

Exploring Dissertations registered with DataCite

Martin Fenner

DataCite Technical Director

<https://orcid.org/0000-0003-1419-2405>

Problem

Many DataCite member organizations register DOIs for dissertations, but this information is somewhat difficult to find in a systematic way. For example

- 127,310 DOIs for resourceType Dissertation
- 188,633 DOIs for resourceType Thesis
- 312,458 Crossref DOIs for work type dissertation

This is of course only a small subset of the total number of dissertations, which is at least 5 million.

How to improve the situation

<p>Decide on a ResourceTypeGeneral Dissertation to clearly identify all dissertations with DataCite DOIs</p>	<p>The Metadata Working Group proposed this in its last call</p>
<p>Combine the DOIs for dissertations from Crossref and DataCite into a single search index</p>	<p>The Common DOI Search task in the FREYA project has started this work</p>
<p>Use the DataCite GraphQL API for a query of (almost) all dissertations in DataCite</p>	<p>The topic of this presentation</p>

User Story

User Story

As a student at the beginning of my PhD thesis, I want to be able to find all dissertations on a given topic.

We used the example query “Shakespeare” that returned 112 dissertations.

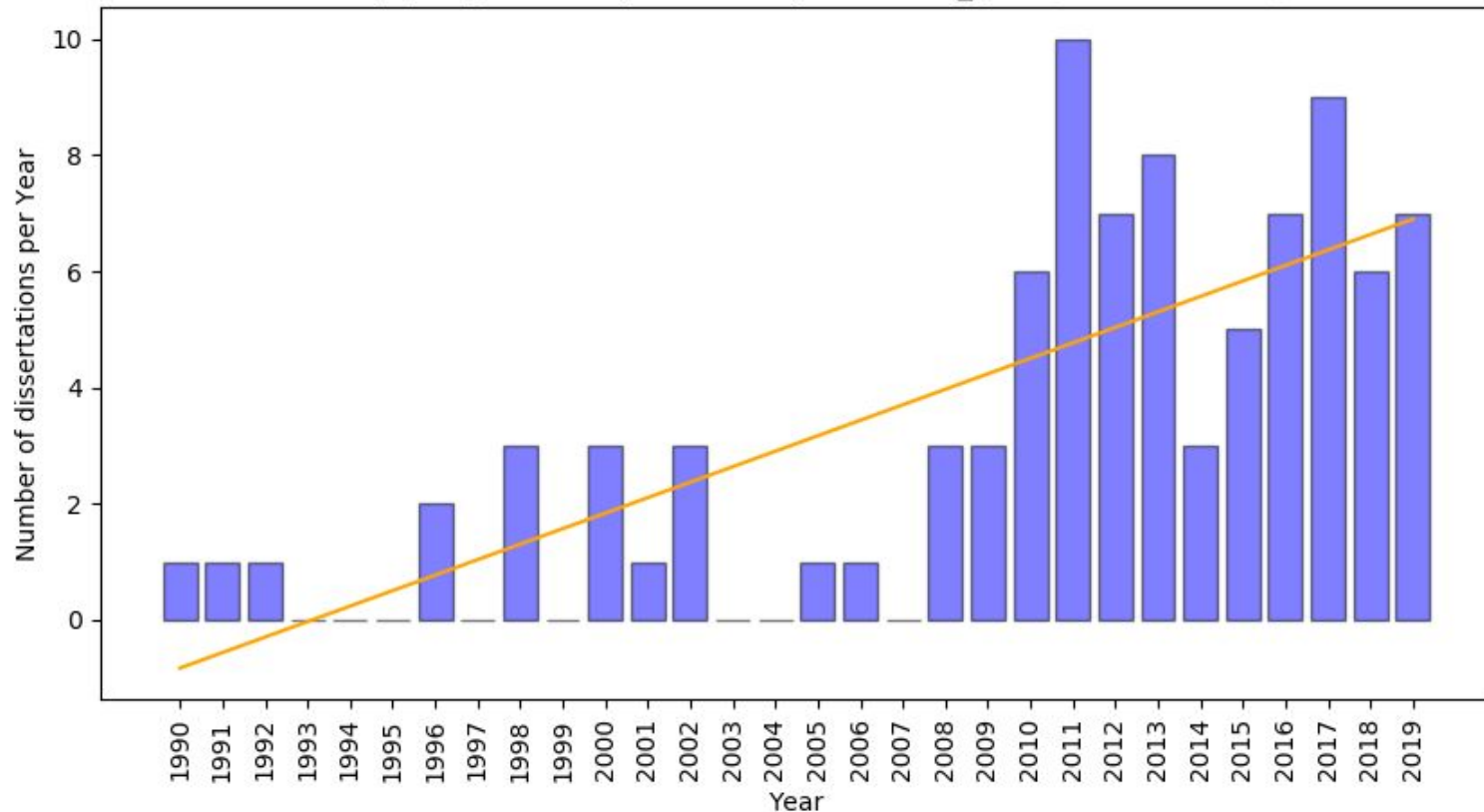
This work was done with the help of

- DataCite GraphQL API
- Jupyter notebooks
- Robert Petryszak (helping in the FREYA project)

<https://github.com/datacite/pidgraph-notebooks-python/tree/master/user-story-5-dissertations>

Number of Dissertations with Keyword “Shakespeare” by Year

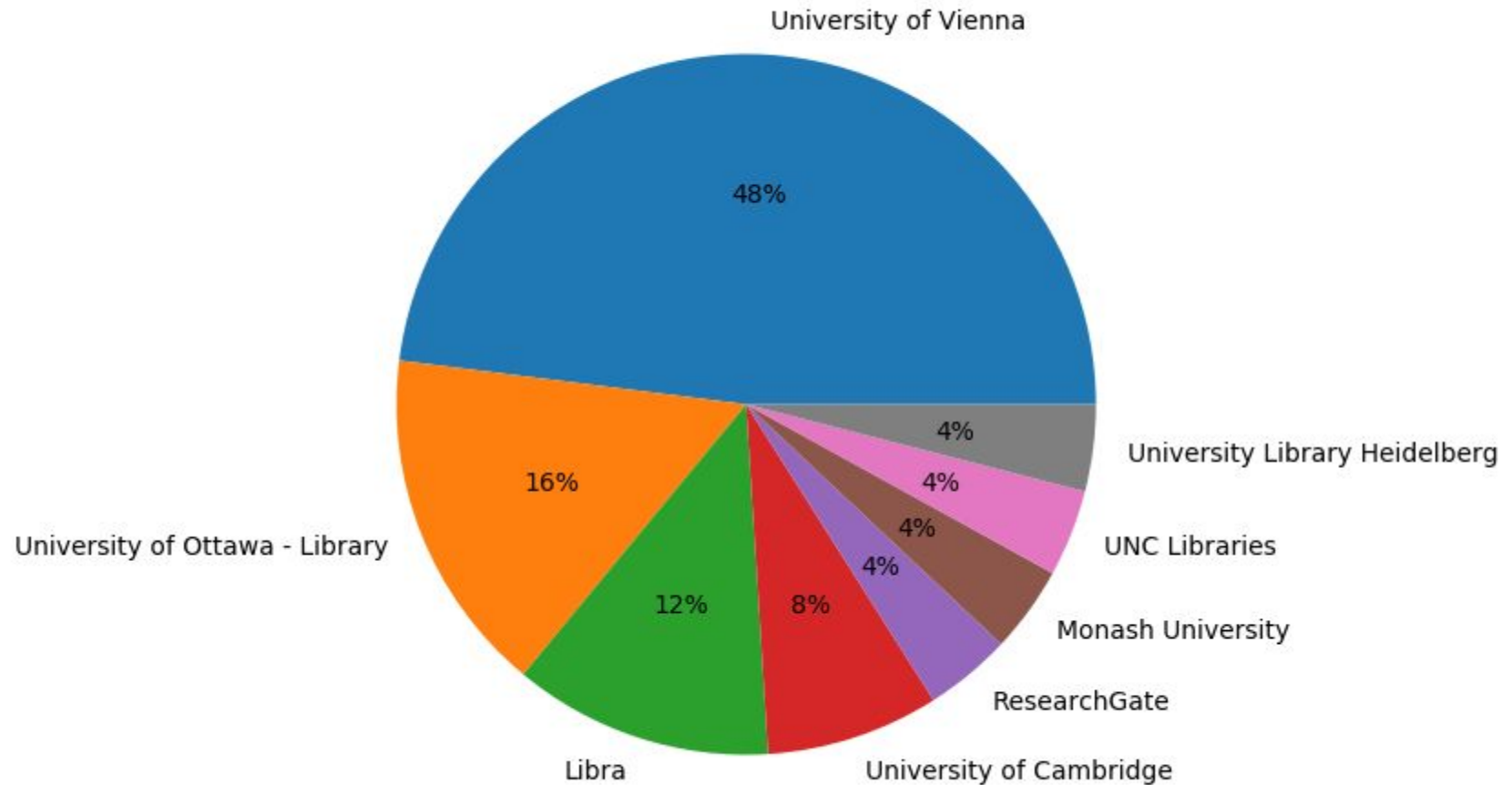
Number of dissertations found by query: 'shakespeare AND publication_year:[1990 TO 2020]' since 1990, with trend line



Word Cloud for Dissertations with Keyword “Shakespeare”



Repositories for Dissertations with Keyword “Shakespeare”



Bibliography for Dissertations with Keyword “Shakespeare”

For further study, the Jupyter notebook is generating a bibtex file containing all 112 dissertations.

The bibtex file can be imported into any reference manager and be used to get the full text of the dissertations, contact information, etc.

We can also make changes to the Jupyter notebook to collect additional information available in DataCite metadata, e.g. language, rights information, co-authors, etc.