

The New Zealand Thesis Project: A Nation's Dissertations

A Wikimedia and academic library collaboration

**WIKIMANIA
SINGAPORE**

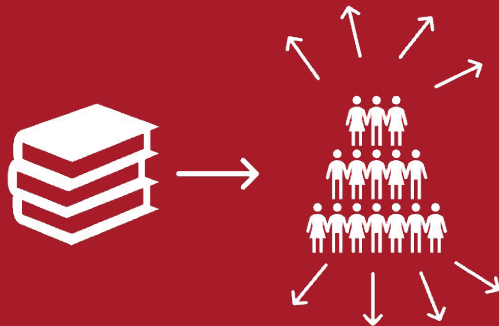
Tamsin Braisher / DrThneed



Aim: to upload bibliographic metadata from our public thesis collections to Wikidata to make it more accessible



66k theses from 13 institutions uploaded



People and main subjects connected



We can ask interesting questions about our data

What is a repository?

- Public collection of outputs
- Exist for theses, reports, scholarly articles, books, conference presentations...
- Some items available for download, some as metadata only
- Data provided by all 8 universities, 5 polytechnics



What is Wikidata?

- Massive linked open database
- > 1 billion items
- > 300 languages
- Machine and human readable
- Contributes to Google searches, voice assistants, infoboxes & more
- Aggregator for identifiers

Thesis links in Wikidata, unlinked author



Institution



Repository

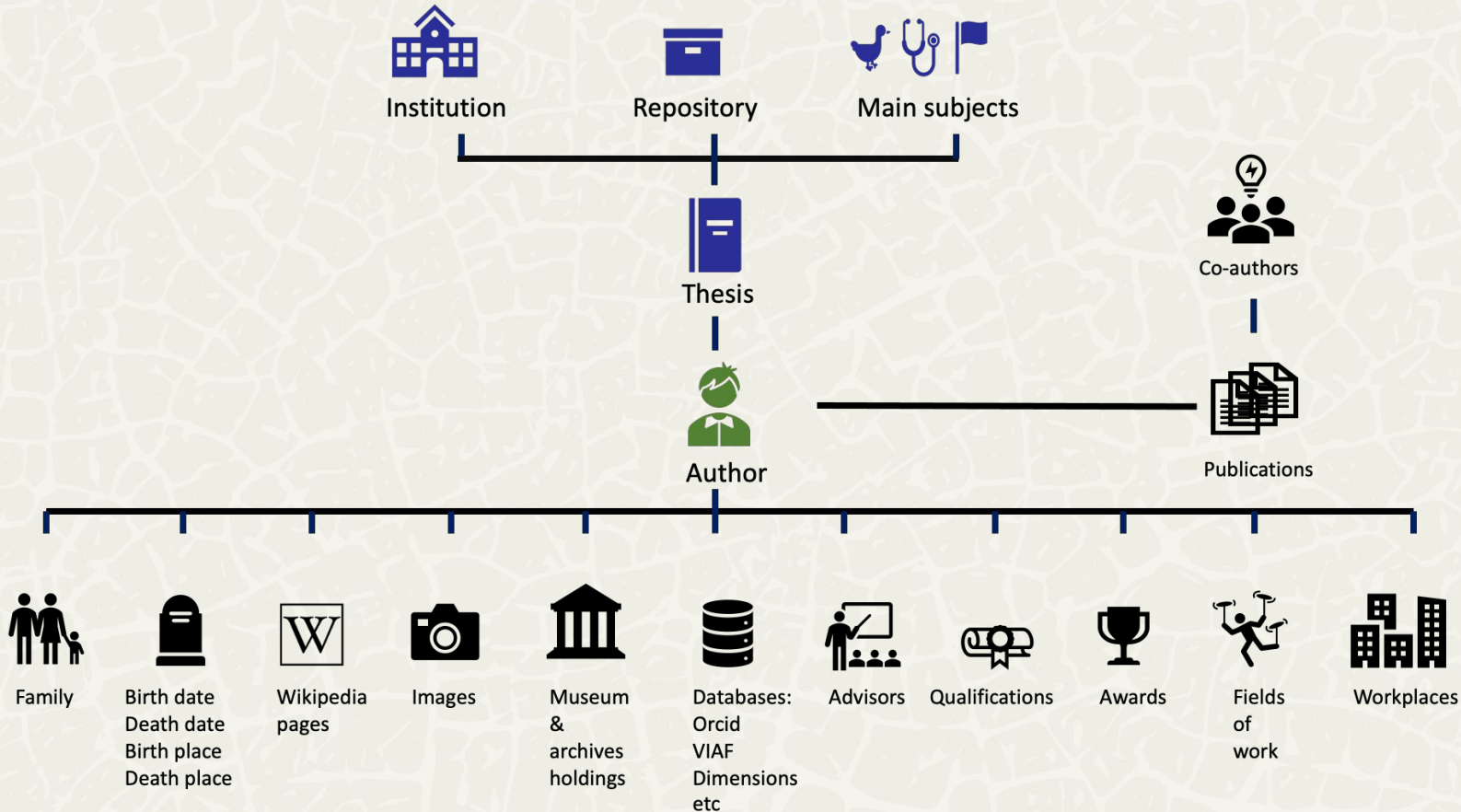


Main subjects



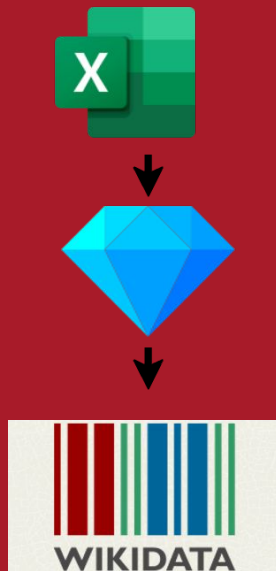
Thesis

Thesis links in Wikidata, linked author



Timeline and process

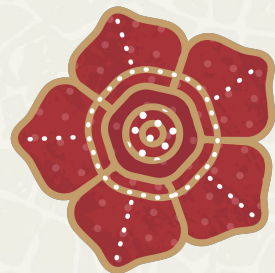
- Institutional buy-in early 2020
- Data aggregation, cleaning, reconciliation March–July
- Upload July–August
- Main subjects, matching of authors continue



- Deborah Fitchett
Academic librarian,
Lincoln University
- Mike Dickison
Wikimedian and librarian
- Tamsin Braisher
Wikimedian, data
wrangler
- Siobhan Leachman
Wikimedian

Dataset description

- >66,000 items
- 13 types of thesis
- Time frame 1907–2022
- Four NZ & Australia-specific controlled vocabularies plus 280,000 rows of freetext keywords
- Handle ID/ DOI
- Several languages
- Missing data, variations in data entry practice
- Qualifications not in Wikidata



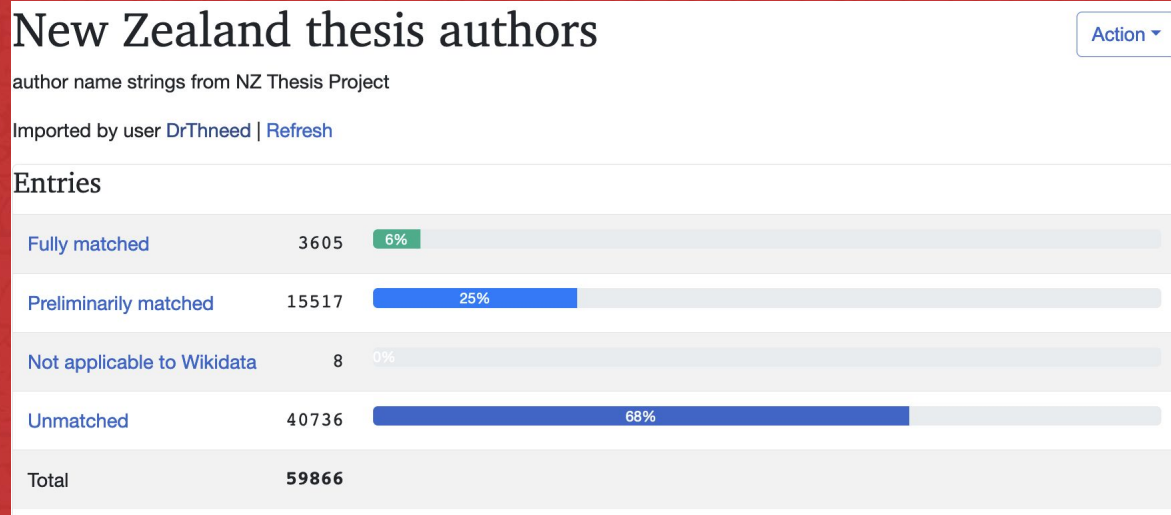
Uploading to Wikidata



- Full schema on project page
- Matched likely subsets of people before upload
- New people items created only when new information
- Unique Wikidata identifiers (QIDs) roundtripped - returned to institutions for inclusion in repository metadata
 - Advisors on 35,000 theses (54%)
 - Linked author for ~10,000 theses (15%)
 - Main subjects for >80% of theses

Matching authors

10% of authors matched before upload of thesis
Mix'n'match catalogue for matching remaining authors



WIKIMANIA
SINGAPORE




New Zealand thesis authors:

C. R. Mason

author of 1987 masters thesis at University of Canterbury titled Rhizomatous legumes for hawkweed dominated grasslands

Preliminarily matched

Cher [Q12003]

Actress, singer, singer-songwriter, composer, film director, character actor, record producer, television actor, model, film actor, manufacturer, and recording artist (*1946) ♀; Primetime Emmy Award, Academy Award for Best Actress, Lucy Award, Grammy Awards, Kennedy Center Honors, and Palme d'Or; member of Sonny & Cher and Allman and Woman; child of John Sarkisian and Georgia Holt; spouse of Gregg Allman and Sonny Bono 

Confirm | **Remove** [all]



**WIKIMANIA
SINGAPORE**



Citing theses on Wikipedia

1400 people in the project are on Wikipedia

Cite thesis on author page, linking back to repository

Link notable authors and advisors



Alexander Gerst

Pages in 35 languages



Beatrice Tinsley

Pages in 33 languages

**WIKIMANIA
SINGAPORE**



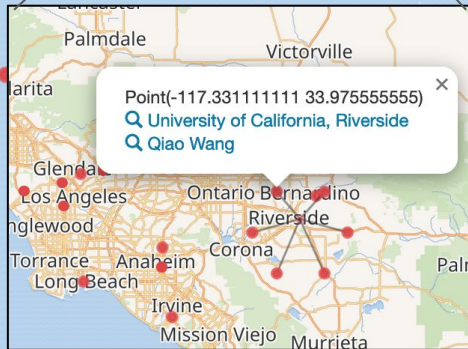
Wikidata visualisations

Run the example queries live at the Wikidata project page
https://www.wikidata.org/wiki/Wikidata:WikiProject_NZThesisProject

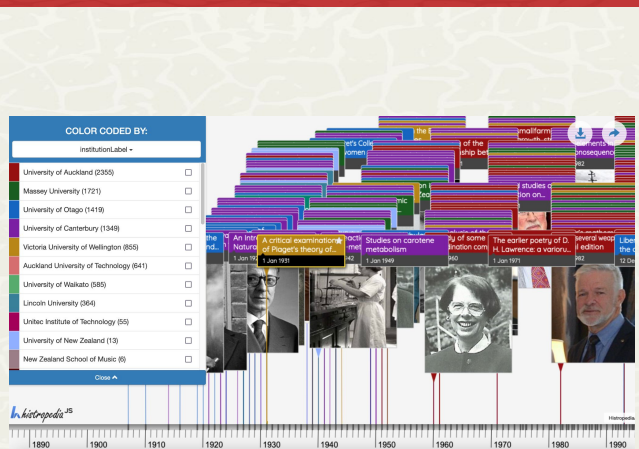
Queries can be run on the entire dataset
or for a specific institution



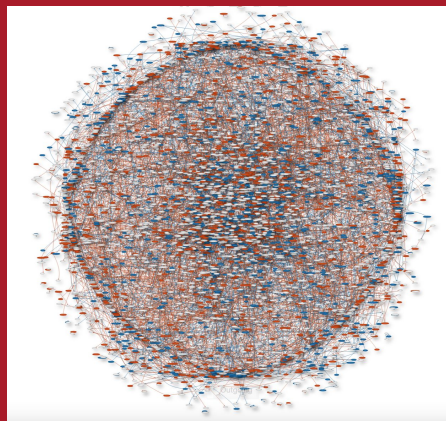
Where have people been employed?



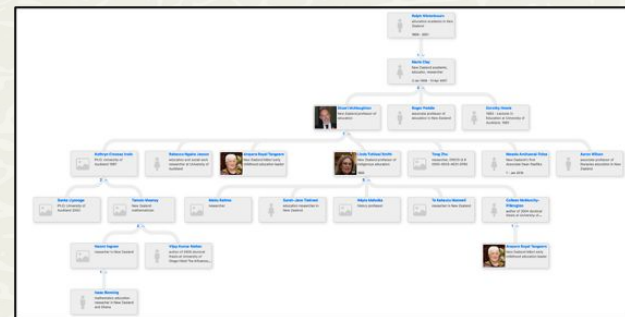
Sample visualisations of thesis data



Histropedia timelines



Doctoral links



Academic family trees

Scholia

organization / Q45135879

Improve data

Department of Zoology, University of Otago (Q45135879)

Table of Contents

- Employees and affiliated
- Co-author graph
- Advisor graph
- Topics that employees and affiliates have published on
- Recent publications
- Uses
- Page production
- Citations
 - Recent citations
 - Most cited papers with affiliated first author
 - Co-author-normalized citations per year
- Awards
- Gender distribution



Employees and affiliated

Past and present employees, affiliated, and members

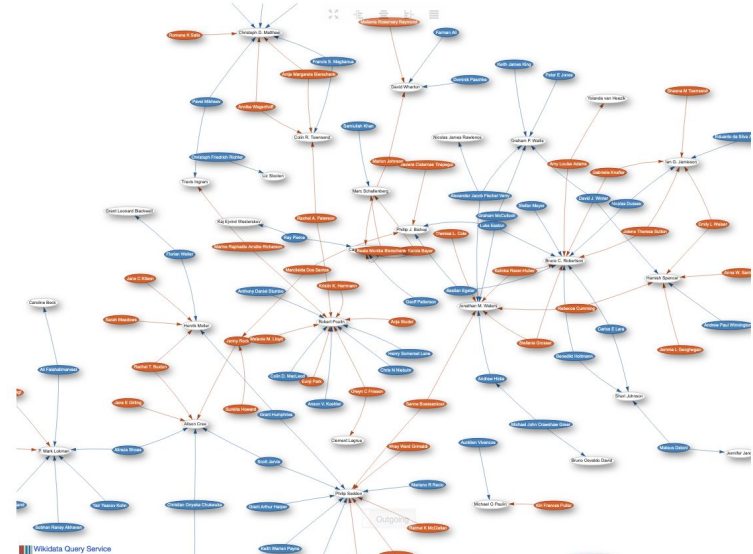
Reload

Show 10 entries

Search:

Works	Wikis	Researcher	Researcher description	Orcid
518	2	Robert Poulin	evolutionary ecologist and parasitologist	0000-0003-1390-1206
216	1	Corey J. A. Bradshaw	Canadian zoologist	0000-0002-5328-7741
158	2	Hamish Spencer	New Zealand evolutionary biologist	0000-0001-7531-597X
139	1	Jonathan M. Waters	zoologist in New Zealand	0000-0002-1514-7916

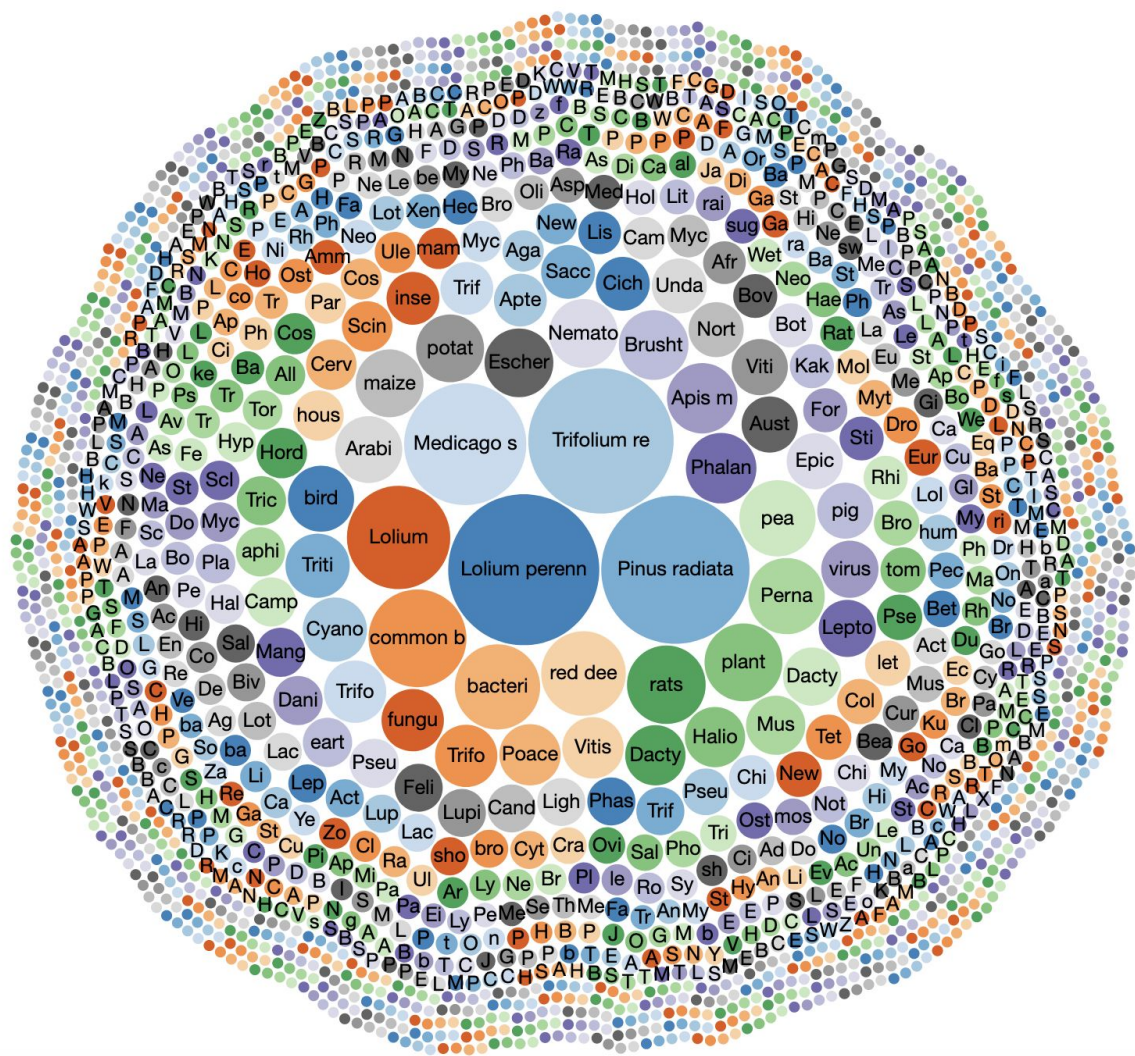
Advisor graph



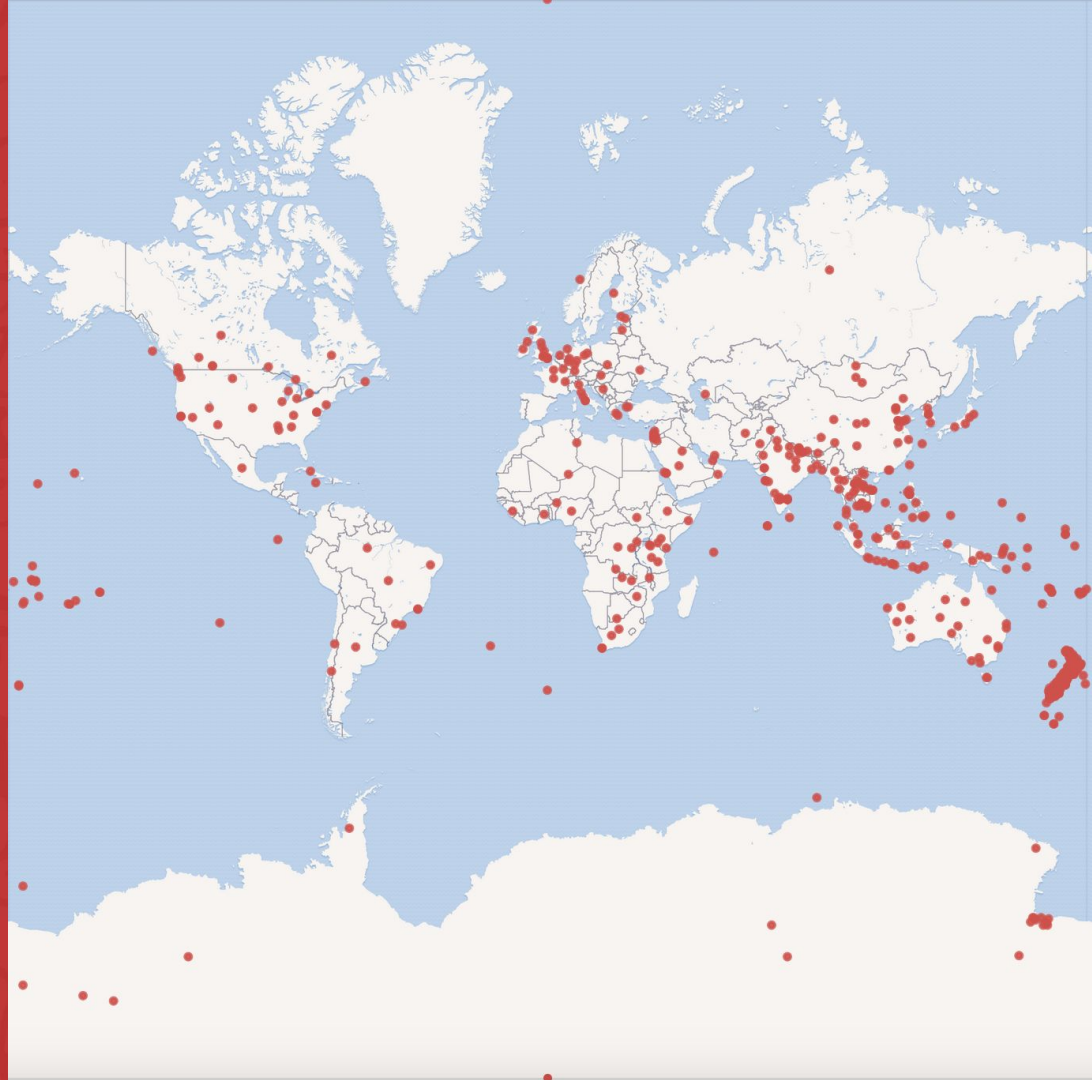
Wikidata Query Service

organization: #45135879-graph.scholia

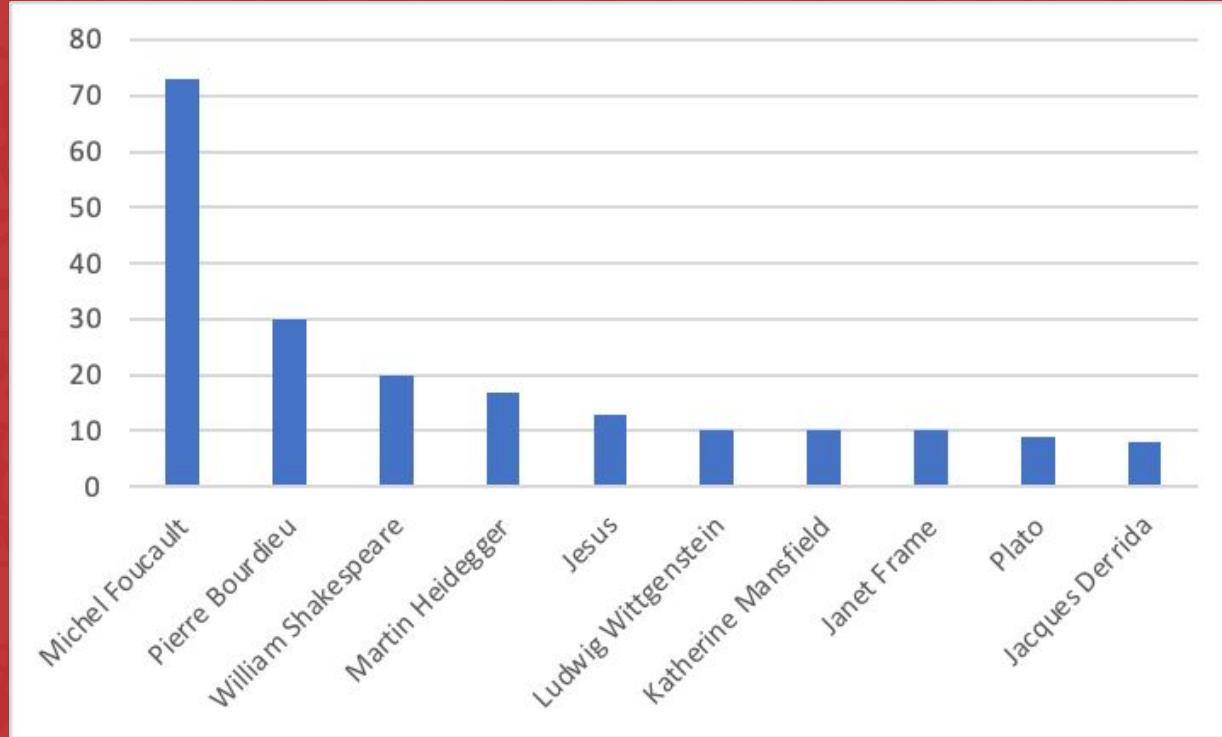
Taxa we study



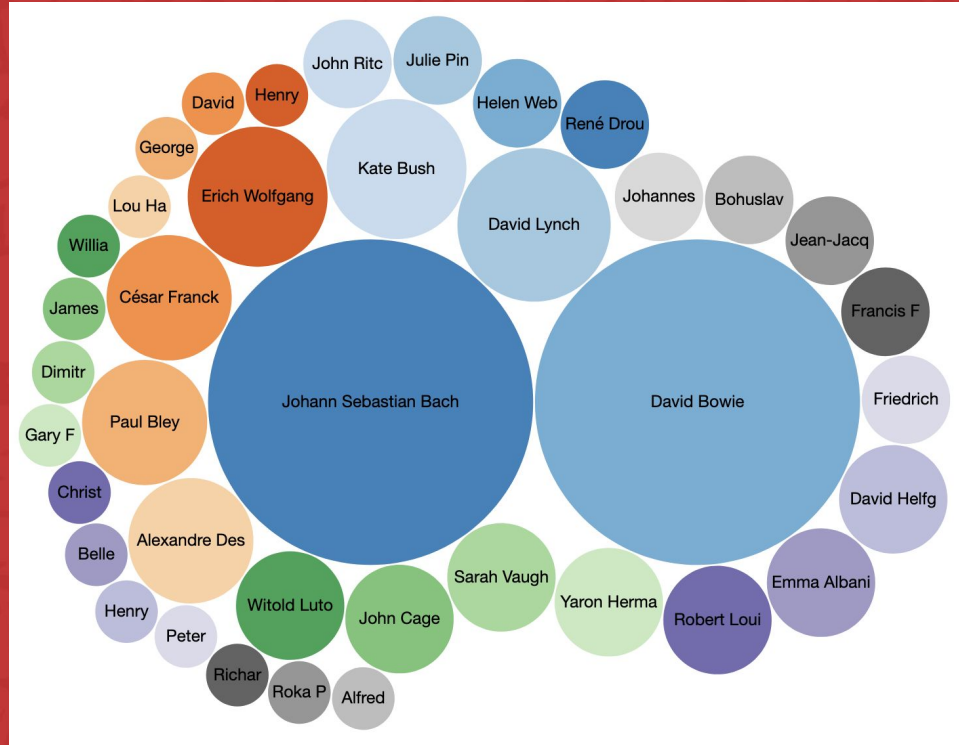
Places we study



Top ten people as main subjects



Bach or Bowie?



Future work

- Project & Wikidata community continue to match authors and add main subjects
- Process for libraries to keep adding theses themselves
- Measure impact in a year or two
- OCLC and ProQuest identifiers for theses
- Connect NZ authors to their theses overseas



Thanks

- Wikidata and Wikipedia community
- WANZ for \$5k data cleaning grant
- WMF for scholarship to Wikimania

Got a repository? Let's talk!

Check out the [Wikidata Thesis Toolkit](#)

**WIKIMANIA
SINGAPORE**



Tamsin Braisher / DrThneed

@KnitMeAThneed on Twitter

@CabbageTree@mastodon.nz



Photo credits

Slide 2 [Wikidata logo](#) public domain

Slide 3 [University of Auckland logo](#), Yvette, CC BY-SA 4.0. [Massey University arms](#), Stanley Bannerman CC BY-SA 4.0. [UC logo](#), University of Canterbury CC0. [VUW logo](#), Victoria University of Wellington, CC0. [University of Waikato image](#), New Zealand Tertiary Union, CC SA 2.0. [University of Otago logo](#), Ulrich Lange, CC0. [Southern Institute of Technology image](#), William Stewart CC BY-SA 2.5. [Otago Polytech logo](#), Otago Polytechnic, CC BY 3.0. [AUT logo](#), Auckland University of Technology, CC0. [Unitec building](#), Ingolfson, CC0. [Lincoln University logo](#), Lincoln University , CC BY 4.0

Slide 4, 5, 19 © Tamsin Braisher, CC BY-SA 4.0

Slide 6 [Excel logo](#) public domain, [OpenRefine logo](#) public domain, [Wikidata logo](#) public domain

Slide 8 [OpenRefine/Wikidata logo](#) by Nikki, public domain

Slide 9 Mixnmatch screenshot CC0

Slide 10 Mixnmatch screenshot CC0. [Cber](#) Amy Stricula CC BY-SA

Slide 11 [Alexander Gerst - Neutral Buoyancy Laboratory Logo](#) by NASA, public domain, [Beatrice Tinsley](#) by Pelopanton, CC-BY-SA 4.0

Slide 12 Image gallery from Wikimedia Commons, various licences. From Sparql query <https://www.wiki.org/wiki/SWDK>

Slide 13 Screenshot Sparql query, CC0 <https://www.wiki.org/wiki/SQfa>

Slide 14 Entitree <https://www.entitree.com/> CC BY-SA 4.0, Histropedia timeline CC BY-SA <http://histropedia.com/faq.html>,
screenshot advisor links Sparql query <https://www.wiki.org/wiki/SVye>

Slide 15 Scholia CC BY-SA <https://scholia.toolforge.org/>

Slide 16/17/18/20 Sparql queries CC0 (<https://www.wiki.org/wiki/SUed>, <https://www.wiki.org/wiki/Sres>, <https://www.wiki.org/wiki/SrpZ>, <https://www.wiki.org/wiki/SvubE>)

Slide 23 [WANZ logo](#) Michael Klajban CC BY-SA 4.0, Photo: Stephen A'Court, under contract by WANZ, CC-BY SA 4.0

Thanks to Adien Gunarta and Naila Rahmah for slide designs and Wikimania 2023 illustrations, CC0

WIKIMANIA
SINGAPORE

