



NZ Thesis Project: A Nation's Dissertations

Introduction to project

- “to upload basic bibliographic metadata from our public thesis collections into Wikidata”
- Idea launched at Aotearoa IR Community Days 2021
- One-pager proposal sent around to get mgmt support
- All 8 unis agreed to provide data - plus ResearchBank (5 polytechs)



- Instructions for extracting data from DSpace
- Preliminary data-cleaning and merged into a single file (66,470 lines)

What is Wikidata?

Linked open database

>300 languages

Machine and human-readable

Openly licensed

Contributes to Wikipedia, voice assistants, Google results, infoboxes etc

Aggregator for identifiers



Overview of the data

- 1907–2022, but only 657 from pre-1962.
- Additional material (papers, reports,...)
- Mostly single authors but some multiple.
- Ten types of thesis (3 types new to Wikidata), mostly doctoral (25,514) and masters (38,736).
- Institutional name changes, shared theses, duplicates, embargoed/withdrawn (Handle ID change)
- Subject keywords (free text, up to 100!), Māori subjects, plus three controlled vocabularies.
- Degree names - very variable forms
- Advisors for 46k theses. Up to 10, many initialized.

Cleaning up

- Cleaning titles with citations in mind - removing extra text from titles (‘: a thesis submitted for the partial fulfilment of the degree of...’)
- Author names mostly surname, first name (but sometimes first name, surname!), some with dates or aliases /pseudonyms or titles
- Many names initialized
- Descriptions - e.g. “1999 doctoral thesis by Janine Smith at the University of Otago”
- Standardised degrees to NZQA-approved name



Uploading to Wikidata

- Cleaned and uploaded (through OpenRefine) a basic set of statements for 66,225 theses
- Statements include
 - Type of thesis
 - Institution it was submitted to
 - Year it was completed
 - Author or author name string
 - Handle ID/DOI
- 7k Wikidata theses matched to an author
- 59k entered with an author name string
- Advisors added on 22k theses
- >9300 people disambiguated in total so far
- Many qualifications created (mostly master's degrees)



QIDs have been returned back to repository owners for entry into database

The genome and beyond: phenotypic determinants of two *Trichoderma cf. atroviride* sister strains

Lange, Claudia

Abstract

Trichoderma cf. atroviride isolate LU132 is used commercially in New Zealand as a biological control agent (BCA) on various economically important plants to protect against plant pathogens. Its broad spectrum of antagonistic activity is linked to a wide range of biological parameters but we have limited knowledge of what specific attributes make this isolate particularly effective. While LU132 is a remarkably efficient BCA compared to *T. cf. atroviride* isolate LU140, LU132 was found to be highly similar to LU140 at the DNA sequence level. This unusual combination of phenotypic variability and... [Show full abstract]

Keywords

biocontrol; genomics; epigenetics; cytosine methylation; *Trichoderma atroviride*; mycoviruses; single nucleotide polymorphism

Fields of Research

06 Biological Sciences; 0605 Microbiology; 0604 Genetics

Date

2015

Type

Thesis

Collections

Bio-Protection Research Centre [366]

Doctoral (PhD) Theses [977]



View/Open

Lange_PhD.pdf

Share this



<https://hdl.handle.net/10182/6675>

Metadata

Expand record

dc.subject.anzsrc 06 Biological Sciences

dc.subject.anzsrc 0605 Microbiology

dc.subject.anzsrc 0604 Genetics

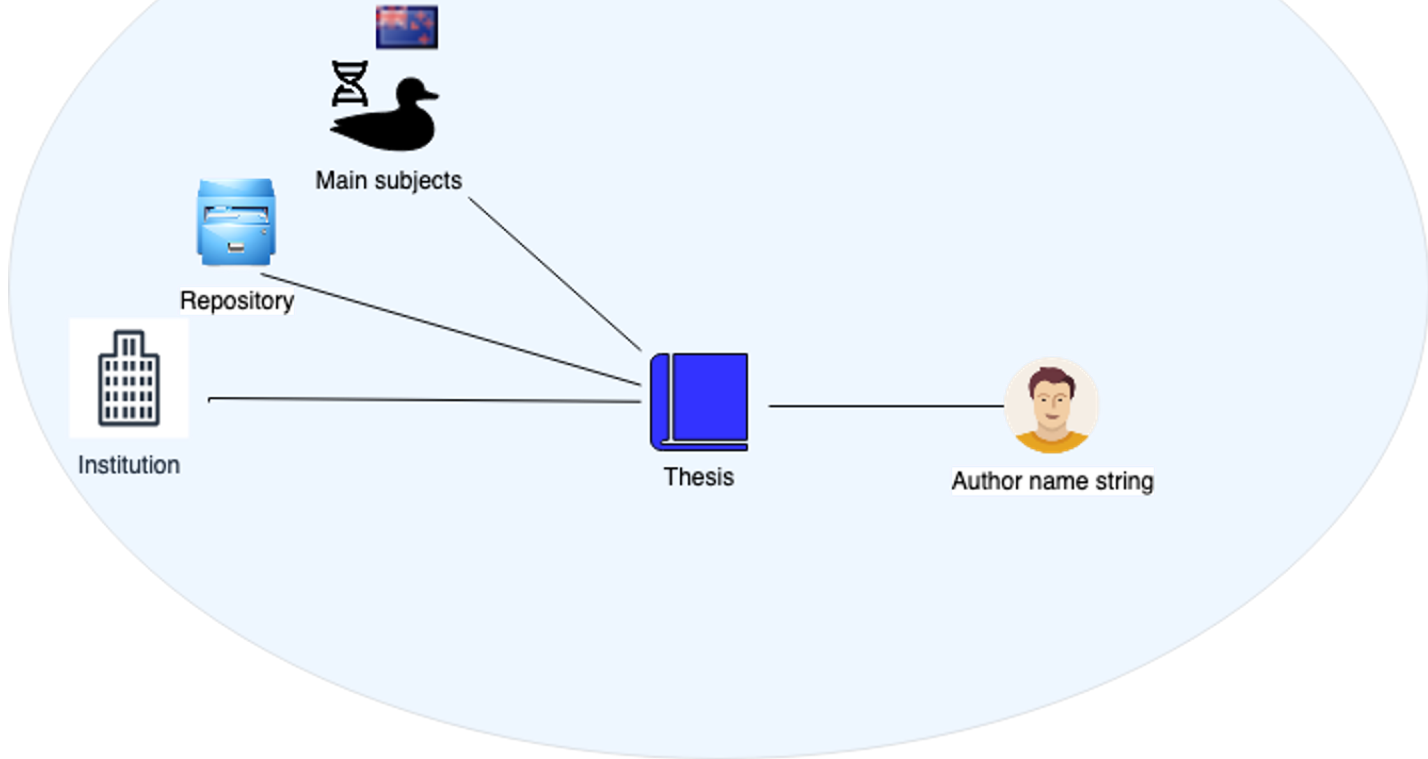
dc.identifier.wikidata Q111964970

Show full item record
Expand record

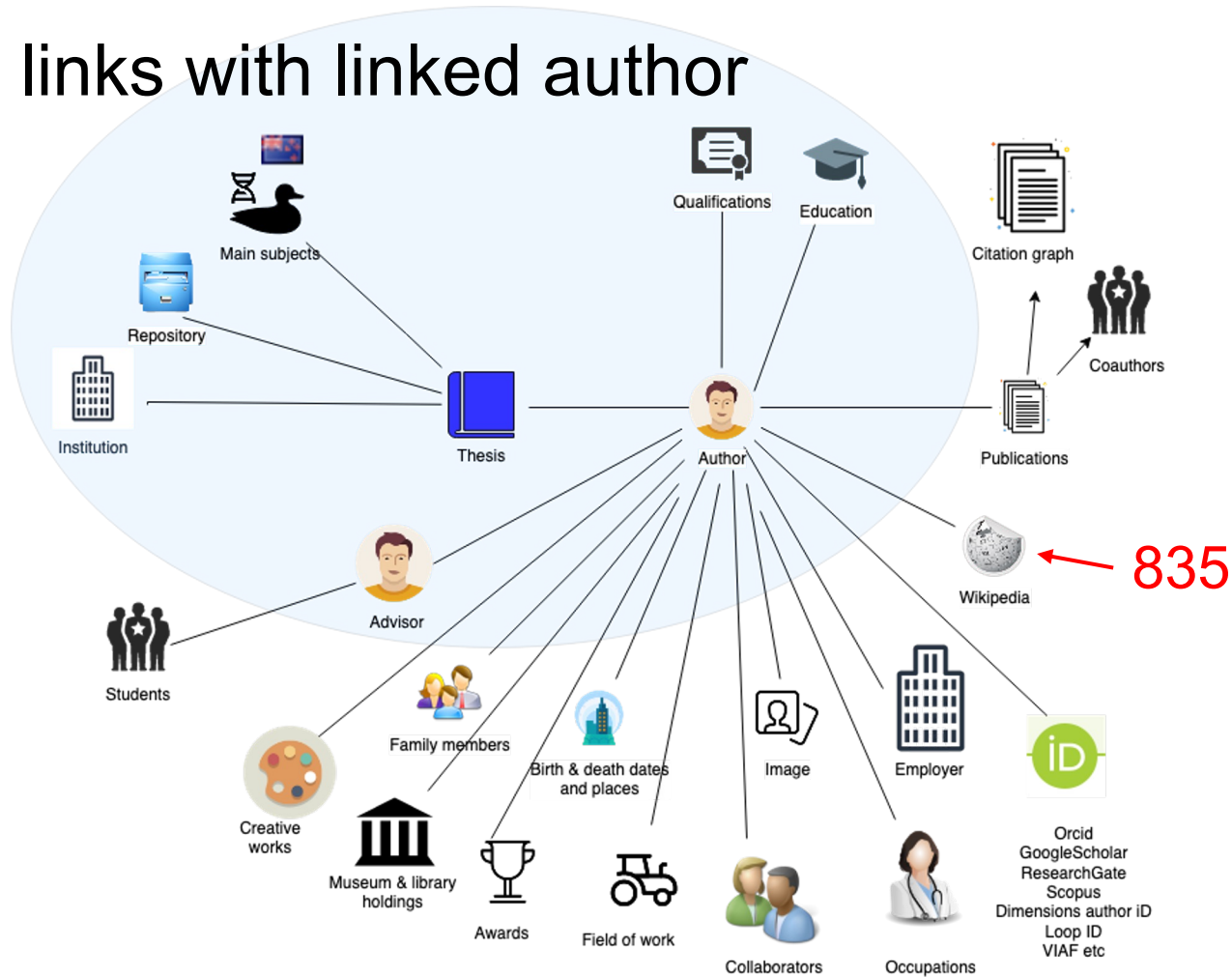
dc.identifier.wikidata Q111991156

otago.wikidata.url <https://www.wikidata.org/wiki/Q111991156>

Thesis links in Wikidata - author name string



Thesis links with linked author



Matching strategies - people

Find subsets of Wikidata to match against first

- FRSNZ fellows
- People who won an award from RSNZ
- People with a NZ honour
- People employed at or educated at a NZ educational institution, or a CRI

cell.cross names between the projects (no fuzziness, so many names missed but also many matched)

Cluster algorithm in OpenRefine found some people with >1 thesis

Ordered alphabetically, looked for duplicates with different spelling

Smartnames query for Orcid, if match, query Wikidata for ORCID id

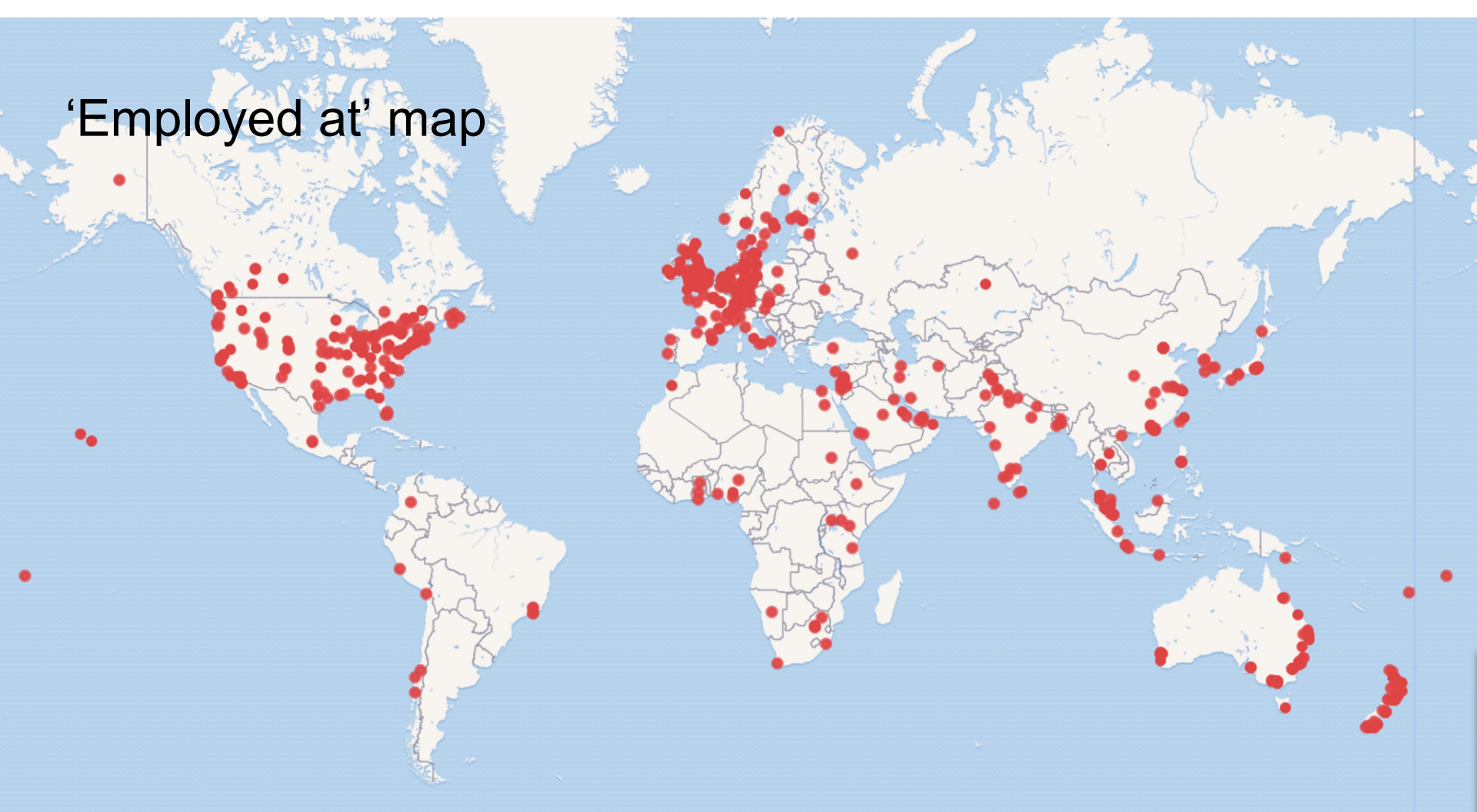


Wikidata visualisations

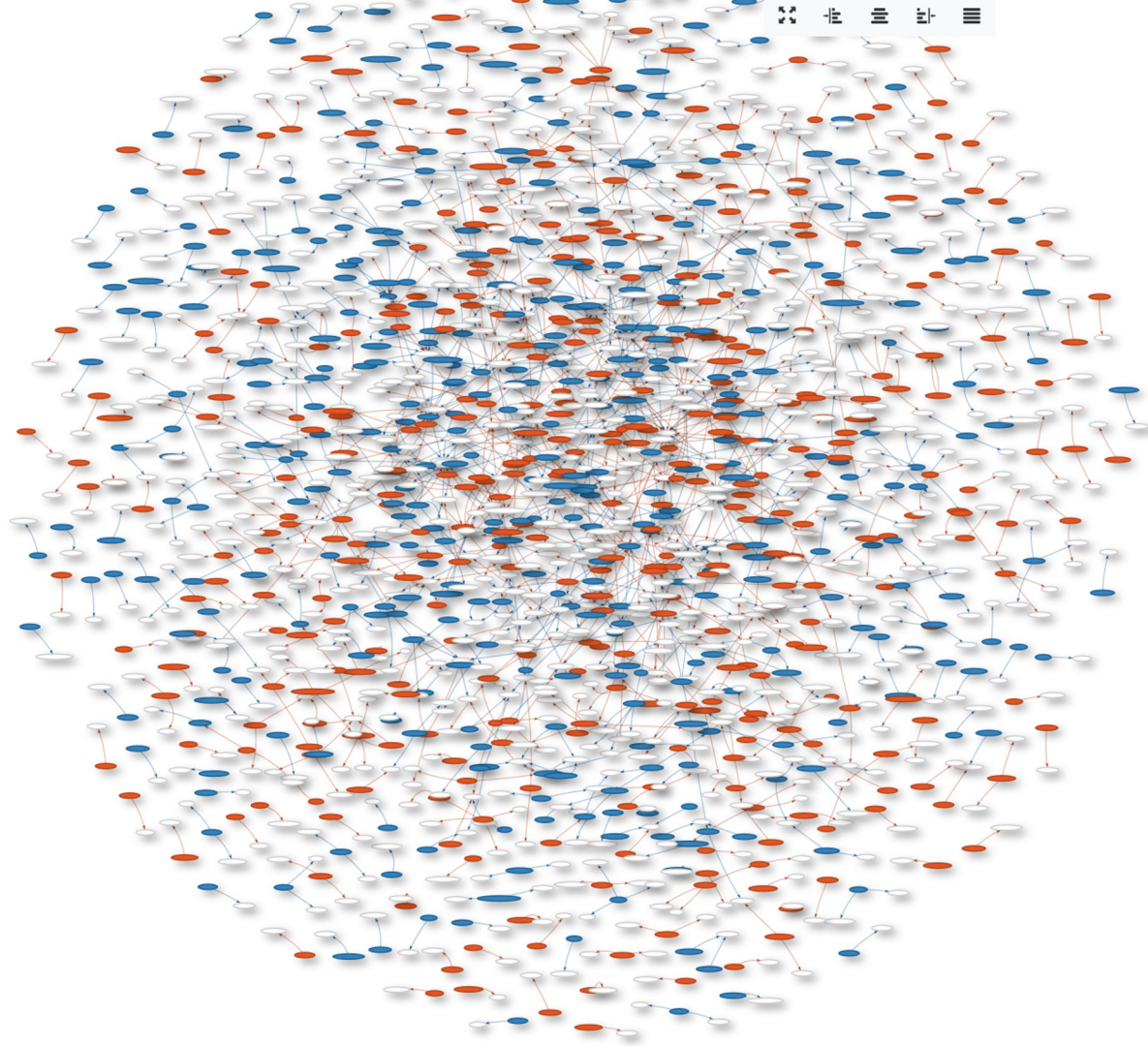
Run the example queries live at the Wikidata project page
https://www.wikidata.org/wiki/Wikidata:WikiProject_NZThesisProject

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

'Employed at' map



Doctoral advisor/student graph



COLOR CODED BY:

institutionLabel ▾

University of Auckland (1846)

Massey University (1080)

University of Canterbury (984)

University of Otago (956)

Victoria University of Wellington (587)

Auckland University of Technology (528)

University of Waikato (409)

Lincoln University (292)

United Institute of Technology (10)

Close ^

hinder the snow

Investigations on
bury soil...

X-ray analysis of an
aloid and some...

A preliminary inquiry
into the general effects...
m of the opti...

Spelling-Analysis and
Ralph Crane: a...

1 Jan 1927

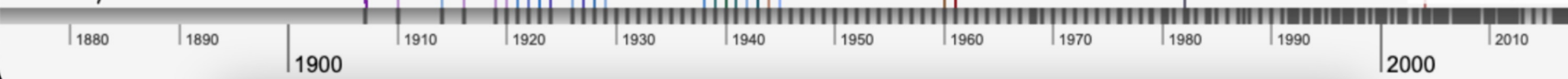
1 Jan 1960

A study on the role of
cytokinins in the...

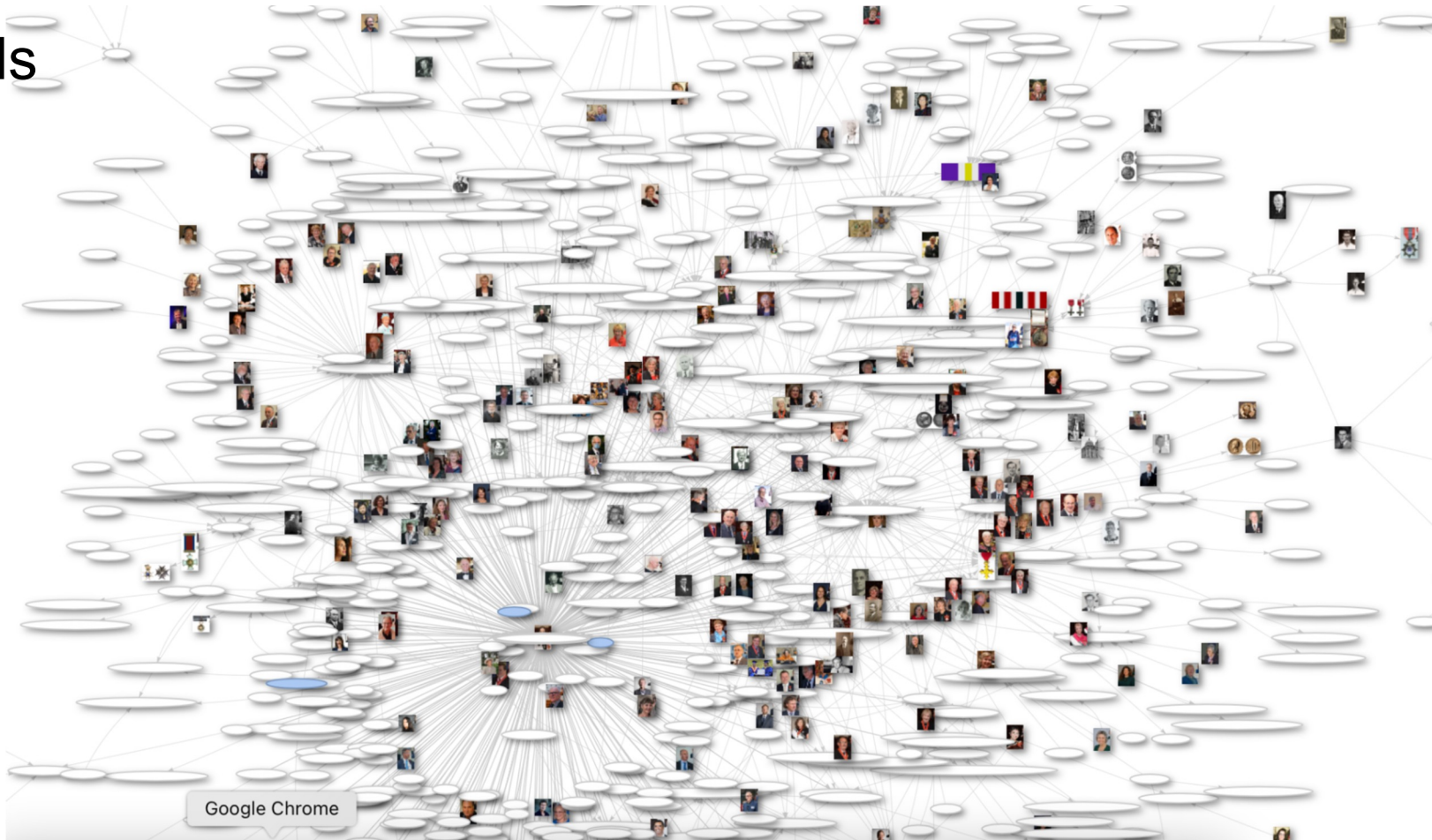
1 Jan 1982

The relationship between
objects and identity

1 Jan 2004



Awards



First country to put a national thesis dataset in Wikidata

Before (March 2022)

- 51 theses in Wikidata
- University of Canterbury had 12

Now

- >66,000 theses
- We are 9 of the top 20 institutions globally

Wikidata:WikiProject Source MetaData/Theses by institution

< Wikidata:WikiProject Source MetaData

[all entities](#) [count](#) [all groupings](#) [without grouping property](#) [below threshold \(1\)](#) [Querying about thesis \(Q1266946\)](#) [Grouping by dissertation submitted to](#)

Top groupings (Minimum 1 items)		Top Prope	
Name	Count	title (P1476)	author (P50)
University of Oxford (Q34433)	16813	99.82% (16783)	3.81% (641)
University of Auckland (Q492467)	14817	100.0% (14817)	12.44% (1843)
Massey University (Q1134606)	11870	100.0% (11870)	9.04% (1073)
University of Canterbury (Q432475)	11287	100.0% (11287)	8.7% (982)
Delft University of Technology (Q752663)	9665	96.69% (9345)	0.21% (20)
University of Leeds (Q503424)	9294	100.0% (9294)	52.74% (4902)
University of Otago (Q1201513)	8564	100.0% (8564)	11.12% (952)
Victoria University of Wellington (Q1150437)	6503	100.0% (6503)	9.03% (587)
Vanderbilt University (Q29052)	5918	99.31% (5877)	11.9% (704)
University of Washington (Q219563)	4988	99.98% (4987)	99.18% (4947)
Auckland University of Technology (Q758659)	4942	100.0% (4942)	10.66% (527)
London School of Economics and Political Science (Q174570)	4125	100.0% (4125)	100.0% (4125)
National University of La Plata (Q784171)	3825	100.0% (3825)	3.87% (148)
University of Waikato (Q606140)	3753	100.0% (3753)	10.9% (409)
Lincoln University (Q1048626)	3124	100.0% (3124)	9.35% (292)
Faculty of Applied Sciences (Q61661775)	2421	99.42% (2407)	0.17% (4)
University of North Carolina at Charlotte (Q1783178)	1445	100.0% (1445)	0.14% (2)
Faculty of Electronic Engineering, Mathematics and Computer Science (Q61660601)	1416	98.73% (1398)	0.07% (1)
Unitec Institute of Technology (Q2495129)	1270	100.0% (1270)	1.02% (13)
Faculty of Mechanical, Maritime and Materials Engineering (Q61662138)	1133	98.76% (1119)	0.09% (1)

Ongoing work: Thesis citations & Mix'n'Match

[Thesis Project Page](#)

(<https://en.wikipedia.org/wiki/Wikipedia:GLAM/NZThesisProject>)

[Google document](#)

Please have a quick read of the instructions tab before starting

Status	Author Wikidata	Author name	Wikipedia article	Thesis Wikidata link	Thesis QID	Thesis Title	Thesis description
Done	http://www.wikidata.org/entity/Q26716088	Hinematau McNeill	https://en.wikipedia.org/wiki/Hinematau_McNeill	http://www.wikidata.org/entity/Q100741873	Q100741873	Te hau ora o nga kaumatau o Tuhoā: a study of Tuhoā kaumatau mental wellness	2007 doctoral thesis by Hinematau McNeill held at Auckland University of Technology
Done	http://www.wikidata.org/entity/Q289186110	Suzanne Pitama	https://en.wikipedia.org/wiki/Suzanne_Pitama	http://www.wikidata.org/entity/Q104432926	Q104432926	'As natural as learning pathology' The design, implementation and impact of indigenous health curricula within medical school	PhD thesis by Suzanne Pitama at the University of Otago in 2013
Done	http://www.wikidata.org/entity/Q46074671	Janice M. Lord	https://en.wikipedia.org/wiki/Janice_Lord	http://www.wikidata.org/entity/Q104435598	Q104435598	The evolutionary ecology of Festuca Novae-Zelandiae in Mid-Canterbury, New Zealand	doctoral thesis of Janice Lord at the University of Canterbury 1992
Done	http://www.wikidata.org/entity/Q103976481	Joyce Watson	https://en.wikipedia.org/wiki/Joyce_Watson_(chemist)	http://www.wikidata.org/entity/Q104626012	Q104626012	A phase rule investigation of the three component system BaO - (CH3CO)2O - H2O at 25 degrees C., and 35 degrees C	Masters thesis at the University of Canterbury
Done	http://www.wikidata.org/entity/Q104639103	Manying Ip	https://en.wikipedia.org/wiki/Manying_Ip	http://www.wikidata.org/entity/Q104986710	Q104986710	From Qing reformer to twentieth-century publisher: the life and times of Zhang Yuanji 1867-1959	1983 doctoral thesis by Manying Ip held at the University of Auckland
Done	http://www.wikidata.org/entity/Q103976497	Aroha Harris	https://en.wikipedia.org/wiki/Aroha_Harris	http://www.wikidata.org/entity/Q105037716	Q105037716	Dancing with the state: Māori creative energy and policies of integration, 1945–1967	2007 doctoral thesis by Aroha Harris at University of Auckland
Done	http://www.wikidata.org/entity/Q64855606	Diane Seward	https://en.wikipedia.org/wiki/Diane_Seward	http://www.wikidata.org/entity/Q105033126	Q105033126	Some Aspects of Sedimentology of the Wanganui Basin, North Island, New Zealand	1974 doctoral thesis by Diane Seward at Victoria University of Wellington
Done	http://www.wikidata.org/entity/Q103976497	Karen Stockin	https://en.wikipedia.org/wiki/Karen_Stockin	http://www.wikidata.org/entity/Q106666293	Q106666293	The New Zealand common dolphin (Delphinus sp.): Identity, ecology and conservation	2008 doctoral thesis by Karen A. Stockin at Massey University
Done	http://www.wikidata.org/entity/Q56956693	Robert McKay	https://en.wikipedia.org/wiki/Robert_McKay_(scientist)	http://www.wikidata.org/entity/Q109161107	Q109161107	Late Cenozoic (13-0 Myr) Glacimarine Sedimentology, Facies Analysis, and Sequence Stratigraphy from the Western Ross E	doctoral thesis
Done	http://www.wikidata.org/entity/Q21390208	Michelle Kelly	https://en.wikipedia.org/wiki/Michelle_Kelly_(marine_sc	http://www.wikidata.org/entity/Q111160507	Q111160507	The order Hadromerida (Porifera: Demospongiae), taxonomy and relationships of the major families	1991 doctoral thesis by Michelle Kelly at the University of Auckland
Done	http://www.wikidata.org/entity/Q7634025	Sue Bradford	https://en.wikipedia.org/wiki/Sue_Bradford	http://www.wikidata.org/entity/Q111290623	Q111290623	A major left wing think tank in Aotearoa: An impossible dream or a call to action?	2014 doctoral thesis by Sue Bradford at Auckland University of Technology
Done	http://www.wikidata.org/entity/Q6483654	Lance McCaskill	https://en.wikipedia.org/wiki/Lance_McCaskill	http://www.wikidata.org/entity/Q111290626	Q111290626	Fertilizers in New Zealand, 1867-1929	1929 masters thesis by L.W. McCaskill at Lincoln University
Done	http://www.wikidata.org/entity/Q14954943	Torchy Atkinson	https://en.wikipedia.org/wiki/Torchy_Atkinson	http://www.wikidata.org/entity/Q111290628	Q111290628	Studies on the dieback of lacebarks, Myxosporium hoheria. n.f.sp	1932 masters thesis by John Dunstan Atkinson at Massey University
Done	http://www.wikidata.org/entity/Q266364481	Brendan Hokowhitu	https://en.wikipedia.org/wiki/Brendan_Hokowhitu	http://www.wikidata.org/entity/Q111290633	Q111290633	Te mana Māori : Te tātari i nga korero parau	2002 doctoral thesis by Brendan J. Hokowhitu at University of Otago
Done	http://www.wikidata.org/entity/Q66057567	Heather Nicholson	https://en.wikipedia.org/wiki/Heather_Nicholson_(geolo	http://www.wikidata.org/entity/Q111963118	Q111963118	Geology of Waiheke Island	1953 master's thesis by Heather Nicholson held at University of Auckland
Done	http://www.wikidata.org/entity/Q8128233	Claudia Orange	https://en.wikipedia.org/wiki/Claudia_Orange	http://www.wikidata.org/entity/Q111963122	Q111963122	A kind of equality: Labour and the Maori people 1935 - 1949	1977 master's thesis by Claudia Orange held at University of Auckland
Done	http://www.wikidata.org/entity/Q290197093	Sereana Naepi	https://en.wikipedia.org/wiki/Sereana_Naepi	http://www.wikidata.org/entity/Q111963126	Q111963126	The Voices of Tacuq. Teaching and Learning Practices in Non-Lecture Settings for Māori and Pasifika Success in the First Ye	2012 master's thesis by Sereana Naepi held at University of Auckland
Done	http://www.wikidata.org/entity/Q30070190	Philippa Black	https://en.wikipedia.org/wiki/Philippa_Black	http://www.wikidata.org/entity/Q111963144	Q111963144	igneous and metamorphic rocks from Tokatoka, Northland.	1964 master's thesis by Philippa Margaret Black held at University of Auckland
Done	http://www.wikidata.org/entity/Q64597521	Salome Tanuvasa	https://en.wikipedia.org/wiki/Salome_Tanuvasa	http://www.wikidata.org/entity/Q111963147	Q111963147	My appreciation	2014 master's thesis by Salome Tanuvasa held at University of Auckland
Done	http://www.wikidata.org/entity/Q7172687	Peter Barrett	https://en.wikipedia.org/wiki/Peter_Barrett_(geologist)	http://www.wikidata.org/entity/Q111963148	Q111963148	The Te Kuli group in the Waitomo-Te Anua area : a study of structures, sedimentation and paleogeography of calcareous sec	1962 master's thesis by Peter J. Barrett held at University of Auckland

Example of Wikipedia article and citation

Michelle Kelly (marine scientist)

From Wikipedia, the free encyclopedia


Michelle Kelly (born 1961), also known as **Michelle Kelly-Borges**, is a New Zealand scientist who specialises in sponges, their chemistry,^[1] their evolution,^{[2][3]} taxonomy, systematics, and ecology.

Early life and education [\[edit \]](#)

Born in Otago in 1961, Kelly lived in Papua New Guinea with her family from 1970 to 1980, and was educated at The Correspondence School. From 1980, she studied at the University of Auckland, and completed a Bachelor of Science degree in 1983, and a Master of Science with honours in 1987.^[4] Her masterate research, supervised by Patricia Bergquist, was an investigation of the systematics and ecology of the sponges of Motupore Island in Papua New Guinea. The title of that thesis was *Systematics and ecology of the sponges of Motupore Island, Papua New Guinea*.^{[4][5]} She then earned a PhD at the University of Auckland in 1991 under the joint supervision of Patricia and Peter Bergquist, with a thesis entitled, *The order Hadromerida (Porifera: Demospongiae), taxonomy and relationships of the major families*,^[6] and in the same year (with Patricia Bergquist) published a paper on the genus, *Tethya*.^[7]

Research career [\[edit \]](#)

From 1991 to 1992, Kelly was a post-doctoral fellow at the Florida Atlantic University Harbor Branch Oceanographic Institute, working with Shirley Pomponi, studying natural products extracted from deep-sea sponges and analysing DNA sequences.^{[4][9][9]} In 1994 she was collaborating with Patricia Bergquist at the University of Auckland.^[10] From 1993, she spent four years with the Natural History Museum, London, carrying out research on sponges in the Indo-Pacific,^{[4][11]} before returning to New Zealand and working as a senior scientist in the Department of Landscape and Plant Science at Unitec Institute of Technology in Auckland and a consultant to the Coral Reef Research Foundation in Micronesia.^[12] At some point she was affiliated with the University of Hawai'i System.^[13] By 2001,^[14] she was working at the National Institute of Water and Atmospheric Research, in Auckland, where she continues to work (as of September 2021).^{[4][15]}

6. [^] Kelly-Borges, Michelle (1991), *The order Hadromerida (Porifera: Demospongiae), taxonomy and relationships of the major families*  (PDF), ResearchSpace@Auckland, hdl ●:2292/1994 [↗](#), Wikidata ● Q111160507 ●

Mix'n'match

New Zealand thesis authors

Action ▾

author name strings from NZ Thesis Project

1 ... 5 6 7 8 9 10 11 12 13 ... 349 All types ▾

John S. Henry	author of 1974 diploma thesis at University of Canterbury titled Glentunnel Domain : landscape development study	<i>Preliminarily matched</i>
John Henry Leech [Q6238866]	Lepidopterist (1862–1900) ♂; Fellow of the Royal Entomological Society ▾	Confirm Remove [all]
Stephen J. Herbert	author of 1977 doctoral thesis at University of Canterbury titled Density and irrigation studies in <i>Lupinus albus</i> and <i>L. angustifolius</i>	<i>Preliminarily matched</i>
Herbert Stephen [Q18912114]	Barrister (1857–1932) ♂; member of Inner Temple; child of James Fitzjames Stephen and Mary Richenda Cunningham; spouse of Mary Hermione Cunningham ▾	Confirm Remove [all]
G. J. Herlihy	author of 1970 masters thesis at University of Canterbury titled The spread of lamb and mutton kill in Southland from the producer viewpoint	<i>Preliminarily matched</i>
Tim Herlihy [Q7803649]	Actor, manufacturer, screenwriter, composer, and film producer (*1966) ♂ ▾	Confirm Remove [all]
B. W. Hewitt	author of 1977 doctoral thesis at University of Auckland titled Suspensions in turbulent flows.	<i>Preliminarily matched</i>
Alex W. Hewitt [Q61809768]	Ophthalmologist ▾	Confirm Remove [all]
P. L. Hewitt	author of 1972 diploma thesis at University of Canterbury titled The influence of mor forming vegetation and time on clay mineral genesis	<i>Preliminarily matched</i>
Alex W. Hewitt [Q61809768]	Ophthalmologist ▾	Confirm Remove [all]

What next

- Deborah & Tamsin working on an OAI -> OpenRefine -> Wikidata process for more automation
- Adding open access, full text URLs & license statements
- Still matching authors, advisors, main subjects
- Events to get alumni or academics involved

Thank you to Wikimedia Aotearoa NZ for a \$5k grant to clean the data



WIKIMEDIA
AOTEAROA
NEW ZEALAND

Questions and Discussion

