ 389 revisions since 2006-10-26 (+11 days), 157 editors, 859 pageviews (30 days), created by: AFdeCH (100) · See full page statistics

**Uta Frith**[4] (*née* **Aurnhammer**; born 25 May 1941)[5] is a German developmental psychologist working at the Institute of Cognitive Neuroscience at University College London. She has pioneered much of the current research into autism[6][7][8][9][10] and dyslexia,[11][12] and has written several books on the two subjects arguing for autism to be considered a mental condition rather than being caused by parenting.[13][14] Her book *Autism: Explaining the Enigma*[15][16] introduces the cognitive neuroscience of autism. She is credited for creating the Sally-Anne test along with fellow scientist Alan Leslie and Simon Baron-Cohen[17][18]. She also pioneered the work with child dyslexia.[13] Among the

**Uta Frith**
DBE, FRS, FBA, FMedSci

45

emma-wesley.co.uk.

63. ^ "Investigating Psychology: CHIPs"⧉. www2.open.ac.uk.

64. ^ "Prof Uta Frith"⧉. UCL IRIS. Retrieved 24 April 2020.

# External links   [ edit source ]

- Official website⧉ ✏
- Profile⧉, ICN Developmental Group
- Profile⧉, UCL IRIS (Institutional Research Information Service)
- Interview with Uta Frith⧉
- "Exploring Autism – A conversation with Uta Frith"⧉, *Ideas Roadshow* (29 March 2013)
- A historical look at the transition from "mentally defective" etc. to autism, August 2014, by Uta Frith⧉

Scholia has an *author* profile for *Uta Frith*.

Scholia has an *author* profile for *Uta Frith*.

46

# How you can participate

○ Open up your data

and/ or

○ Do anything with Wikidata!
- · say hello / join an activity at a Wikidata WikiProject
- · create profiles of people at your org
- · participate in WikiCite

# Credits

- Scholia is supported by the Alfred P. Sloan Foundation under Grant G-2019-11458.
- Presentation template by SlidesCarnival.
- Wikimedia community for providing the shoulders that Scholia stands on.

! ☮ 💙 🌐

Questions?

Lane Rasberry, rasberry@virginia.edu

Daniel Mietchen, daniel.mietchen@virginia.edu