

Theses and Wikidata: Squishing a Sample

- Tamsin Braisher/DrThneed

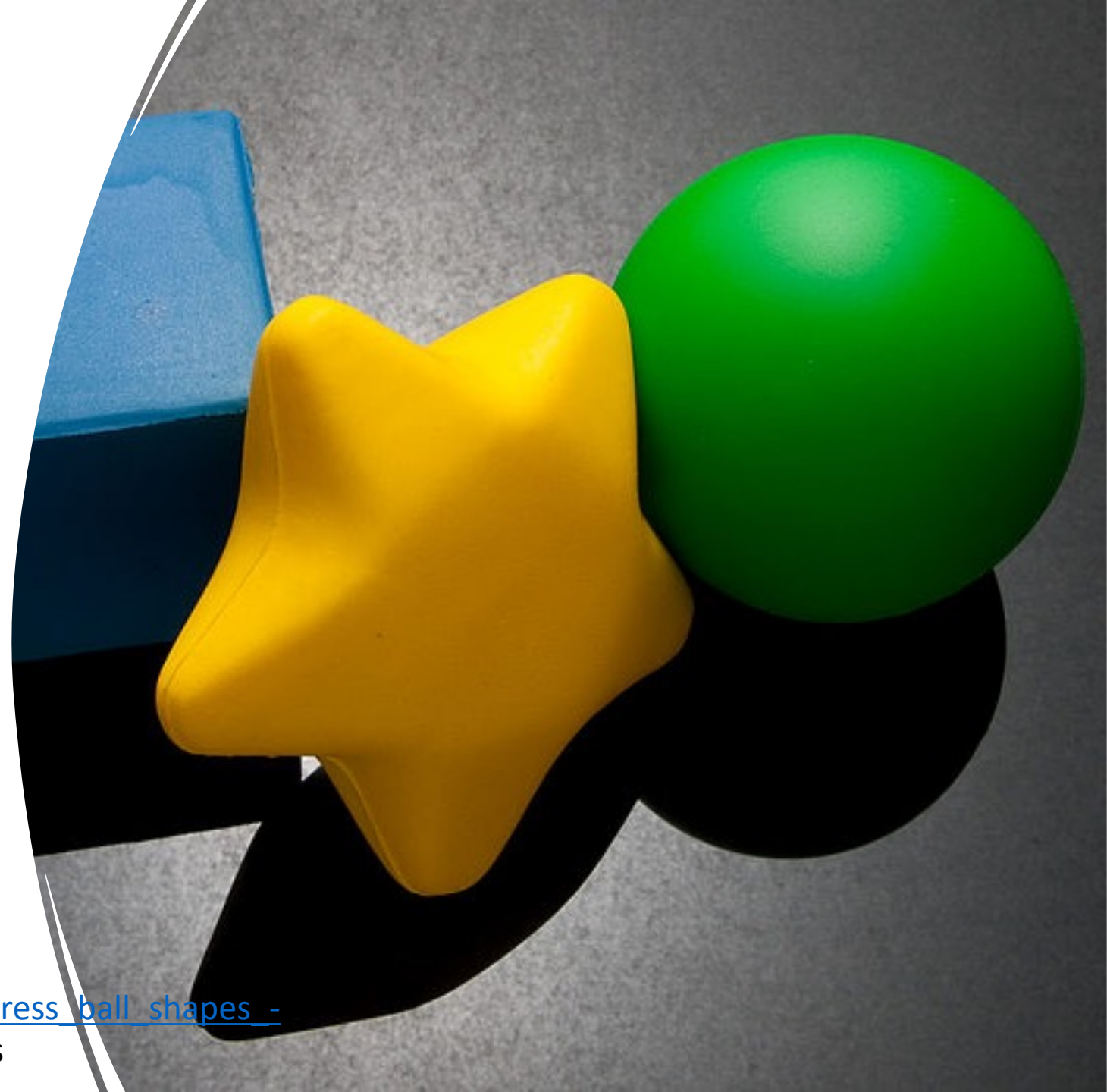


Image: [https://commons.wikimedia.org/wiki/File:Stress_ball_shapes_-_1_\(3693412091\).jpg](https://commons.wikimedia.org/wiki/File:Stress_ball_shapes_-_1_(3693412091).jpg) CC BY-SA 2.0 Stephen Edmonds

Squishing the sample data

- Data clean up – sample 116 items (of 66,467 from 13 institutions)
- Matching to Wikidata, uploading
- Querying the sample
- What's next

Data clean up - title

- Removing extra text from titles
 - “:|ba thesis ...”
 - “: Thesis ...”
 - “A thesis”
 - “: a dissertation”
 - “: [Thesis”
- Upper case titles were selected by regex and case converted, with proper nouns reinserted manually.
- Sentence case desired for Wikidata, but some titles are Title Case. All 116 sample theses were changed, but takes too long to be feasible for the entire data set.

Possible duplicates

- Duplicates: ~70 items withdrawn and reissued (after embargo?), will remove withdrawn ones from dataset.
- Others we have queried with some of you, e.g. thesis held at two institutions, ones that just appear twice in DSpace with different handles

Data clean up - names

- Reversed author names to match Wikidata norm of First name Last name
- Pseudonyms alternative names
 - Animal, Chirp, 392 etc (23 total)
 - Alternative name e.g Asian/English, married names
 - Sometimes same name duplicated (with double pipe) *filter out*
 - Multiple authors (e.g. design)
 - 2 authors with ORCID identifier (78 in entire set)



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

Geographical data
Create a new Lexeme
Recent changes
Random Lexeme

Wikidata Edit
VEF: Settings
VEF: AdmEntity
VEF: Article
VEF: Award
VEF: Book / Journal
VEF: Entity
VEF: External ID

Item [Discussion](#) [Read](#) [Labels list](#) [View history](#)

Katrine Hursthouse (Q111323900)...

ORES predicted quality: D (2.08)

conservationist in New Zealand

Katrine Fearon Hursthouse | Animal | KF Hursthouse | Katrine Fearon Miller | Kae Miller

[Recoin: Most relevant properties which are absent](#)

[In more languages](#)

Configure

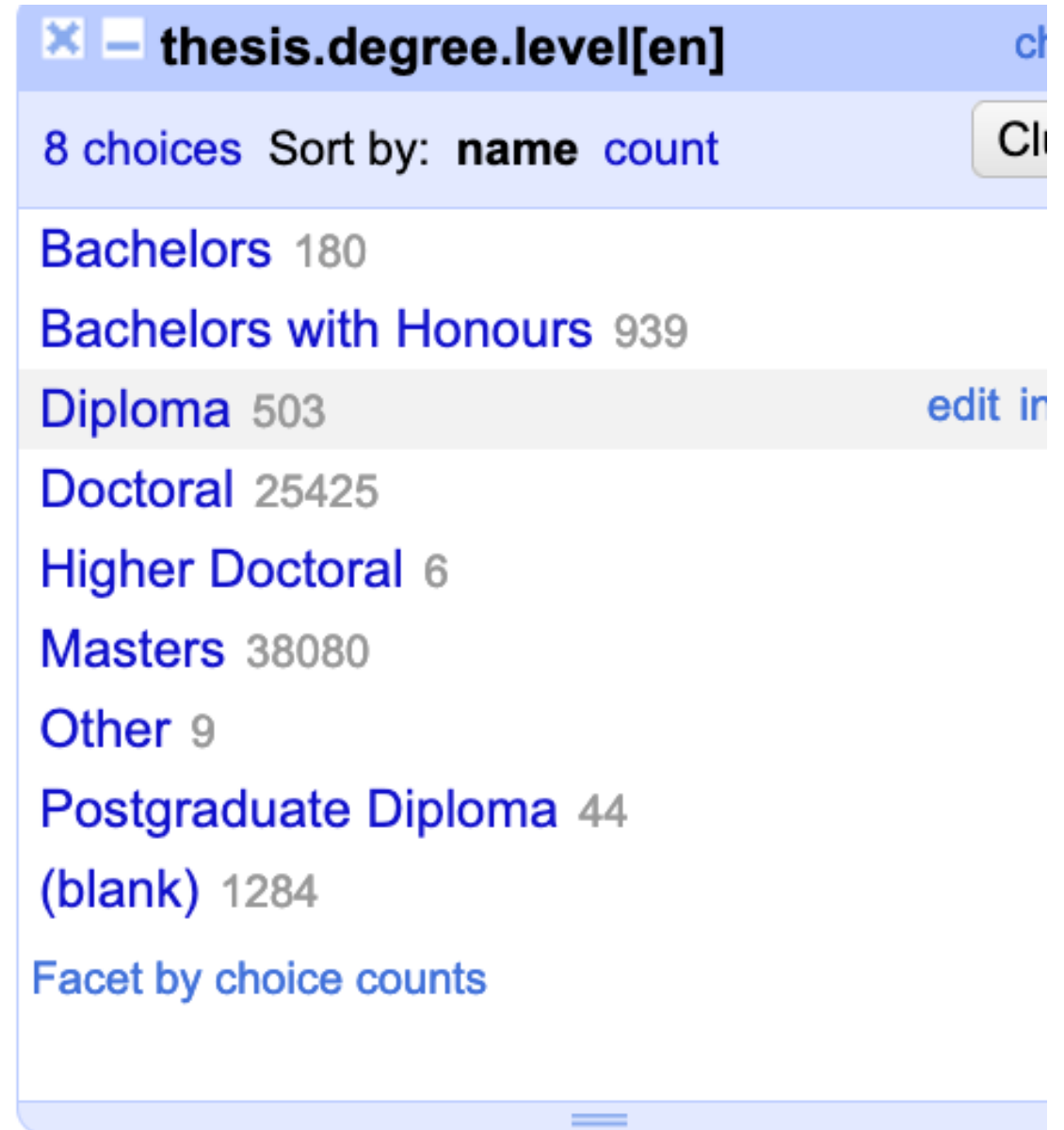
| Language | Label | Description | Also known as |
|----------|-----------------------------------|---|--|
| English | Katrine Hursthouse by DrThneed | conservationist in New Zealand by DrThneed | Katrine Fearon Hursthouse Animal KF Hursthouse Katrine Fearon Miller Kae Miller by DrThneed |

Mix'n'match entries for this item

- ✓ [Turnbull Names 2 Katrine Fearon Miller](#)
([entry in catalog](#)) [Hursthouse, Kae] - 1911:1994 - Conservationist. See INNZ, and ATL Biographies Index. Photographed by news photographer, J J Short ca 1983

Data cleaning

- Type of thesis (thesis degree level)
 - Sample: doctoral, masters, and one undetermined, entered as “thesis”
 - 1284 are blank, type can be inferred from degree name instead
 - 109 are blank for both columns, will check manually. Any that can't be determined will be entered as 'thesis'



thesis.degree.level[en] 8 choices Sort by: name count

| | | |
|------------------------|-------|-------------------------|
| Bachelors | 180 | |
| Bachelors with Honours | 939 | |
| Diploma | 503 | edit in |
| Doctoral | 25425 | |
| Higher Doctoral | 6 | |
| Masters | 38080 | |
| Other | 9 | |
| Postgraduate Diploma | 44 | |
| (blank) | 1284 | |

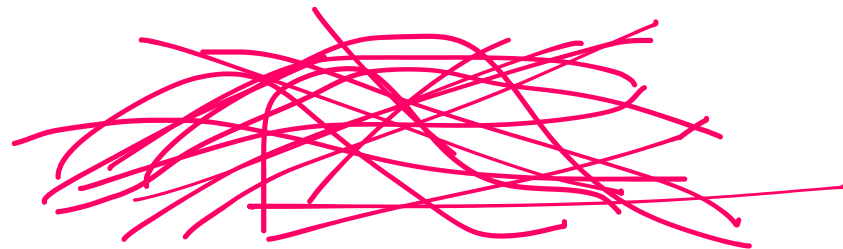
Facet by choice counts

Repository

- Created a Wikidata item for each repository, used for “publisher” field. Mapped from the institution that supplied the data.

“Dissertation submitted to”

- More complicated than it appeared!
 - Not consistent across the columns
 - 1. grantor current name
 - 2. grantor old name
 - 3. institution supplying data
 - 4. extra title information
 - 5. publisher
- Submitted to: “University of New Zealand”
 - qualifier: “named as” e.g. Canterbury Agricultural College, University of New Zealand
 - Will go very carefully with big dataset



Building descriptions:
“1998 master’s thesis by A
Smith at University of Otago”

- Built from year
- Type of thesis
- Author name (Wikidata form)
- The institution that holds the data

Q | clarke canterbury ⓧ **Search**

Advanced search: Sort by relevance ⓧ ⌵

Search in: (Main) ⓧ Property ⓧ ⌵

[Social media: a new religion? \(Q111290530\)](#)
2020 master's thesis by Ashleigh Nicola **Clarke** at University of **Canterbury**
10 statements, 0 sitelinks - 22:11, 22 March 2022

[Clarke Pass \(Q32160807\)](#)
gap in New Zealand
5 statements, 1 sitelink - 13:07, 6 August 2017

Digital Hymnology: [Canterbury Dictionary of Hymnology, Hymn Tune Index, Hymnary.org \(Q58165703\)](#)
article by Martin V. Clarke published 4 August 2016 in Nineteenth Century Music Review
9 statements, 0 sitelinks - 00:01, 22 January 2022

[Clarke Monument In Churchyard About 2.5 Metres South Of Tower, Church Of St Thomas Of Canterbury \(Q26629716\)](#)
monument in Lovington, South Somerset, Somerset, UK
8 statements, 0 sitelinks - 02:17, 21 February 2022

Resulting Wikidata entry



| Language | Label | Description | Also known as |
|----------|---|---|---------------|
| English | Te mana o te Ao Turoa hei tupu i te koru by DrThneed | 2020 thesis by Wiki Walker at University of Canterbury by DrThneed | |

[All entered languages](#)

Statements

| | | | |
|--|--|--------------|--|
| instance of by DrThneed | doctoral thesis ... | 0 references | |
| title by DrThneed | Te mana o te Ao Turoa hei tupu i te koru (Maori) | 0 references | |
| author name string by DrThneed | Wiki Walker ... | 0 references | |
| publisher by DrThneed | UC Research Repository ... | 0 references | |
| dissertation submitted to by DrThneed | University of Canterbury ... | 0 references | |
| country of origin by DrThneed | New Zealand ... | 0 references | |
| language of work or name by DrThneed | English ... | 0 references | |
| publication date by DrThneed | 2020 ... | 0 references | |

Identifiers

| | | | |
|--------------------------|----------------|--------------|--|
| DOI by DrThneed | 10.26021/10801 | 0 references | |
| Handle ID by DrThneed | 10092/101748 | 0 references | |

Reconciliation

- 116 dissertations
 - 16 reconciled to an existing author
 - Advisors also reconciled
 - Enabled connection of authors to supervisors
- Student item:
 - “educated at”
 - “academic thesis”
 - “doctoral advisor” or “student of”
 - Advisor item:
 - “doctoral student” or “student”
 - Thesis item:
 - Qualifier “doctoral advisor” on “dissertation submitted to” statement
 - “Author” instead of “author name string”

| ▼ author first last | ▼ Name to reconcile | ▼ Date |
|--|--|---------|
| Janet Stephenson | Janet Stephenson Choose new match | 2005 |
| Jane Eleanor Girling | Jane E Girling Choose new match | 1998-06 |
| Guillaume Aucher | Guillaume Aucher Choose new match | 2008 |
| Brendan J. Hokowhitu | Brendan Hokowhitu Choose new match | 2002 |
| Wendy Zhang Choose new match | Wendy Zhang Choose new match | 2021 |

| ▼ author first last | ▼ Name to reconcile | ▼ Date |
|-----------------------|--|-----------|
| Sue Bradford | Sue Bradford Choose new match | 2014 |
| Hayden Joel Pritchard | Hayden J. Pritchard Choose new match | 2017 |
| A. H. Flay | Albert Flay Choose new match | 1928 |
| L.W. McCaskill | Lance McCaskill Choose new match | 1929 |
| I. D. Blair | Ian Douglas Blair Choose new match | 1935 |
| John Dunstan Atkinson | Torchy Atkinson Choose new match | 1932 |
| Verity Frances Oliver | Verity F Oliver Choose new match | 2010-04-0 |



Queries for each institution all of them (title only) <https://w.wiki/4yXT>

- Auckland Uni <https://w.wiki/4yYB>
- AUT <https://w.wiki/4yYE>
- Canterbury <https://w.wiki/4yit>
- Waikato <https://w.wiki/4yXc>
- VUW – Te Herenga Waka <https://w.wiki/4yiu>
- Otago Uni <https://w.wiki/4yix>
- Otago Polytechnic <https://w.wiki/4yiy>
- EIT <https://w.wiki/4yiz>
- Lincoln Uni <https://w.wiki/4yj3>
- Massey <https://w.wiki/4yj4>
- SIT <https://w.wiki/4yj5>
- Toi Ohomai <https://w.wiki/4yj6>
- Unitec <https://w.wiki/4yj7>
- University of New Zealand <https://w.wiki/4yis>

What next

- Feedback – questions?
Problems?
- Processing and uploading the main dataset ~66k dissertations
- Expanding items
 - Access status - digitised or not
“full work available at URL”, license
 - Advisor names
 - Key words/main subjects
 - Degree program (list of degrees?)
- Wikidata Q numbers for you to put in Dspace, to make updating easy
- Using the data!



How can the data be used?

- Citations - #1Lib1ref (May), how to use these Wikidata items on Wikipedia
- Linking to author items
 - People with alternative names
 - Session on connecting theses to authors, creating new authors?
- Scholia

Percival Clennell Fenwick

MĀORI ›

Born 1870 - Died 6 July 1958



SERVICE NUMBER N/N AWMM
3/158A AWMM
15/118 AWMM

ARMED FORCE / BRANCH [Army](#) AWMM

LAST RANK South Africa [Surgeon Captain](#) AWMM
[New Zealand Mounted Rifles, 2nd Contingent](#) AWMM
WW1 [Lieutenant Colonel/Military](#) AWMM
[New Zealand Medical Corps](#) AWMM

WAR [South African War, 1899-1902](#)



Scholia

<https://scholia.toolforge.org/>

organization / Q1201513

Improve data

University of Otago (Q1201513)

The University of Otago is a public university based in Dunedin, Otago, New Zealand. It scores highly for average research quality, and in 2006 was second in New Zealand only to the University of Auckland in the number of A-rated academic researchers it employs. In the past it has topped the New Zealand Performance Based Research Fund evaluation. [\(Read more on English Wikipedia\)](#)

Related: [University of South Carolina](#) · [American University](#) · [University of Canterbury](#) · [Victoria University of Wellington](#) · [University of South Florida](#) · [National University of Colombia](#) · [University of Hawaii](#) · [George Mason University](#) · [University of British Columbia](#) · [Queen Mary University of London](#)

@otago | <https://ror.org/01jmx844>

Table of Contents

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



Employees and affiliated

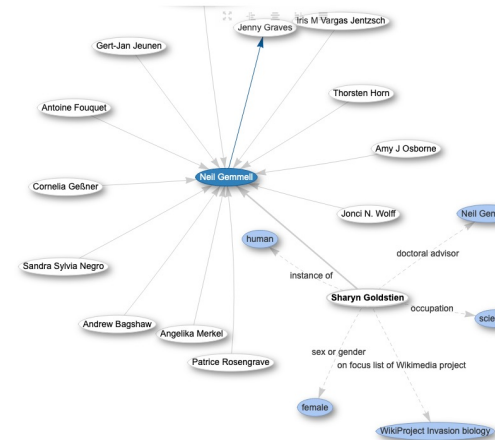
Past and present employees, affiliated, and members

Show 10 entries

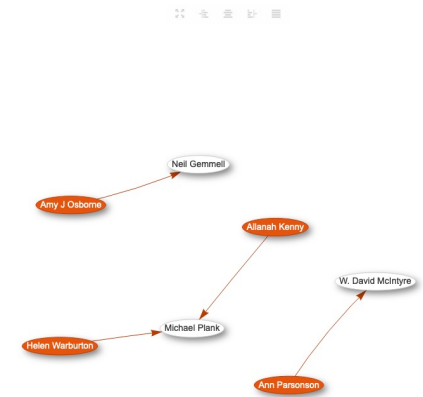
Search:

| Works | Wikis | Researcher | Researcher description | Orcid |
|-------|-------|--------------------------------|-----------------------------|-------------------------------------|
| 667 | 0 | Chris Frampton | New Zealand biostatistician | 0000-0003-0603-5661 |
| 448 | 3 | Murray Brennan | New Zealand surgeon | 0000-0003-2358-4371 |

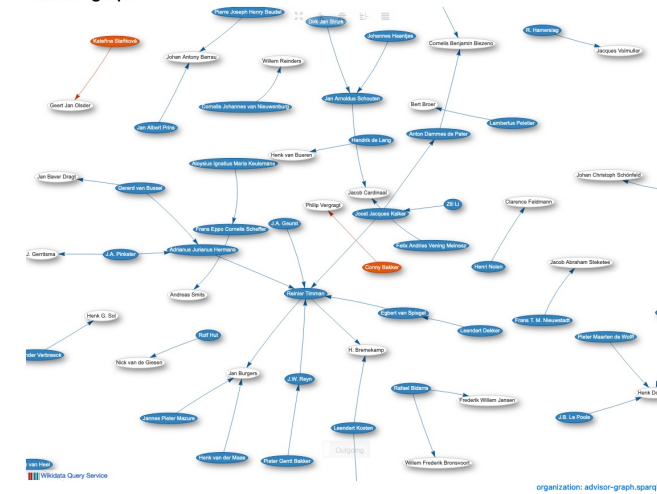
Academic tree



Advisor graph



Advisor graph



organization: advisor-graph.sparql