WikiCite and Scholia
A Linked Open Data approach to exploring the scholarly literature and related resources

Daniel Mietchen, Finn Årup Nielsen, Egon Willighagen

Data Science Institute, University of Virginia, **daniel.mietchen@virginia.edu**, @EvoMRI

Cognitive Systems, DTU Compute, Technical University of Denmark, @fnielsen
Department of Bioinformatics - BiGCaT, NUTRIM, Maastricht University, @EgonWillighagen

The research ecosystem as a network of linkable data

Research takes place in a sociotechnical ecosystem that connects researchers, institutions, funders, databases, locations, publications, methodologies and related concepts with the objects of study and the world around them. Schemas for describing such concepts are growing in breadth and depth, number and popularity, as are mechanisms to persistently and uniquely identify the concepts, the schemas, their relationships or any of their components. In parallel, more and more data — and particularly metadata — are being made available under open licenses, which facilitates discoverability, reproducibility and reuse, as well as data integration.

Wikidata

Wikidata is a community-curated open knowledge base in which concepts covered in any Wikipedia — and beyond — can be described in a structured fashion as Linked Open Data that can be mapped to RDF and queried using SPARQL as well as various other means. The Wikidata community of close to 20,000 monthly contributors oversees a corpus that currently comprises about 50 million 'items', i.e. entries about concepts. These items cover a broader range of topics than Wikipedia and are annotated and linked via about 5000 'properties' that describe relationships between items or between items and external entities or that express specific values. The items and properties have persistent unique identifiers, to which labels and descriptions can be attached in about 300 natural languages. For instance, Q3919 represents the item for 'Gaborone' and Q6786626 'maternal health', while P274 stands for the property of 'chemical formula', and P225 for 'taxon name'.

WikiCite

Amongst many other topics, Wikidata also contains information about researchers and many components of their research ecosystems, including a growing body of publications and databases, particularly in the life sciences, which can be used as references in Wikidata or beyond. The curation of this reference-centric part of Wikidata is overseen by the WikiCite initiative, which extends from scholarly publications to patents, course cases, cell lines and a range of other resources that are being cited.

Scholia

A range of open-source tools is available to interact with Wikidata — to enter information, curate and query it. One of them is Scholia, a frontend to Wikidata's SPARQL endpoint. Available via **https://tools.wmflabs.org/scholia/**, it can be used to explore research publications and how they relate to authors, institutions, funders and other parts of the research ecosystem, as well as to taxa, metabolic networks, or geolocations. Scholia can thus be used as a starting point for exploring how information about scholarly research is represented in Wikidata and how it can be explored, curated and reused.

Sample visualizations
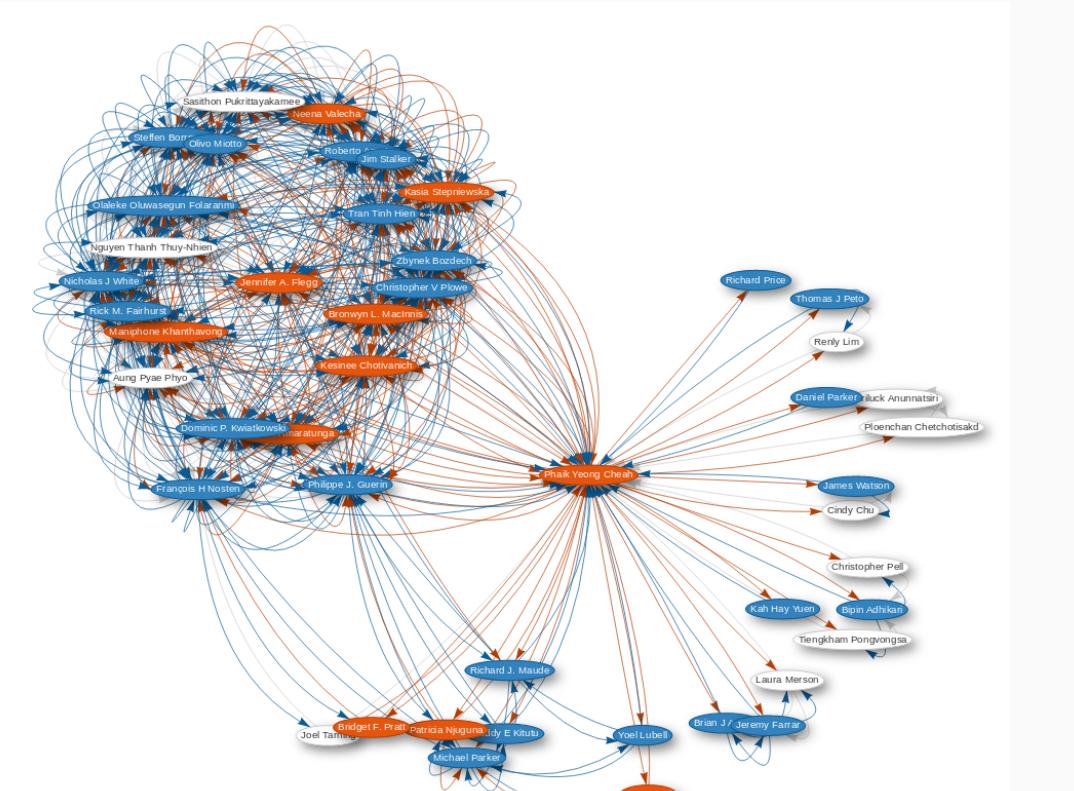
Partial co-author graph for Phaik Yeong Cheah
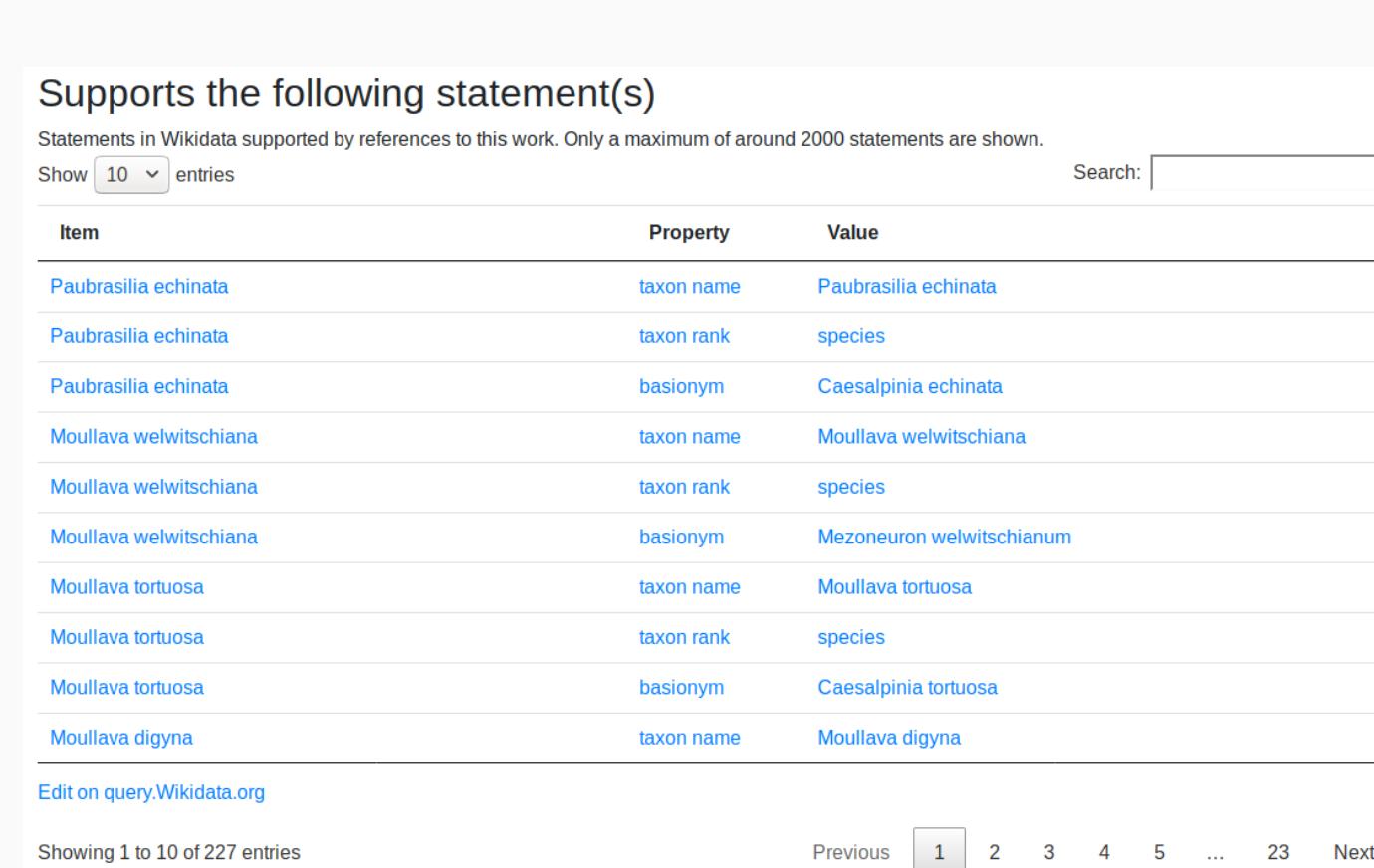
Wikidata statements supported by a given paper
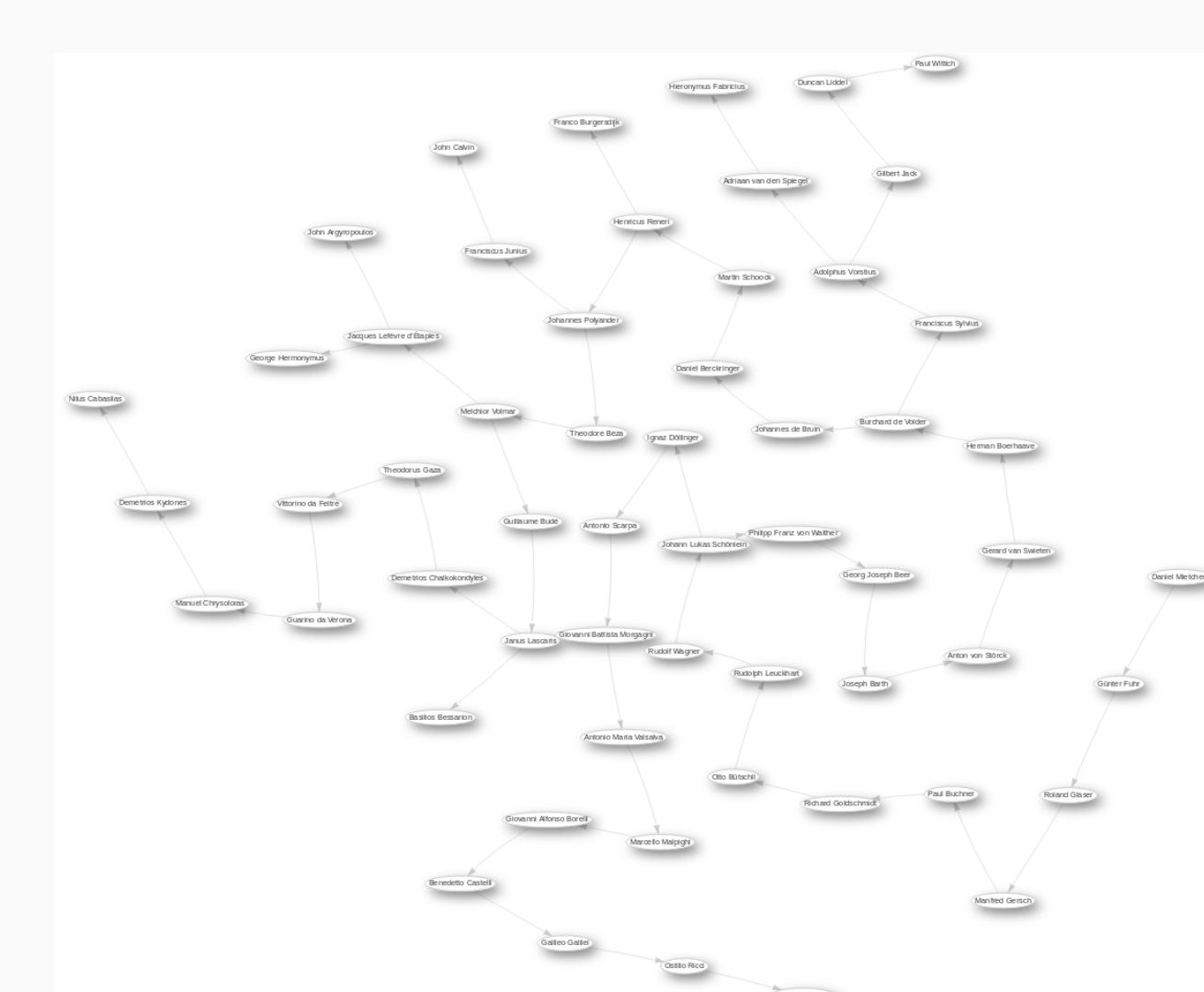
Partial academic tree for one of us

Recent publications by people affiliated with the University of Leeds
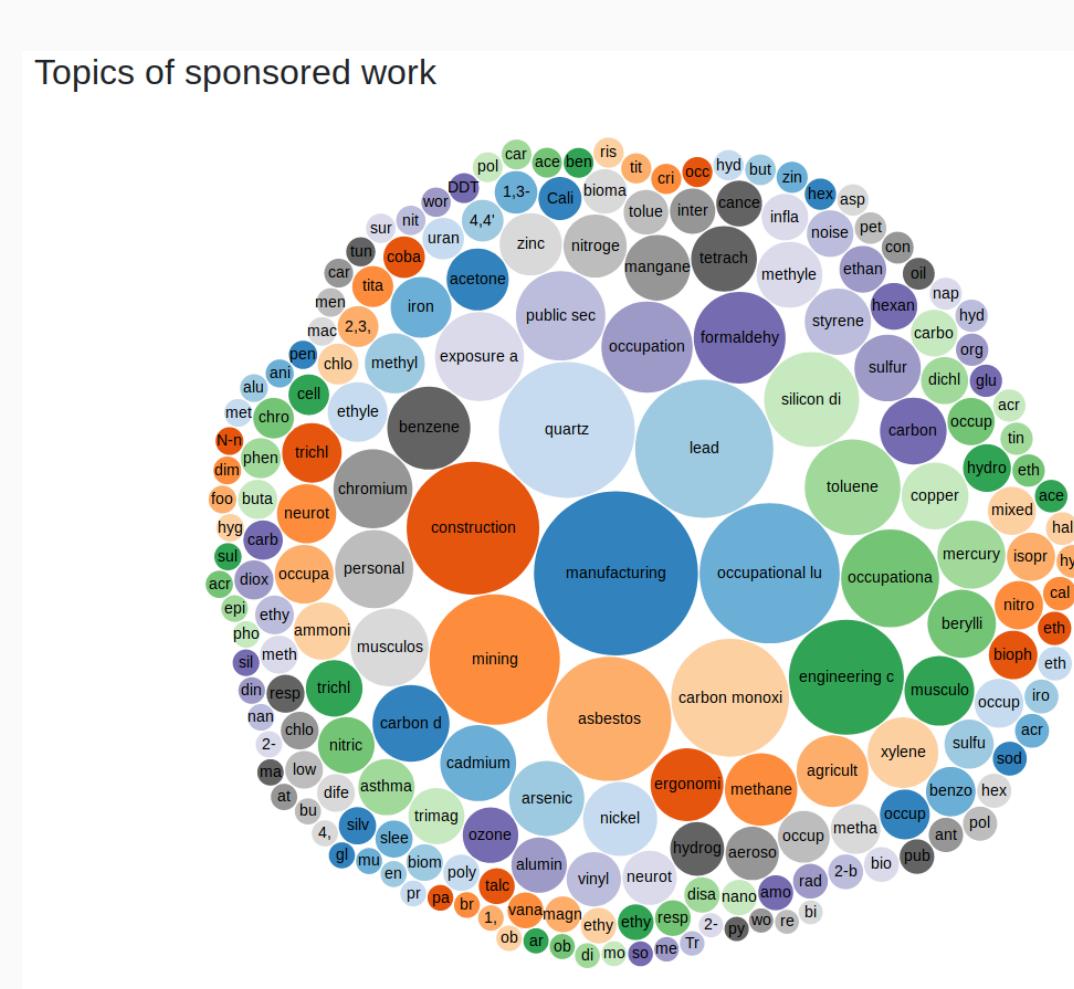
Topics funded by the U.S. National Institute of Occupational Safety and Health
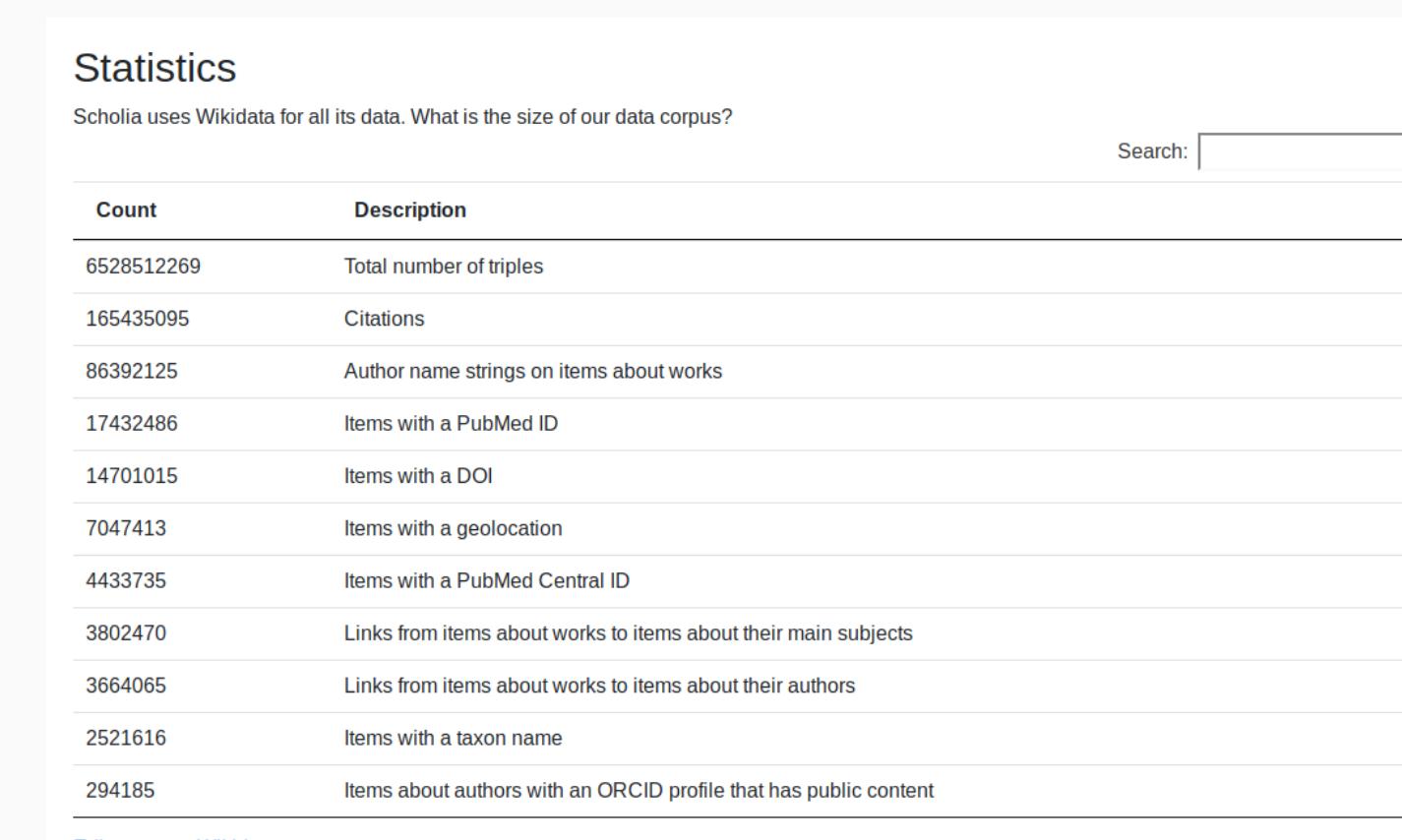
Basic statistics of WikiCite content in Wikidata

Zika corpus

Scholia's topic profile for the Zika virus.

Further reading

For publications about Scholia, see **https://tools.wmflabs.org/scholia/topic/Q45340488**, its page about itself.