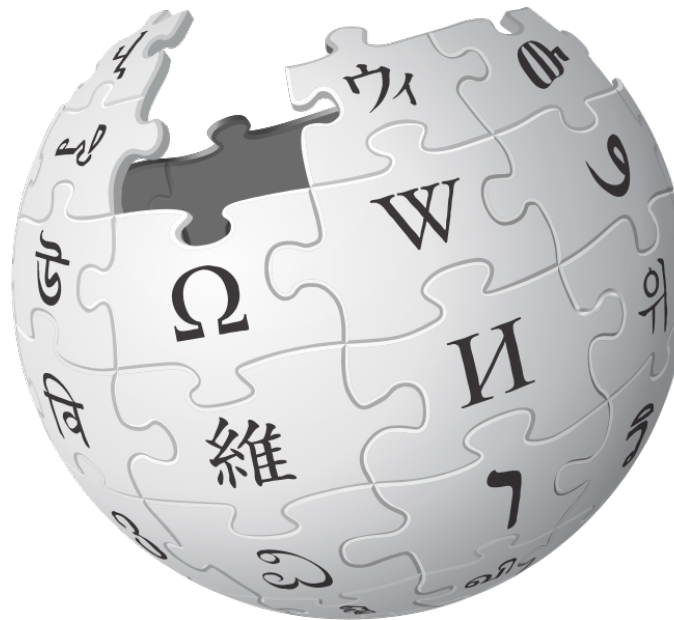


Wikipedia — a serious platform for researchers?

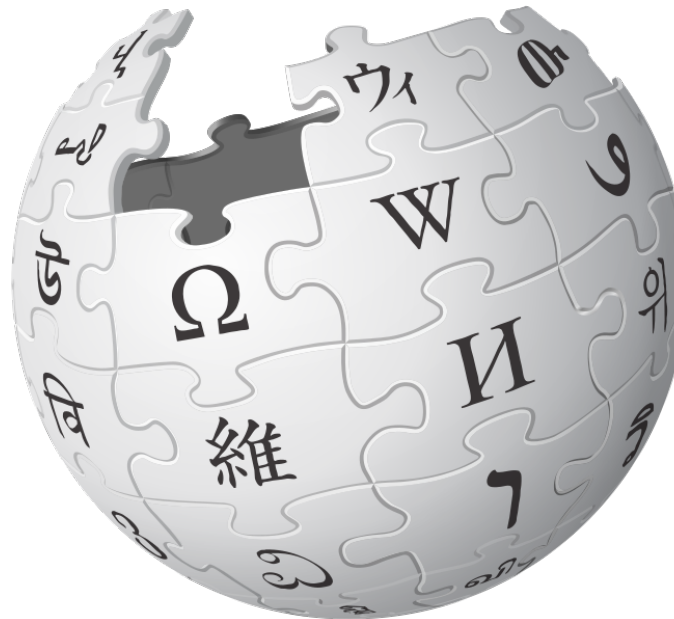
Finn Årup Nielsen

Cognitive Systems, DTU Compute, Technical University of Denmark

13 December 2018



What is this?



What is this?

Logo Nohat (concept by Paullusmagnus); Wikimedia. CC BY-SA. Trademark by Wikimedia Foundation

An online encyclopedia

Yes, you can read it like a scientific review article.

A publishing platform

And you can write like a scientific article or blog post.

A social media platform

Wikis were one of the first Web 2.0 platforms: With Wikipedia, You can login a talk and discuss with other users, usually with a more civil tone than other parts of the social media ecosystem.

A part of the free and open software community

There is a strong focus on free software use and Open licence, — in line with the Open Science movement.

Linux, Apache, PHP, Javascript, Python. Creative Commons or GPL licenses. OGG media format because of patents in MPEG.

A corpus

Used in state-of-the-art machine learning algorithms.

A project

Wikipedia is continuously evolving with people interacting.

Examples: “Lisbeth eller Lisbet Palme?” and “Digtet holder kun på 15 strofer”

An annotated search engine

Perhaps Wikipedia is not an citable encyclopedia, but an annotated list with pointers to where there real information is, e.g., in scientific articles.

Wikipedia as a corpus

Explicit semantic analysis for semantic relatedness (Gabrilovich and Markovitch, 2006) . . . and see our review (Mehdi et al., 2017).

Facebook AI Research's fastText at <https://fasttext.cc/>: “We are publishing pre-trained word vectors for 294 languages, trained on Wikipedia using fastText.” (Bojanowski et al., 2016)

Google's BERT deep learning model: “For the pre-training corpus we use the concatenation of BooksCorpus (800M words) (Zhu et al., 2015) and English Wikipedia (2,500M words).” (Devlin et al., 2018)

Danish model: “We downloaded the Danish Wikipedia XML article dump from <https://dumps.wikimedia.org/> and used the mwparserfromhell Python module to extract text from 351,186 raw article wiki-pages.” (Nielsen and Hansen, 2017)

Editing Wikipedia

Editing Biofilm

Content that violates any copyrights will be deleted. Encyclopedic content must be verifiable. Work submitted to Wikipedia can be edited, used, and redistributed—by anyone—subject to certain terms and conditions.

```
Chemistry}} |year= 2012 |volume= 84 |issue= 2 |pages= 377–410 |doi=
10.1351/PAC-REC-10-12-04 |url= http://pac.iupac.org/publications/pac/pdf
/2012/pdf/8402x0377.pdf |last1= Vert |first1= Michel |last2= Doi |first2=
Yoshiharu |last3= Hellwich |first3= Karl-Heinz |last4= Hess |first4= Michael
|last5= Hodge |first5= Philip |last6= Kubisa |first6= Przemyslaw |last7=
Rinaudo |first7= Marguerite |last8= Schué |first8= François}}</ref>
}}
A ''biofilm'' comprises any syntrophic consortium of [[microorganism]]s in
which [[cell (biology)|cells]] stick to each other and often also to a
surface.<ref name=":0">{{Cite journal|last= López|first= Daniel|last2=
Vlamakis|first2= Hera|last3= Kolter|first3= Roberto|date= 2010|title=
Biofilms|journal= Cold Spring Harbor Perspectives in Biology|volume= 2|issue=
7|page= a000398|doi= 10.1101/cshperspect.a000398|issn= 1943-0264|pmc=
2890205|pmid= 20519345|via= }}</ref><ref name=":1" /> These adherent cells
become embedded within a slimy [[extracellular matrix]] that is composed of
[[extracellular polymeric substance]]s (EPS).<ref name=":0" /><ref name=":1"
/> The cells within the biofilm produce the EPS components, which are
typically a [[polymer]]ic conglomeration of extracellular
[[polysaccharide]]s, [[protein]]s, [[lipid]]s and [[DNA]].<ref name=":0"
/><ref name=":1" /><ref name=":5">{{cite journal |last1= Aggarwal |first1=
S|last2= Stewart |first2= P|last3= Hozalski |first3= R |date= January 2016
|title= Biofilm Cohesive Strength as a Basis for Biofilm Recalcitrance: Are
Bacterial Biofilms Overdesigned?url= http://journals.sagepub.com/doi/abs
/10.4137/MBI.S31444?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&
rfr_dat=cr_pub%3dpubmed |journal= Microbiology Insights|volume= 8
|issue=Sun1 2 |pages= 29–32 |doi= 10.4137/MRT.S31444 |access-date= June 29
```

Edit summary (Briefly describe your changes)

Create an account: Otherwise your IP address will be recorded. You get a private watchlist, a user page where you can present yourself, a discussion page where people can contact you, a editing history.

Choice: Raw edit and visual editor.

Begin from sources and use citations.

Stigmergy!

... but I cannot write ...

Wikimedia Commons



Ascaris male 200x section by Massimo brizzi. CC BY-SA 4.0. Photo from **Wiki Science Competition**

You can contribute to the media archive for Wikimedia wikis and others at <https://commons.wikimedia.org>

Over 50 million files in various formats: Images (photos, plots, illustrations, icons, ...), video, audio, 3D, data files.

Media files must be Creative Commons BY-SA or similar: If you use them remember to attribute author and license!

Wikimedia Commons: Photos



Figure 4 from [Evidence of Authentic DNA from Danish Viking Age Skeletons Untouched by Humans for 1,000 Years](#). Melchior et al. (Melchior et al., 2008). CC-BY. Used in the Danish Wikipedia article [Galgedil](#).

Wikimedia Commons: Video

File:Elephant Koshik vocalizing - 126327009.ogv

From Wikimedia Commons, the free media repository

Download Use this file Use this file Email a link Info



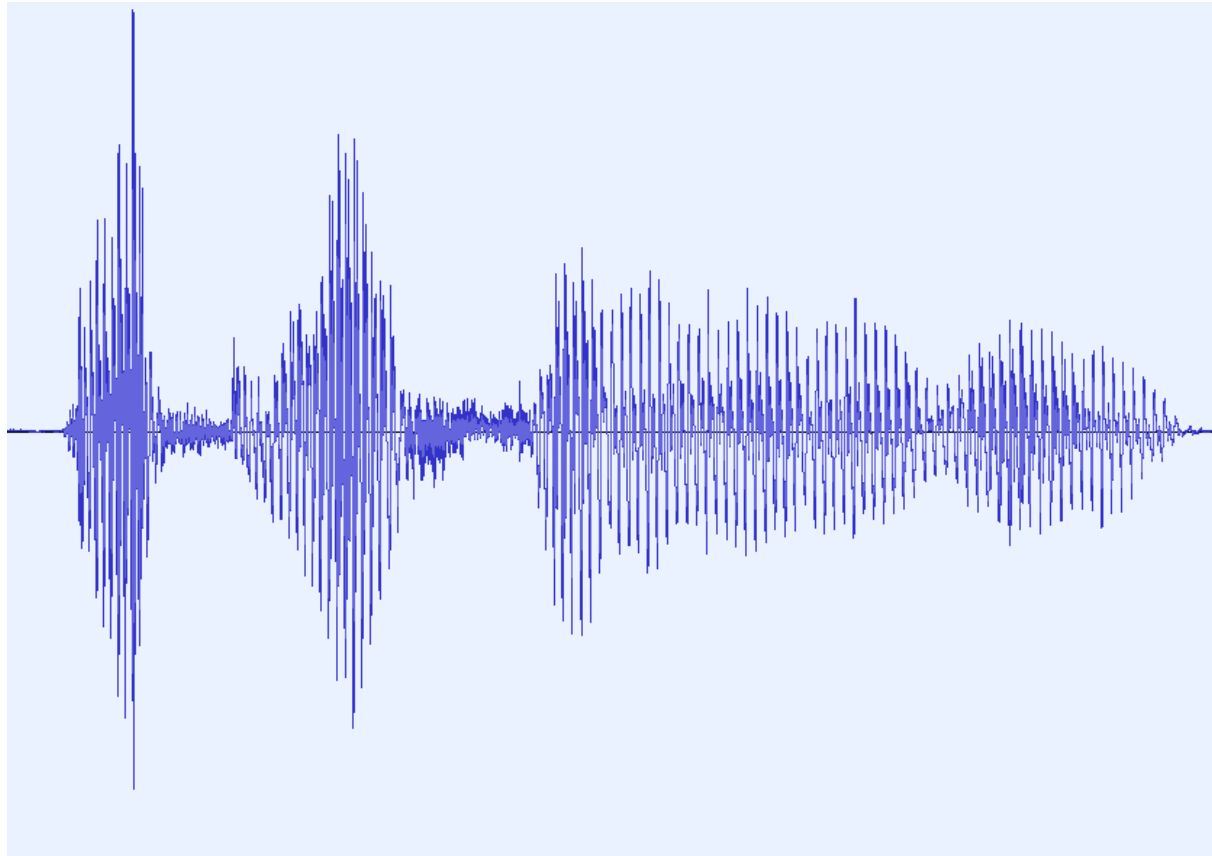
No higher resolution available. [Detect labels](#)

[Elephant_Koshik_vocalizing_-_126327009.ogv](#) (Ogg multiplexed audio/video file, Theora/Vorbis, length 12 s, 640 × 480 pixels, 2.45 Mbps overall)

Example: The Korean speaking elephant **Kosik**: https://commons.wikimedia.org/wiki/File:Elephant_Koshik_vocalizing_-_126327009.ogv

From *An Asian Elephant Imitates Human Speech*, Current Biology, 2012 ([Stoeger et al., 2012](#)).

Wikimedia Commons: Audio



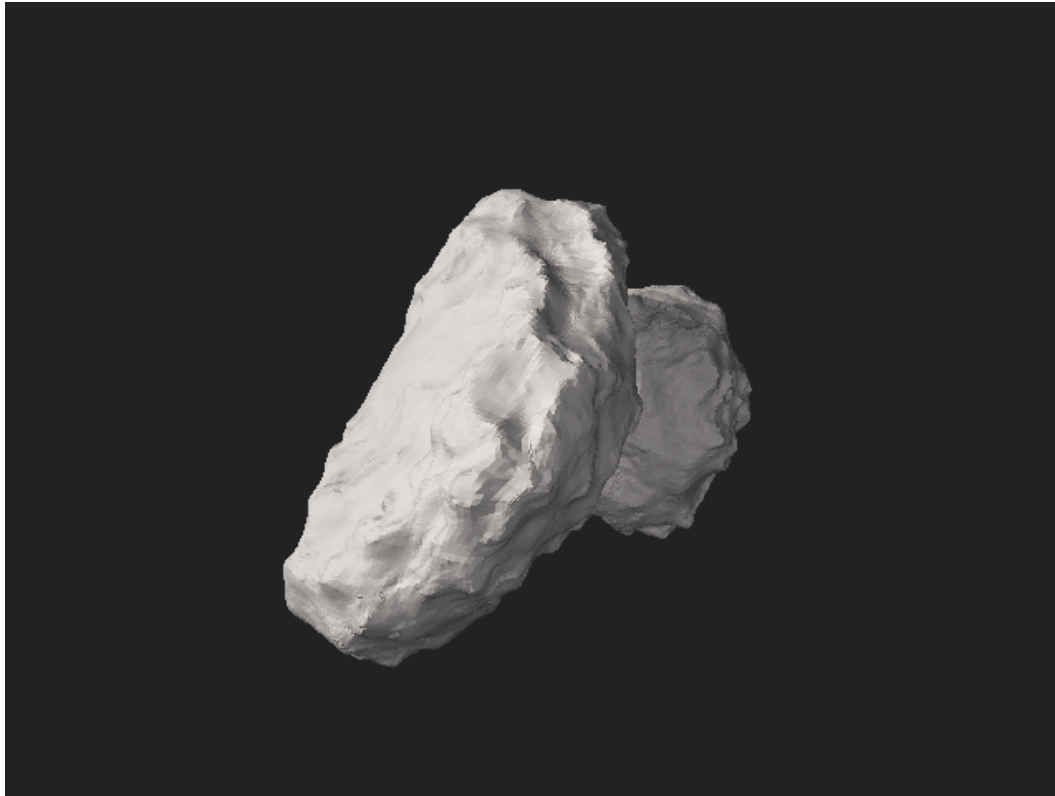
Ethnologisches Museum Berlin: [I C 1479 b x](#). Sound from an ethnographic artifact.

[Ukrainian Art Song Project](#)

Audio files with speech and pronunciations, e.g., [“Abbruchgenehmigung”](#).

Screenshot of time series of [De-Abbruchgenehmigung.ogg](#) by [jeuwe](#) CC BY-SA

Wikimedia Commons: 3D



3D files in the STL format, see examples see the category https://commons.wikimedia.org/wiki/Category:STL_files_by_source

Sculpture example from Statens Museum for Kunst: [Diskoskasteren](#)

There are science files from ESA and NASA, for instance, [67P-Churyumov-Gerasimenko.stl](#).

Wikimedia Commons: map and table data

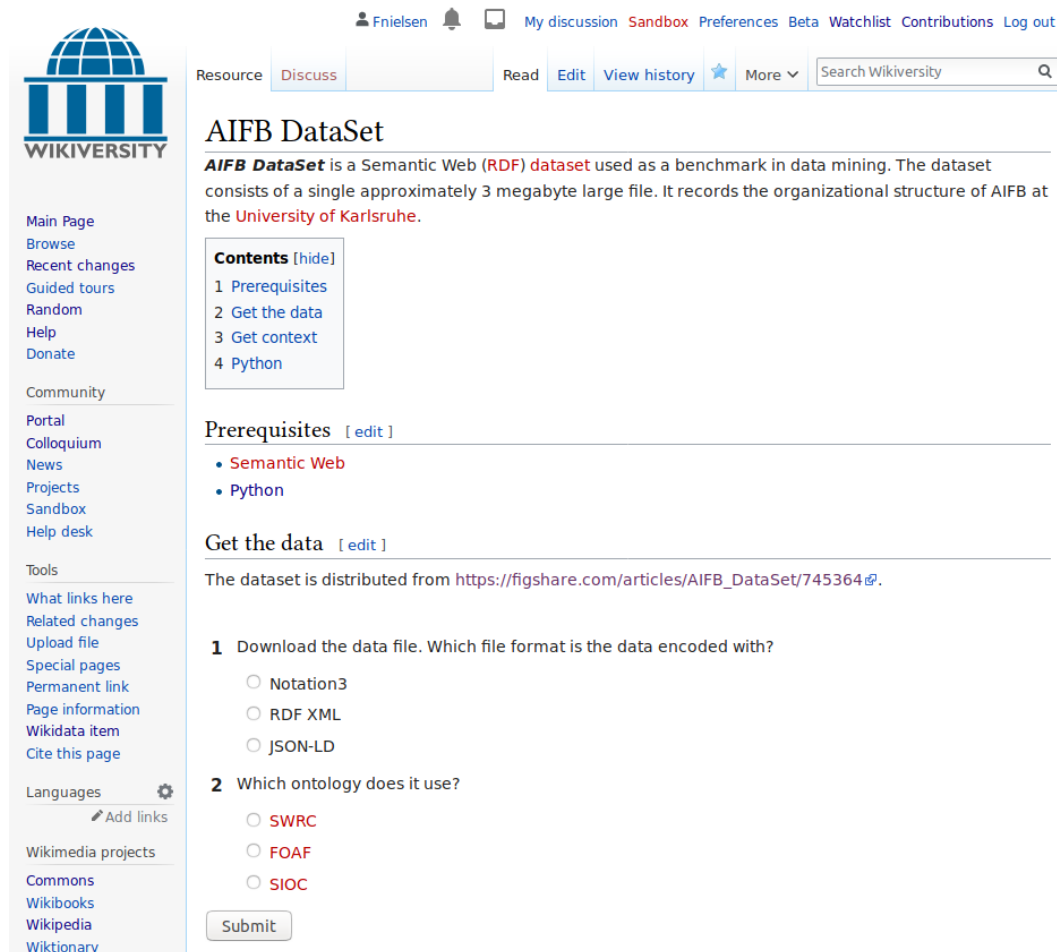


Specify geographic shape.
Example: **Manhattan**.

Tabular data, e.g., for
weather history and popu-
lation size. Example: **New
York weather history**

Map by OpenStreetMap contributors.

Wikiversity



Resource **Discuss** Read **Edit** View history More ▾ Search Wikiversity

AIFB DataSet

AIFB DataSet is a Semantic Web (RDF) dataset used as a benchmark in data mining. The dataset consists of a single approximately 3 megabyte large file. It records the organizational structure of AIFB at the **University of Karlsruhe**.

Contents [hide]

- 1 Prerequisites
- 2 Get the data
- 3 Get context
- 4 Python

Prerequisites [edit]

- **Semantic Web**
- Python

Get the data [edit]

The dataset is distributed from https://figshare.com/articles/AIFB_DataSet/745364.

1 Download the data file. Which file format is the data encoded with?

Notation3

RDF XML

JSON-LD

2 Which ontology does it use?

SWRC

FOAF

SIOC

Submit

Wikiversity at is presumably the least visible “big” Wikimedia wiki . . . and still trying to define itself in terms of scope and style.

“ . . . project devoted to learning resources, learning projects, and research for use in all levels, types . . . ”

Quizzes possible to make, but the responses are not collected.

Example: AIFB DataSet: https://en.wikiversity.org/wiki/AIFB_DataSet.

...but as a serious researcher I do not want to contribute, because it is difficult to get scholarly credit, the text I write is not citable and people might revert what I have written ...

Parallel publishing

The screenshot shows two overlapping web pages. The top page is a journal article titled "Approximate Bayesian Computation" by Mikael Sunnåker et al., published in January 2013. It has 39,208 views and 34 shares. The bottom page is the corresponding Wikipedia article for "Approximate Bayesian computation", which has 241 revisions since 2007, 86 editors, 34 watchers, and 4,388 page statistics. The Wikipedia article text states: "Approximate Bayesian computation (ABC) constitutes a class of computational methods rooted in Bayesian statistics that can be used to estimate the posterior distributions of model parameters. In all models of statistical inference, the likelihood function is of central importance, and thus quantifies the support data lend to particular values of the model parameters and to choices among different models. For simple models, an analytical formula for the likelihood function can typically be derived. However, for more complex models, an analytical formula might be difficult to derive, and the likelihood function might be computationally very costly to evaluate. ABC methods bypass the evaluation of the likelihood function. In this sense, ABC methods widen the realm of models for which statistical inference can be considered. ABC methods are mathematically well-founded, but they inevitably make assumptions and approximations whose impact need to be carefully assessed. Furthermore, the wider application domain of ABC exacerbates the challenges of parameter estimation and model selection. ABC has rapidly gained popularity over the last years and in particular in the analysis of complex problems arising in biological sciences, e.g. in population genetics, ecology, epidemiology, and systems biology."

Journal(s?) exist that allow authors to write peer-reviewed articles for inclusion in both the journal and Wikipedia.

Example: PLOS Computational Biology: “Topic pages” (Mietchen et al., 2018)

Here the English Wikipedia article **Approximate Bayesian computation** vs the originally published (Sunnåker et al., 2013).

WikiJournals



WikiJournal of Medicine, 2014, 1 (2)
doi: 10.15347/wjwm/2014.012
Review Article

WikiJournal of Medicine, WikiJournal of Science, WikiJournal of Humanities

Insights into abdominal pregnancy

Gwinyai Masukume

Editor's note

This article provided a great deal of valuable evidence that was not mentioned in the Wikipedia article on abdominal pregnancy, and the Wikipedia article has subsequently been expanded with text from this publication. However, because of this purpose, it has never been the aim of this article in itself to be a complete review of the subject, and many aspects of abdominal pregnancy are not included herein.

This article also provides an example of how to contribute to Wikimedia projects such as Wikipedia by means of academic publishing.

Introduction

While rare, abdominal pregnancies have a higher chance of maternal mortality, perinatal mortality and morbidity compared to normal and ectopic pregnancies, but on occasion a healthy viable infant can be delivered.^[1]

Because tubal, ovarian and broad ligament pregnancies are as difficult to diagnose and treat as abdominal pregnancies, their exclusion from the most common definition of abdominal pregnancy has been debated.^[2]

Others - in the minority - are of the view that abdominal pregnancy should be defined by a placenta implanted into the peritoneum.^[3]

Symptoms and signs

Abdominal pregnancy does not have any specific symptoms and signs so much so that in about half of instances it is missed, only being discovered during surgery; because of the "vague" yet serious nature of the symptoms, signs and results of medical tests patients

Risk factors

Risk factors are similar to tubal pregnancy with sexually transmitted disease playing a major role.^[7] However, about half of those with ectopic pregnancy have no known risk factors - known risk factors include damage to the Fallopian tubes from previous surgery or from previous ectopic pregnancy and tobacco smoking.^[8]

Mechanism

Typically an abdominal pregnancy is a secondary implantation which means that it originated from a tubal (less common an ovarian) pregnancy and re-implanted.^[9] Other mechanisms for secondary abdominal pregnancy include uterine rupture, rupture of a uterine rudimentary horn and fimbrial abortion.^[13]

Diagnosis

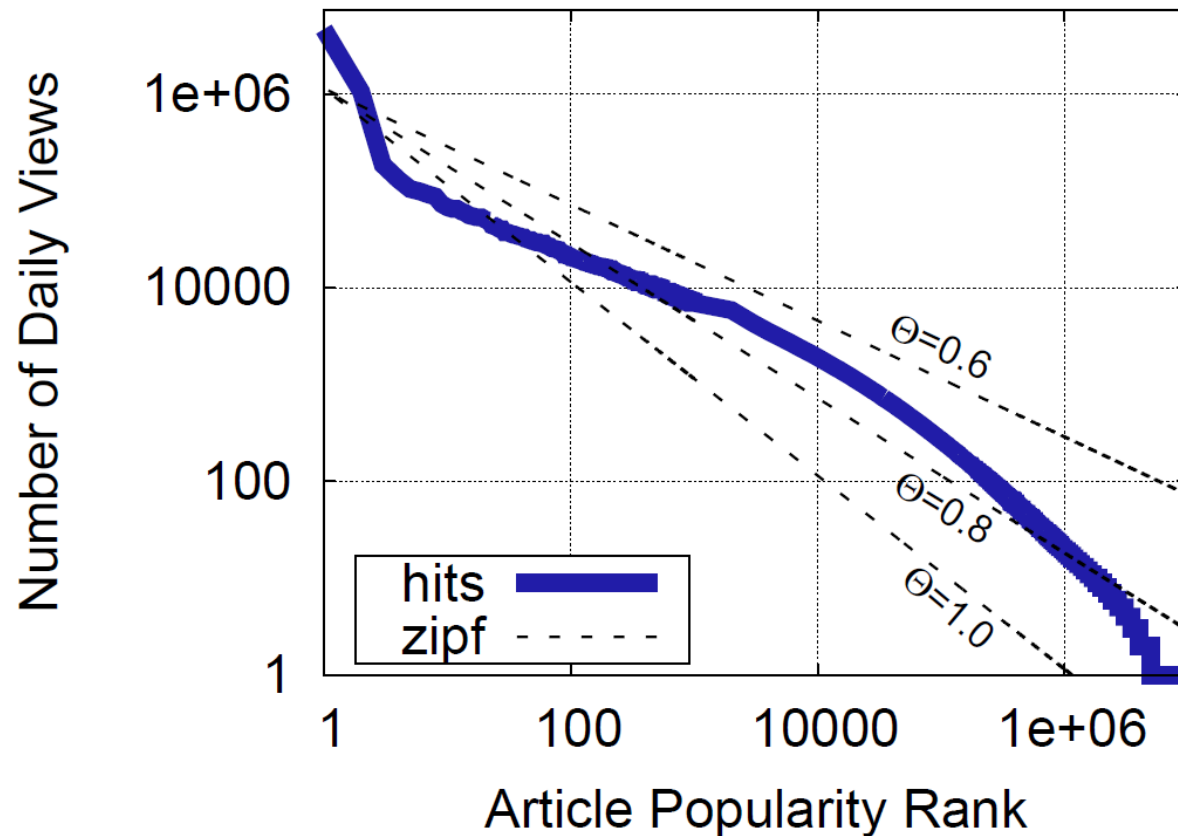
Suspicion of an abdominal pregnancy is raised when the baby's parts can be easily felt, or the lie is abnormal, the

OpenAccess journals with no cost for reader nor author and open peer-review.

Examples: Insights into abdominal pregnancy (Masukume, 2014).

Interesting, but also somewhat experimental.

Page views



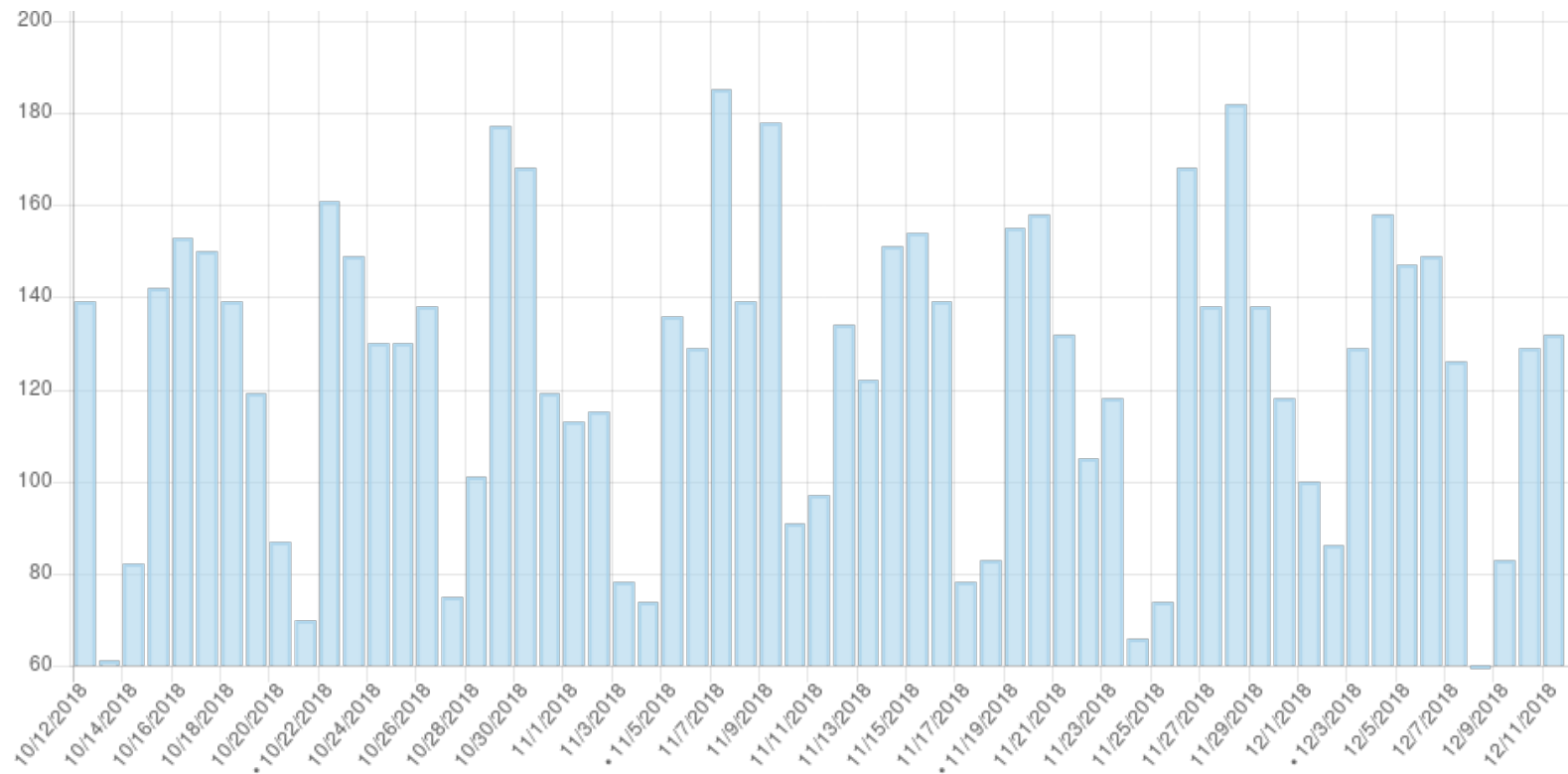
Wikipedias are among the most viewed sites in the world.

Distribution among pages highly skewed: Do not expect your article about a special topic to be viewed much.

Statistics is available as aggregate: <https://stats.wikimedia.org/>.

Wikipedia view distribution by article rank by Andrew G. West. GPL 1.2. Figure 5 from (West et al., 2011).

Page views



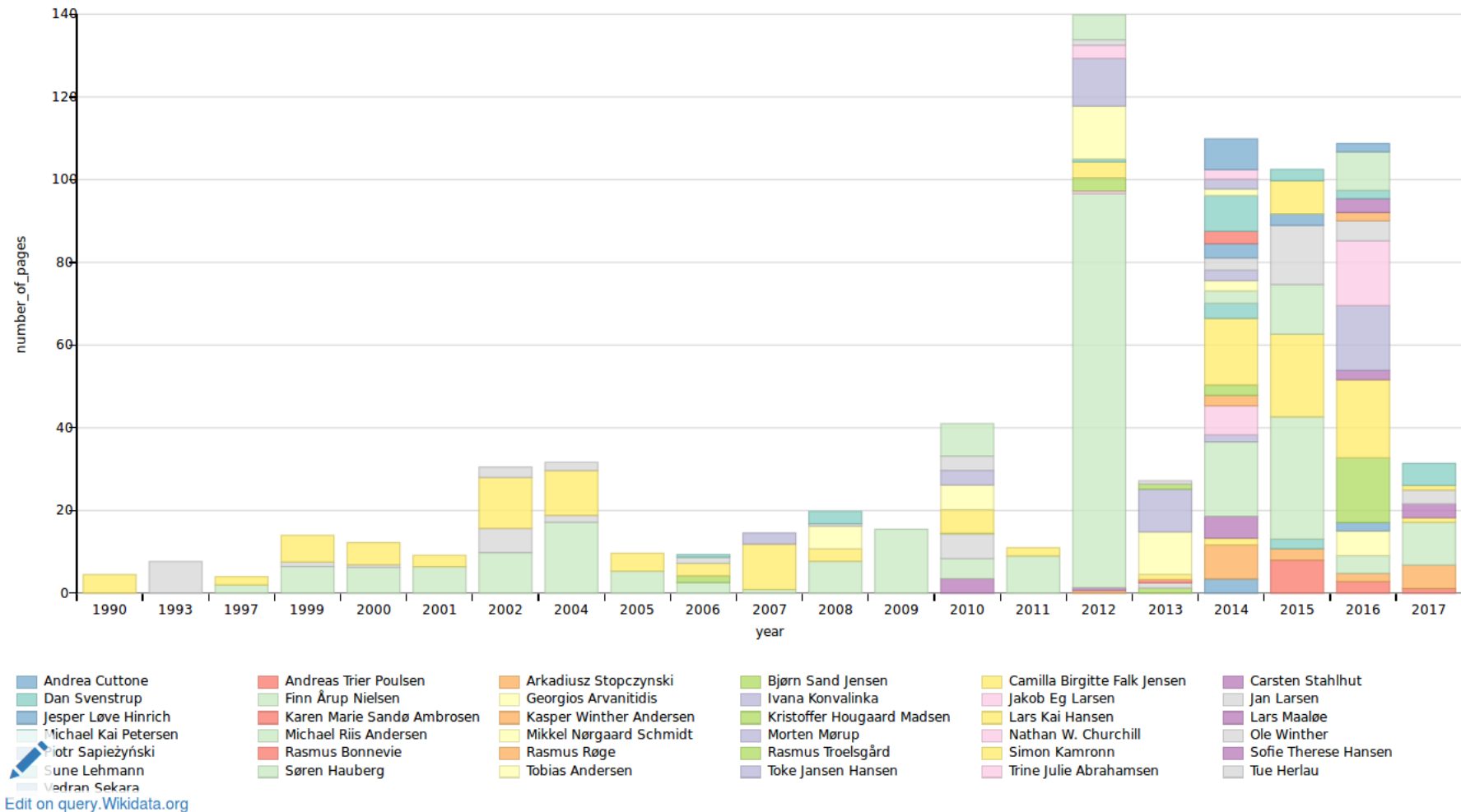
Individual article page views: <https://tools.wmflabs.org/pageviews/>, e.g., here for **Ratio distribution** on the English Wikipedia: **124 daily average**.

Paa Memphis Station: 27; 5-HTTLPR: 87

Scholia

Page production

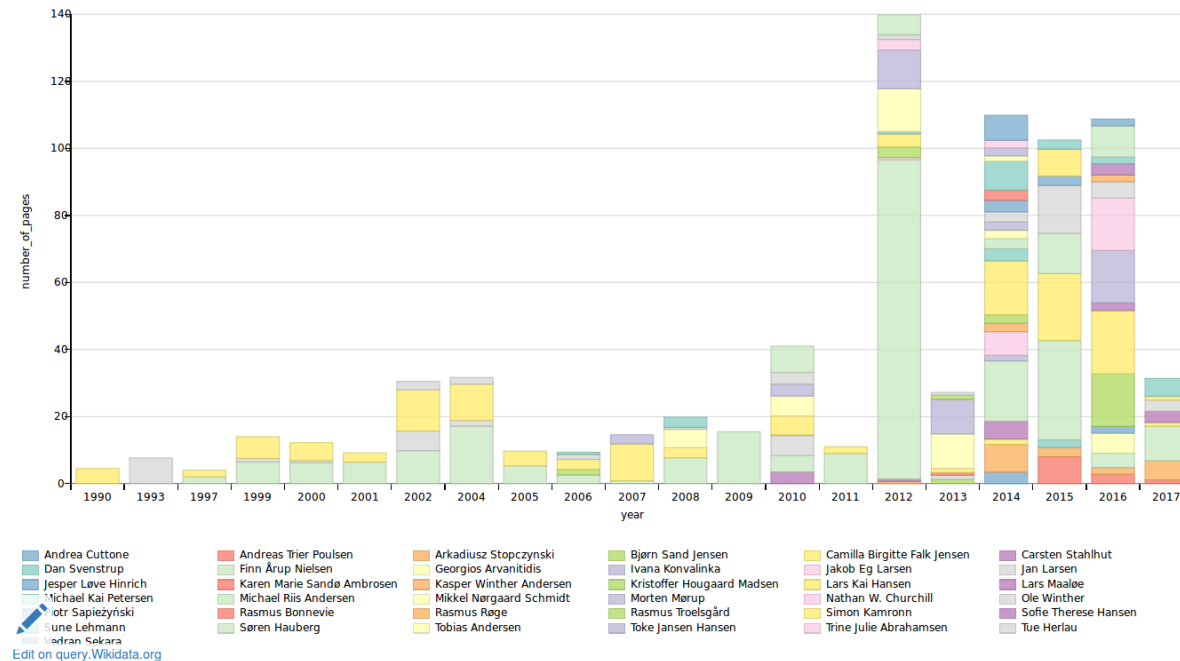
Scientific article page production per year per author. The number of pages for a multiple-author paper is distributed among the authors. The statistics is only for papers where the "number of pages" property has been set.



Scholia

Page production

Scientific article page production per year per author. The number of pages for a multiple-author paper is distributed among the authors. The statistics is only for papers where the "number of pages" property has been set.



Scholia is a webservice from <https://tools.wmflabs.org/scholia/> and a Python package from <https://github.com/fnielsen/scholia>.

The webservice generates overview of science with *Wikidata Query Service* and is built with the Flask web framework, HTML, Bootstrap, Javascript and templated SPARQL.

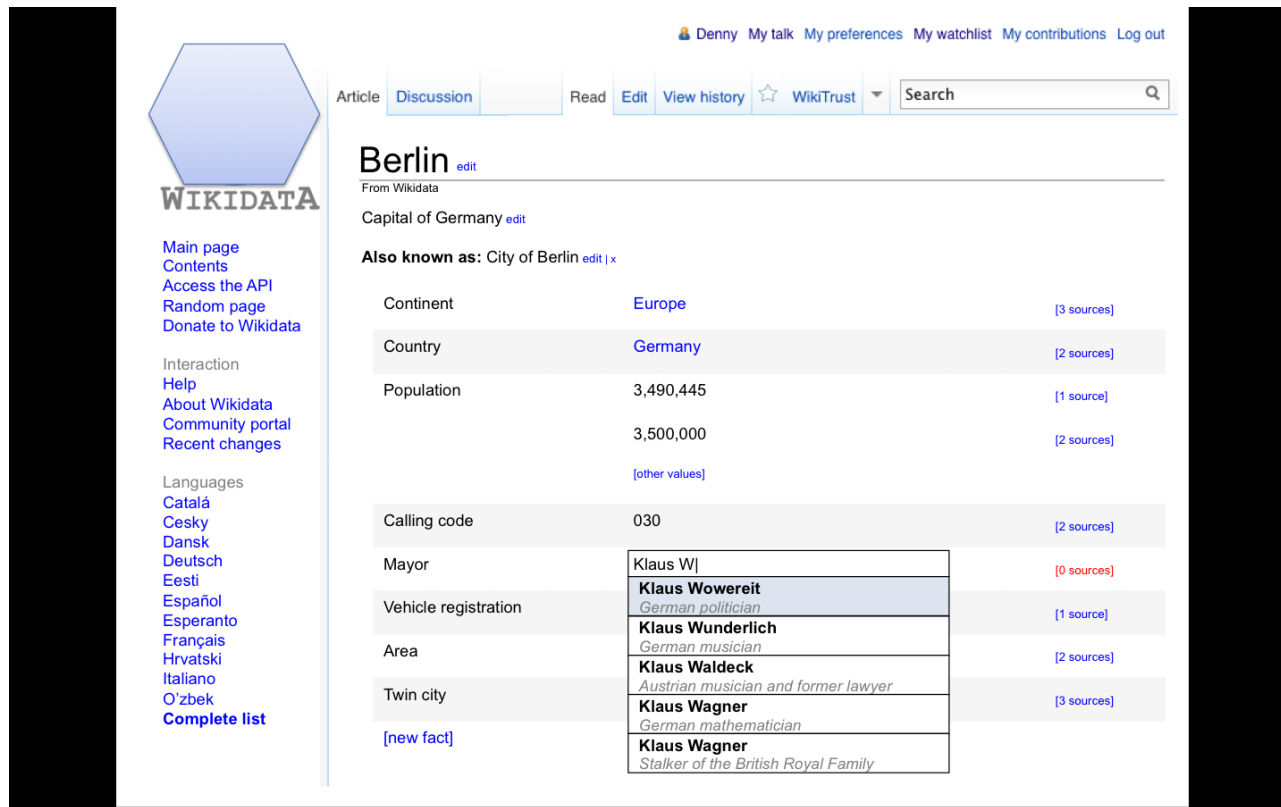
For researcher profiles, scientometrics, bibliographic reference management, information discovery (find relevant papers, scientific meetings, researchers, funding opportunities, ...).

Where does the data comes from?



WIKIDATA

Wikidata



“Wikidata: Verifiable, Linked Open Knowledge That Anyone Can edit” (Dario Taraborelli)

CC0-licensed data available on website, API, SPARQL endpoint or dump files.

Each page is an “item” with labels, aliases, properties and property values, as well as Wikipedia links.

Wikidata site UI mockup from 2012 for **Berlin (Q64)**.

Wikidata Query Service

The screenshot shows the Wikidata Query Service interface. At the top, there are navigation tabs for 'Examples', 'Help', and 'Tools', along with a language selector set to 'English'. A 'Query Helper' panel is open on the left. The main area displays a SPARQL query:

```

1 #defaultView:BubbleChart
2 SELECT ?count ?venue (SAMPLE(?venue_label_) AS ?venue_label)
3 WITH {
4   SELECT (COUNT(?work) as ?count) ?venue WHERE {
5     ?work wdt:P50 wd:Q18618629 .
6     ?work wdt:P1433 ?venue .
7   }
8   GROUP BY ?venue
9 } AS %counts
10 WHERE {
11   INCLUDE %counts
12   ?venue rdfs:label ?long_venue_label FILTER(LANG(?long_venue_label) = 'en')
13   OPTIONAL { ?venue wdt:P1813 ?short_name . }
14   BIND(COALESCE(?short_name, ?long_venue_label) AS ?venue_label_)
15 }
16 GROUP BY ?venue ?count
17 ORDER BY DESC(?count)

```

Below the query, the results are displayed as a bubble chart. The chart shows seven bubbles of varying sizes and colors, representing different venues and their counts. The bubbles are labeled as follows:

- ISWC 2006 (purple)
- WWW'06 (dark purple)
- ISWC 2014 (light green)
- SSS (light blue)
- WWW '12 (medium blue)
- WWW '16 (green)
- Hypertext (orange)

The chart is titled 'Bubble chart' and shows '14 results in 54 ms'. There are also options for 'Code', 'Download', and 'Link'.

Wikidata Query Service (WDQS) is the SPARQL endpoint for the RDF-transformed data in Wikidata: <https://query.wikidata.org/>

There is a “Query Helper” for non-programmatic formation of SPARQL queries, predefined prefixes, identifier lookup.

Several results output formats: table, bubble chart, line chart, graphs, etc.

WikiCite



Bay Area WikiSalon Feb 2017 by Pax Ahimsa Gethen. CC BY-SA 4.0

“WikiCite: Building the sum of all human citations” (Dario Taraborelli)

Use Wikidata to hold metadata about works (scientific articles, book, etc.)

Properties: authors, publication date, where it is published, reviewed by, editor, main subject, language, retracted by, erratum, volume, issue number, page range, number of pages, type or genre (retraction notice, retracted paper), series, publisher, and a lot of identifiers: DOI, ACM, Semantic Scholar, PMCID, PMID, arXiv, etc.

WikiCite Statistics

Count	Description
6110672735	Total number of triples
121065663	Citations
77862349	Author name strings on items about works
17160242	Items with a PubMed ID
13835584	Items with a DOI
6889517	Items with a geolocation
4390875	Items with a PubMed Central ID
3516037	Links from items about works to items about their main subjects
2868187	Links from items about works to items about their authors
2519365	Items with a taxon name
186519	Items about authors with an ORCID profile that has public content

Wikidata (**updated!**) statistics on WikiCite data from October 2018. Currently presented on the main page of Scholia.

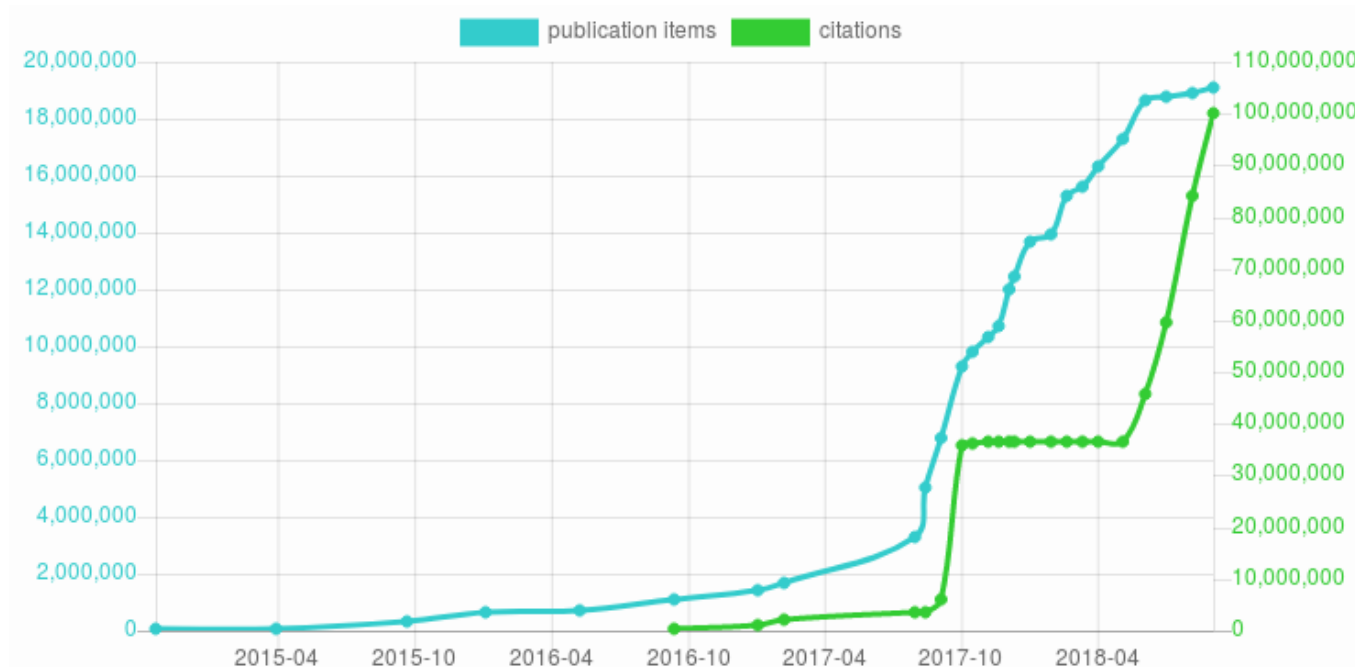
121 million citations.

17 million PubMed links.

14 million DOI links.

187 thousand ORCID links.

Jakob Voß' WikiCite statistics



Jakob Voß' Wikicite statistics that is update regularly.

<http://wikicite.org/statistics.html>

Number of publications and citations in Wikidata.

Note the staircase curve of the citations. My guess is that this shape is due to prolific James Hare using Europe PubMed Central initially and then switching to CrossRef for citations.

Scholia's aspects

Scholia Author Work Organization Location Event Award Topic Tools Help

Work

Scientific articles, conference articles, books, ...

Examples

Works	Comparisons
<ul style="list-style-type: none"> The Alzheimer's disease-associated amyloid beta-protein is an antimicrobial peptide 	<ul style="list-style-type: none"> "Protein measurement using the Folin phenol reagent" and "Novel method for the detection of beta-amyloid protein"

General
Disease
Taxon
Gene
Protein
Pathway
Chemical
Chemical class
Use

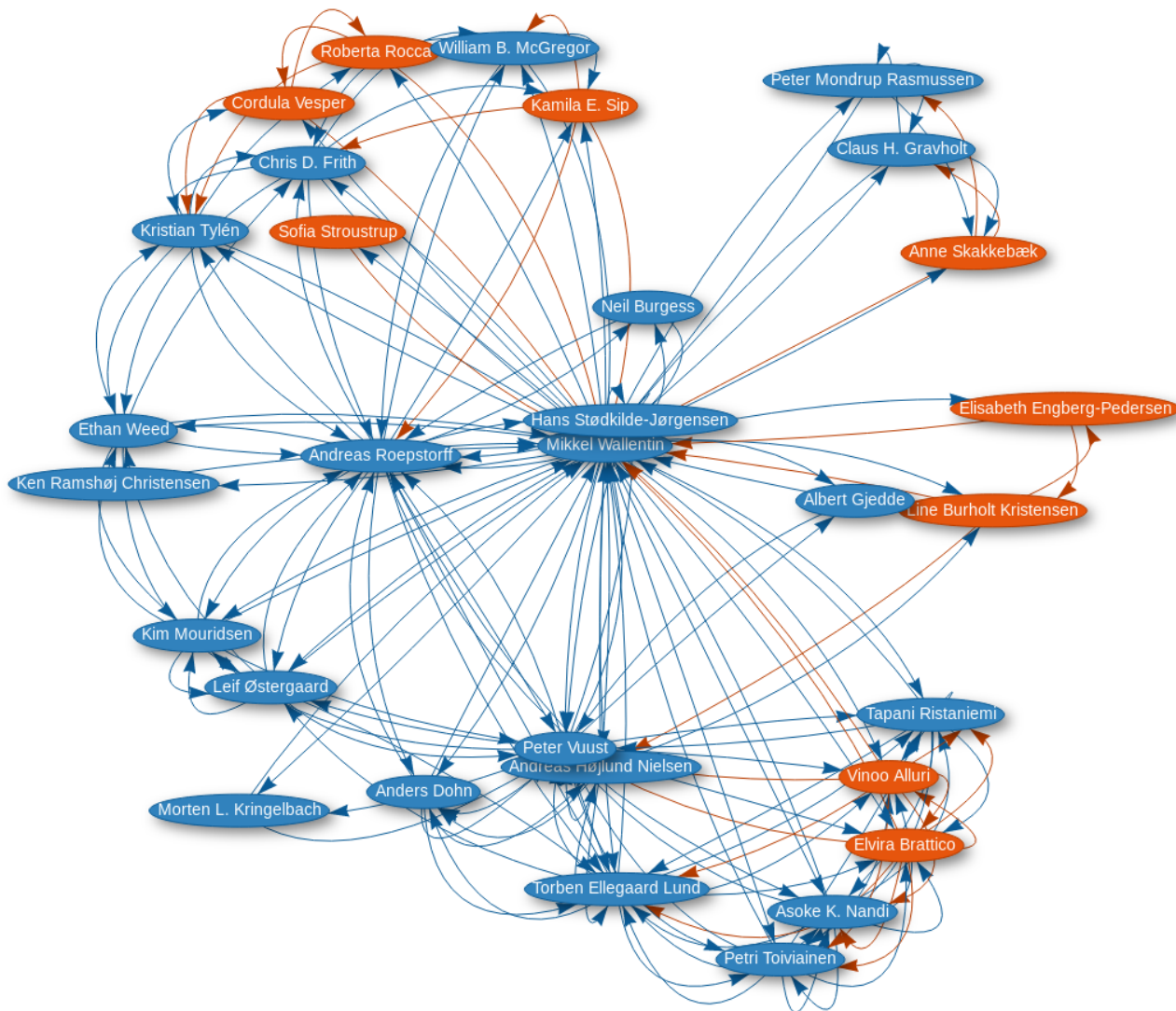
Scholia shows Wikidata data in *aspects*, author, work, organization (e.g., university, research group), venue (journal or conference), series, publisher, sponsor, location, event, award, topic, chemical, disease, etc.

For instance, the *Technical University of Denmark* may be viewed as a publisher, topic, organization, sponsor and location.

Author aspect: Co-author graph

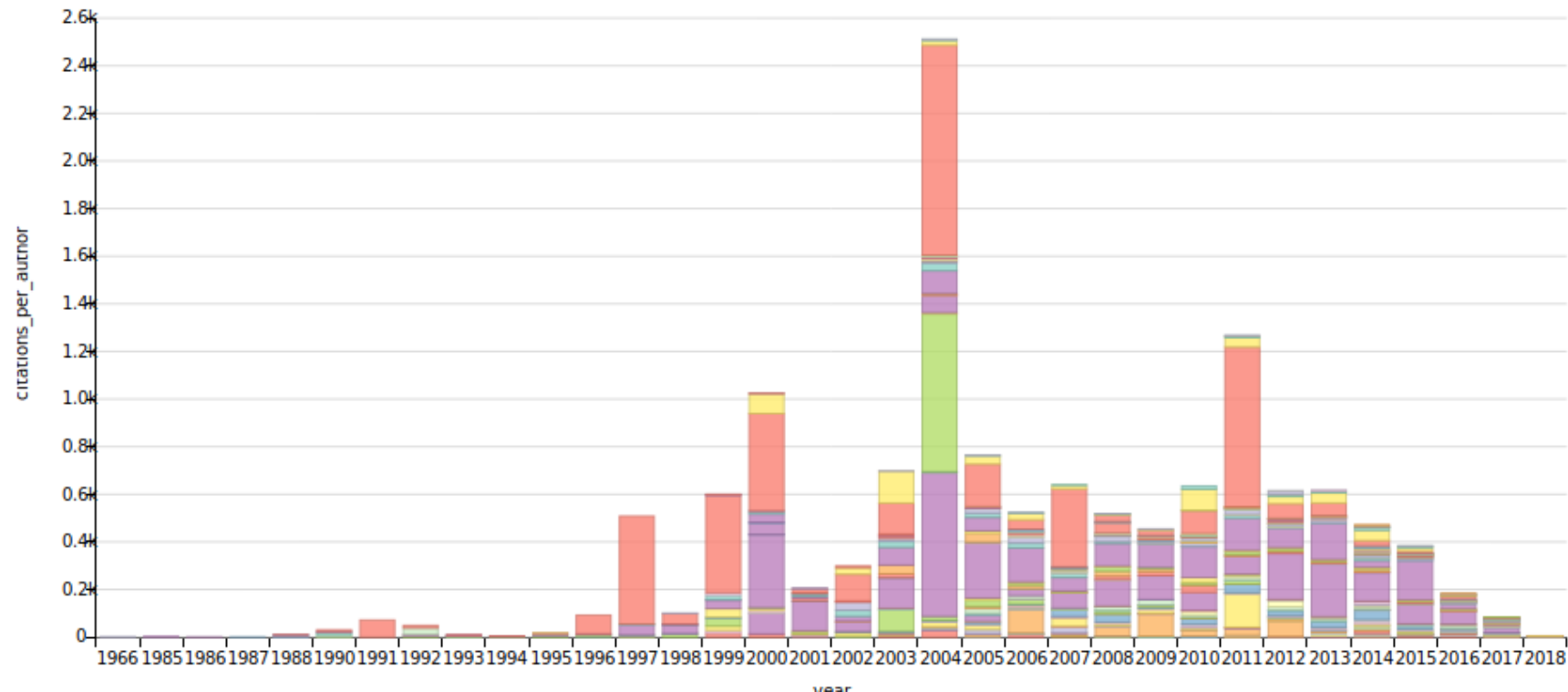
The egocentric co-author graph in Scholia's author aspect for the researcher **Mikkel Wallentin**, Aarhus University.

Colored according to gender.



Organization aspect: Citations

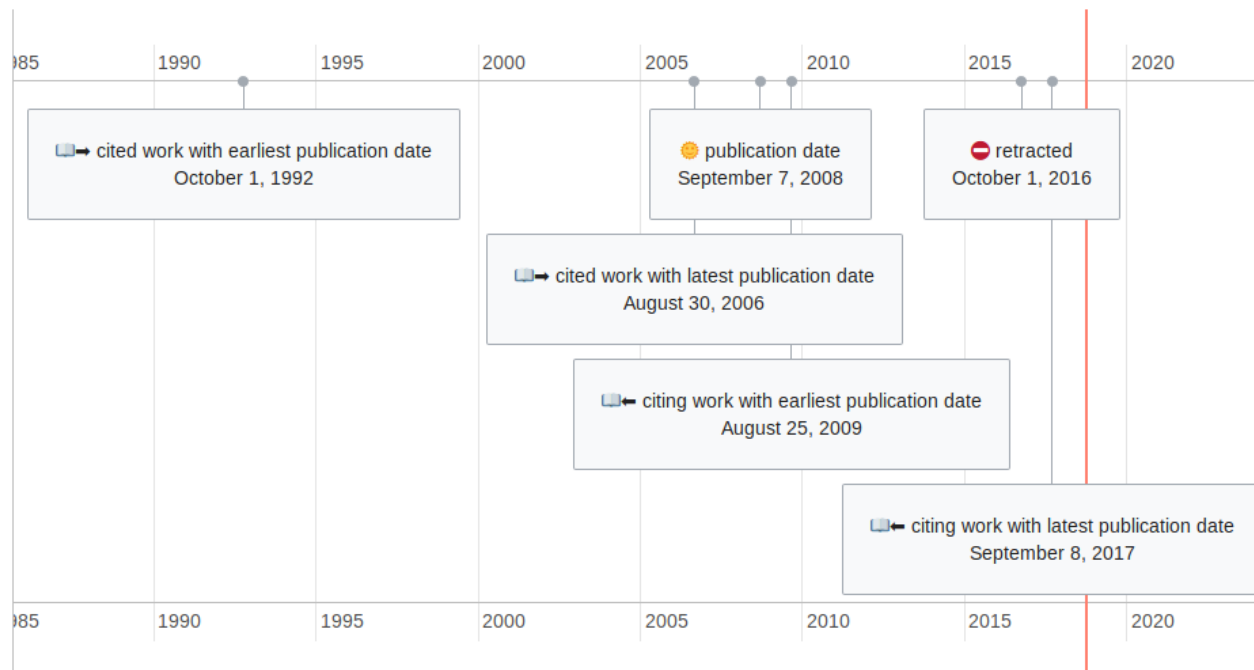
Co-author-normalized citations per year



Co-author normalized citations per year for Technical University of Denmark: Number of citations per year divided by number of co-authors on cited paper.

Work aspect: Retractions

Timeline

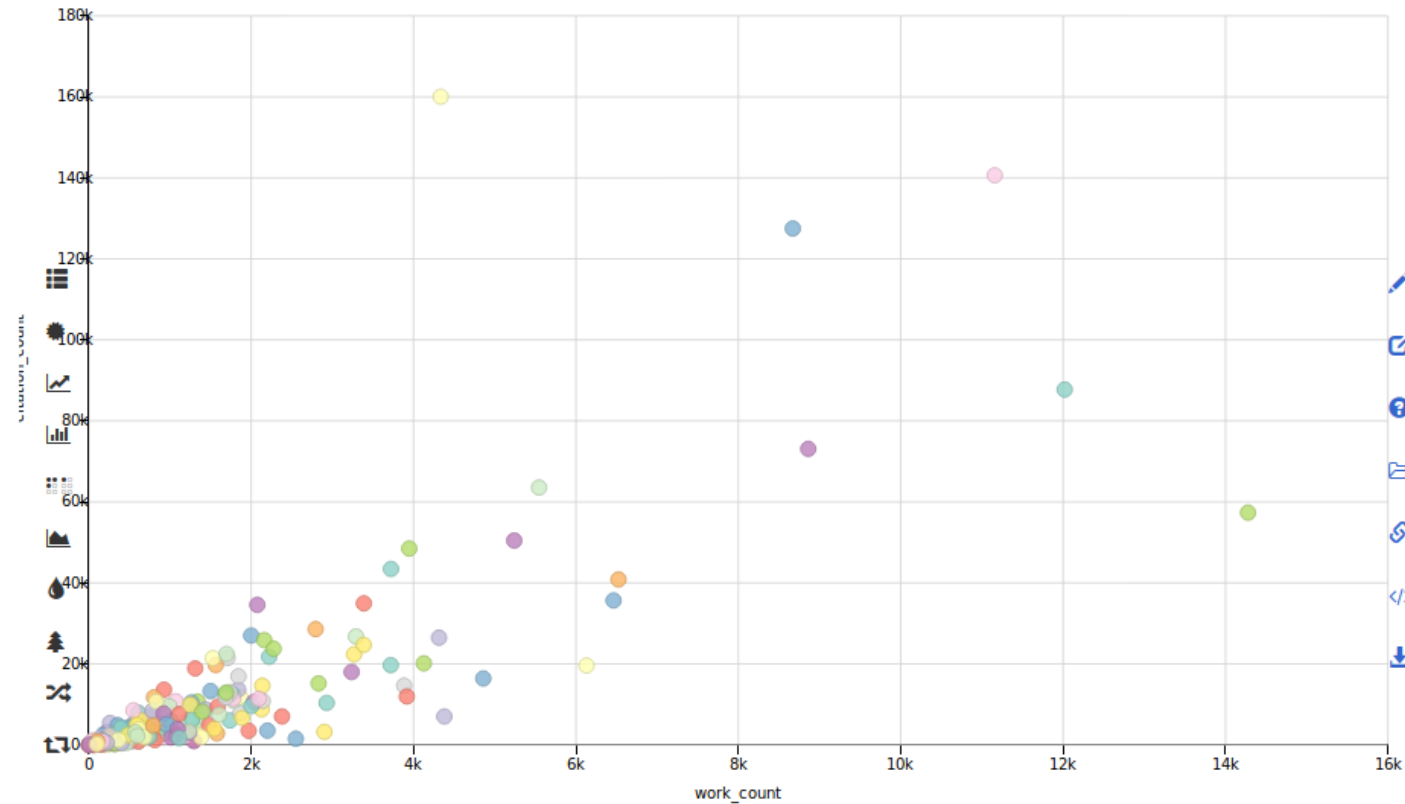


Wikidata can specify retracted papers, retraction notices and their connection.

By combining citation and retraction information we can find papers citing another paper after it has been retracted.

Currently, Scholia visualizes such information in a timeline. Here *Identification of Aurora-A as a direct target of E2F3 during G2/M cell cycle progression*: “For example, silencing E2F3 prevented entry into G2/M in ovarian cancer cells [61].” (received April 2016, accepted August 2017)

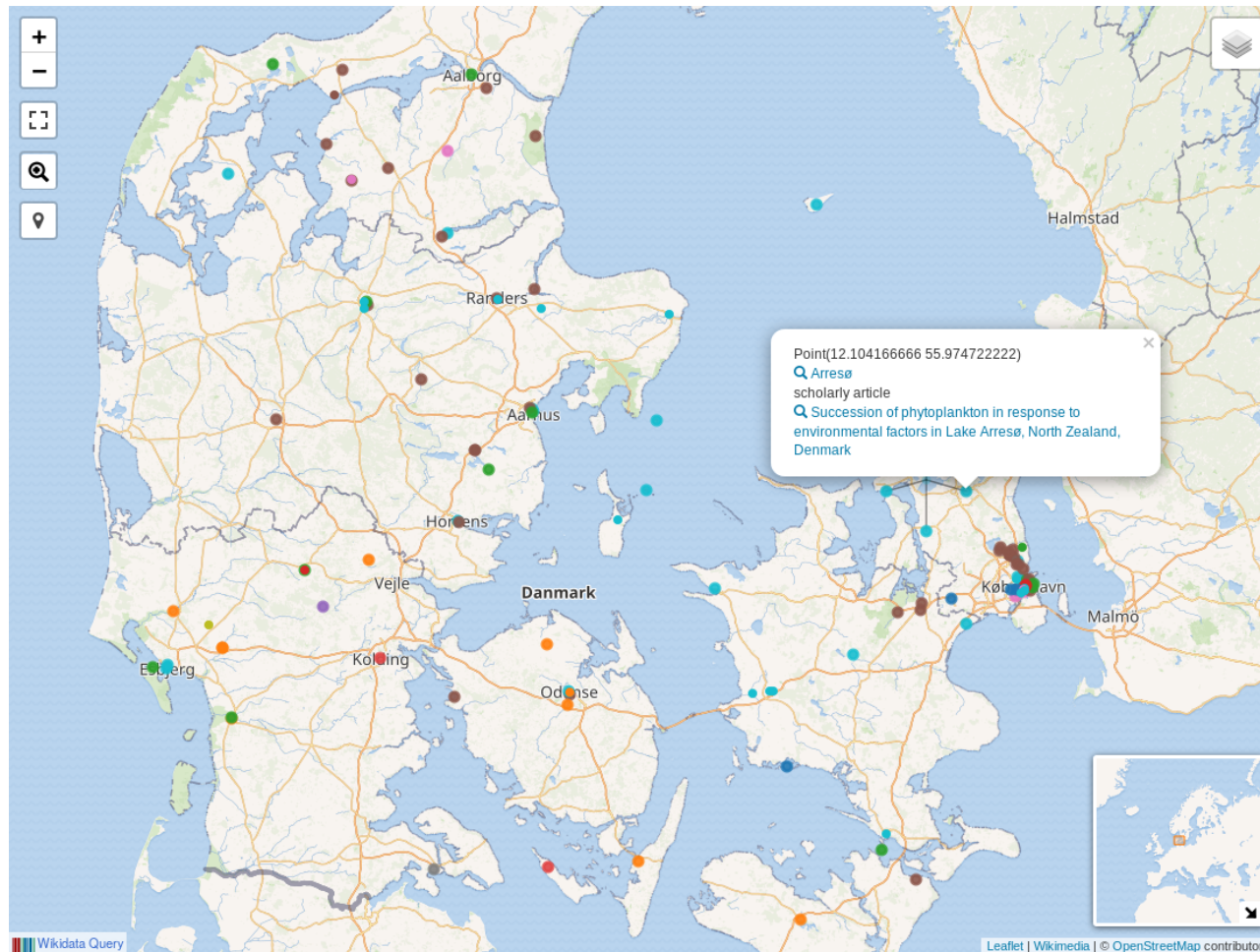
Publisher aspects



Scatter plot of number of citations as a function of number of works published in journals published under the **BioMed Central** brand.

The top left one is *Genome Biology*, the lower right *Critical Care*.

Country aspect



Locations in **Denmark** that is the main subject of a work (**Nielsen et al., 2018**).

Example popup: *Succession of phytoplankton in response to environmental factors in Lake Arresø, North Zealand, Denmark.*

Similar maps can be created for narrative locations.

Project aspect: Research projects in Scholia

Citations per budget

Show entries

Search:

Cites per_million	Citations	Budget	Currency	Short name	Project
207.40053358079109	894	4310500	euro	NANOMMUNE	Comprehensive assessment of hazardous effects of engineered nanomaterials on the immune system
193.09230169599405	54	279659	euro	ENRHES	Engineered Nanoparticles: Review of Health and Environmental Safety
126.71418448584886	19	149943.75	euro	SILKENE	SILKENE: Bionic silk with graphene or other nanomaterials spun by silkworms
88.94785719449311	429	4823050.42	euro	NEURONANO	Do nanoparticles induce neurodegenerative diseases? Understanding the origin of reactive oxidative species and protein aggregation and mis-folding phenomena in the presence of nanoparticles
64.33839298625732	84	1305596.8	euro	NANOTRANSKINETICS	Modelling basis and kinetics of nanoparticle interaction with membranes, uptake into cells, and sub-cellular and inter-compartmental transport
57.69595026013908	304	5269000.66	euro	ENPRA	Risk Assessment of Engineered Nanoparticles
49.57705673070313	195	3933271.01	euro	NANOTEST	Development of methodology for alternative testing strategies for the assessment of the toxicological profile of nanoparticles used in medical diagnostics
39.87060659140868	51	1279137.8	euro	MODNANOTOX	Modelling nanoparticle toxicity: principles, methods, novel approaches
36.2593836519345	118	3254330	euro	NANOTOES	Nanotechnology: Training Of Experts in Safety
29.24248324571952	365	12481840.1	euro	MARINA	Managing Risks of Nanoparticles

Research project aspect (Willighagen et al., 2018a).

If works are linked up to the project (by Wikidata's *sponsored by* property) we can make unusually statistics.

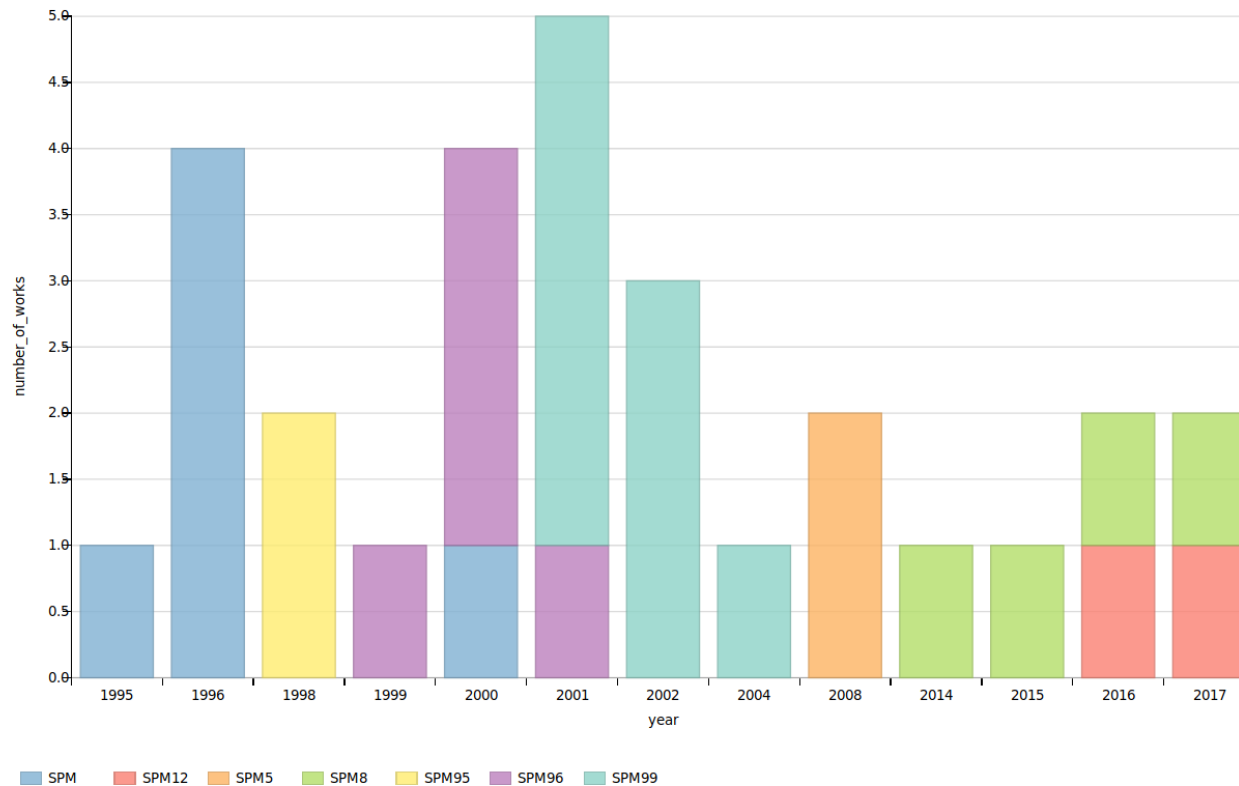
Here *citations per million budget*.

(The schema for projects and grants is not quite settled)

Use aspect

Usage over time

Works using the resource over time.

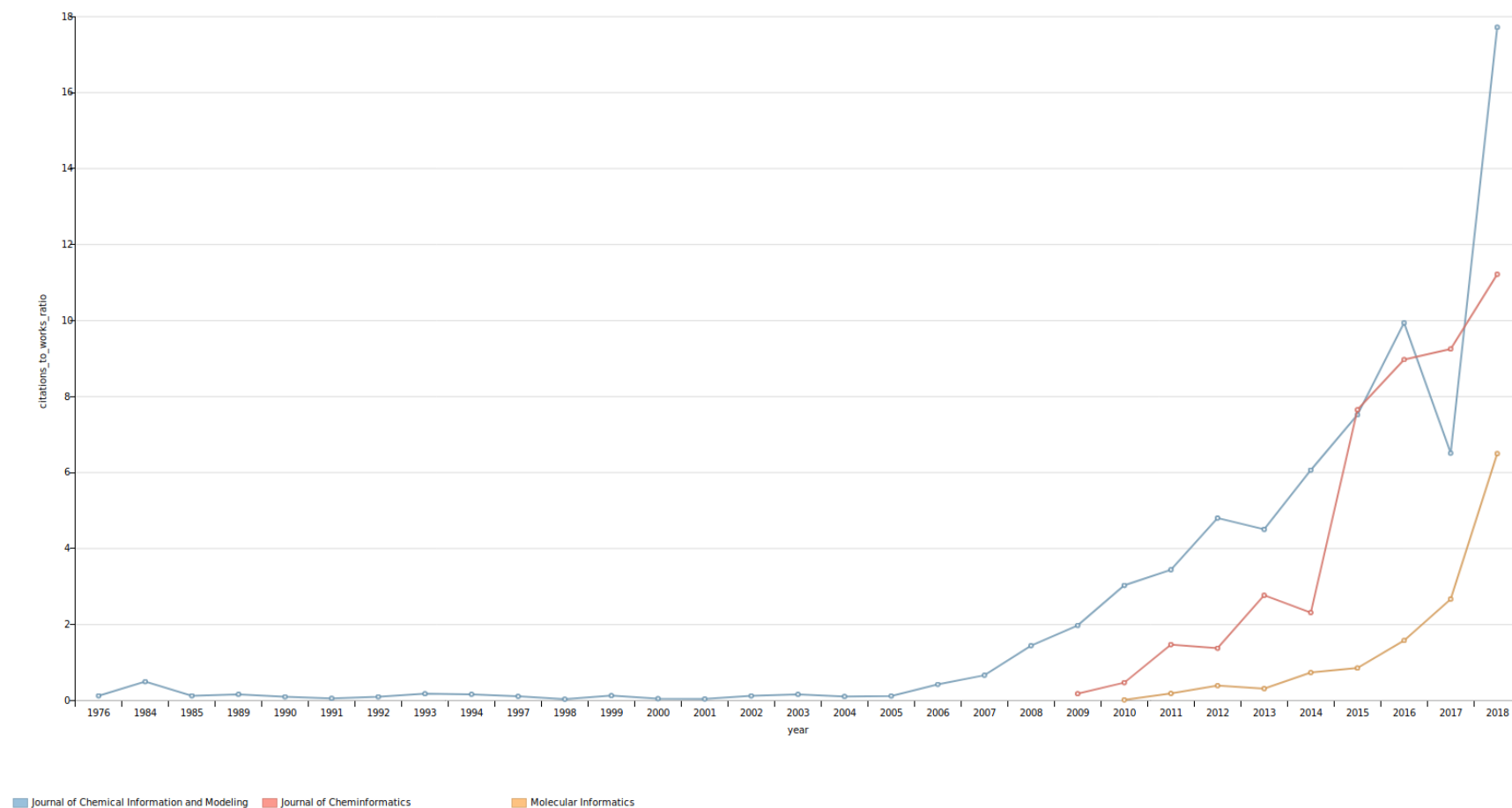


Bar chart for usage of SPM software (functional neuroimaging software) over time with different software versions indicated by color.

Uses the *describes a project that uses* property.

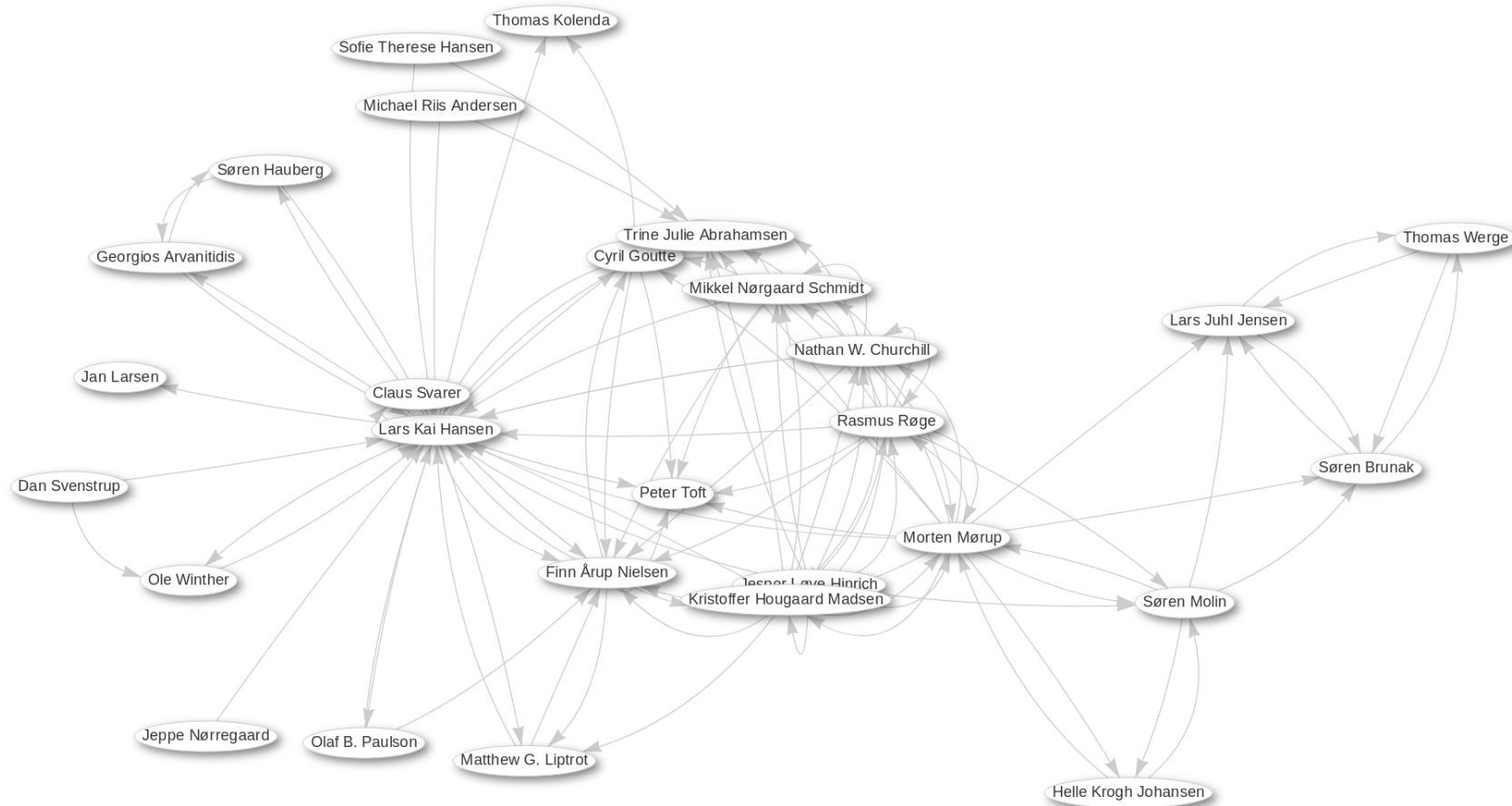
Such data is likely not available in directly machine readable format.

Comparison of multiple items



Multiple countries, e.g., some **Southern and Eastern African countries** or **cheminformatics journals** (here Willighagen's *citations to work ratio*).

Scholia's “subaspects”



Cocitation network for machine learning researchers in Denmark:
</scholia/country/Q33/topic/Q2539>.

Geodata and Scholia

Nearby researchers

Show entriesSearch:

Score	Author	Example work
24.178268894199626	Ulrike Cress	A productive clash of perspectives? The interplay between articles' and authors' perspectives and their impact on Wikipedia edits in a controversial domain
9.818634462803981	lassen Halatchliyski	A productive clash of perspectives? The interplay between articles' and authors' perspectives and their impact on Wikipedia edits in a controversial domain
1.604942154393766	Jason Weston	Reading Wikipedia to Answer Open-Domain Questions
0.16670001484301264	Denny Vrandečić	Revisiting reverts: accurate revert detection in Wikipedia
0.08335000742150632	Rudi Studer	Semantic Wikipedia
0.04167500371075316	Maria Koutraki	Wikipedia Infobox Type Prediction Using Embeddings
0.04167500371075316	Harald Sack	Wikipedia Infobox Type Prediction Using Embeddings

Wikipedia researchers near Tübingen: Weight information in Wikidata by the geographical distance and topic of authored works (Nielsen et al., 2018).

</scholia/location/Q3806/-topic/Q52>.

Nearby (in space and time) events also possible.

Related diseases with Wikidata Query Service

Genetically associated diseases

Other diseases with reported genetic association via genes, ordered according to number of co-associated genes.

Show entries

Search:

Count	Disease	Genes
14	bipolar disorder	NPAS3 // CACNA1C // ANK3 // MSRA // PTPRN2 // IFT88 // KCNMB2 // PHF8 // CNTNAP2 // ERC2 // COMMD10 // RIN2 // NLRC5 // MYO18B
5	obesity	PTPRN2 // CNTNAP2 // CTNNA3 // RIN2 // CSMD1
5	mental depression	NPAS3 // CDH13 // RORA // IFT88 // MYO18B
4	periodontitis	CDH13 // ERC2 // CSMD1 // NKAIN2
4	Alzheimer	RELN // CNTNAP2 // CSMD1 // NKAIN2
3	asthma	RORA // NOTCH4 // CTNNA3
2	coronary artery disease	TNIK // CSMD1
2	amyotrophic lateral sclerosis	ANK3 // KCNMB2
2	morbid obesity	TCF4 // SDCCAG8
2	major depressive disorder	CACNA1C // ANK3
2	multiple sclerosis	RELN // CSMD1
1	celiac disease/ allergic disorder	NKAIN2
1	smallpox	CSMD1
1	intracranial aneurysm	CNNM2
1	nicotine dependence	CTNNA3

Count some form of co-occurrences with a SPARQL query in the Wikidata Query service.

Scholia is doing this for diseases and proteins with tailor-made SPARQL. Here for the disease **schizophrenia**.

Shows genetically associated diseases via the **P2293** (genetic association) property.

Wembedder

Frontolimbic Serotonin 2A Receptor Binding in Healthy Subjects Is Associated with Personality Risk Factors for Affective Disorder (Q20984691)

Related: [Seasonal changes in brain serotonin transporter binding in short serotonin transporter linked polymorphic region-allele carriers but not in long-allele homozygotes](#) · [A nonlinear relationship between cerebral serotonin transporter and 5-HT\(2A\) receptor binding: an in vivo molecular imaging study in humans](#) · [Mining the posterior cingulate: Segregation between memory and pain components](#) · [Cerebral 5-HT2A receptor binding is increased in patients with Tourette's syndrome](#) · [Wikipedia in the eyes of its beholders: A systematic review of scholarly research on Wikipedia readers and readership](#) · ["The sum of all human knowledge": A systematic review of scholarly research on the content of Wikipedia](#) · [Cerebellar heterogeneity and its impact on PET data quantification of 5-HT receptor radioligands](#) · [Good Friends, Bad News - Affect and Virality in Twitter](#) · [The Center for Integrated Molecular Brain Imaging \(Cimbi\) database](#) · [A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs](#)

Finding related items based on word2vec-based knowledge graph embedding ([Nielsen, 2017](#)).

Here for [a scientific article](#).

In this case, the similar articles found are (probably) mostly related to coauthorship relations.

But a newer embedding would probably be much affected by the citation relations between papers.

Related items by co-citations

Count	Work
27	Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact
11	Twitter Predicts Citation Rates of Ecological Research.
10	How the scientific community reacts to newly submitted preprints: article downloads, Twitter mentions, and citations
9	Altmetrics: Value all research products
9	Characterizing social media metrics of scholarly papers: the effect of document properties and collaboration patterns
8	Tweeting birds: online mentions predict future citations in ornithology.
8	I Like, I Cite? Do Facebook Likes Predict the Impact of Scientific Work?
7	The differential impact of scientific quality, bibliometric factors, and social media activity on the influence of systematic reviews and meta-analyses about psoriasis.
7	A systematic identification and analysis of scientists on Twitter.
6	Social media release increases dissemination of original articles in the clinical pain sciences

Example with *Do altmetrics work? Twitter and ten other social web services.*

Counts citations back and forth, one step and two step with the SPARQL fragment:

```
wd:Q21133507
(^wdt:P2860 |
wdt:P2860) /
(^wdt:P2860 |
wdt:P2860)?
?work .
```

How do we get data into Wikidata?

Wikidata input

#	Item	main subject
1	Trapping the Tiger: Efficacy of the Novel BG-Sentinel 2 With Several Attractants and Carbon Dioxide for Collecting <i>Aedes albopictus</i> (Diptera: Culicidae) in Southern France Q22330695	<div>Asian tiger mosquito</div> <div>Culicidae</div> <div>Chikungunya Virus</div>
2	New vascular plant records for the Canadian Arctic Archipelago Q22583137	
3	Demography of some non-native isopods (Crustacea, Isopoda, Oniscidea) in a Mid-Atlantic forest, USA Q22675943	demographics
4	An Asiatic Chironomid in Brazil: morphology, DNA barcode and bionomics Q22675958	Brazil
5	Occurrence of <i>Diopatra marocensis</i> (Annelida, Onuphidae) in the eastern Mediterranean Q22680870	

Manual input on the <https://www.wikidata.org> website.

Magnus Manske's tools: SourceMD including its ORCIDator and resolver, Quickstatements, TABernacle (left screenshot). Relatively quick for each researcher if ORCID profile has DOI publications.

Other approaches: Fatameh, programmatic upload, e.g., with [WikidataIntegrator](#).

Scholia has arXiv and NeurIPS scraping.

Wikidata input example

Technical University of Denmark on Google Scholar

Take Frank Aarestrup as he is **not in Wikidata**

“Opret et nyt emne” (new item) on Wikidata

Find Frank Aarestrup on **ORCID**

Set the ORCID iD on Wikidata.

Go to Magnus Manske’s **sourceid** tool and copy-paste the Q-identifier:
Now Manske will automagically set up Aarestrup’s ORCID publications.

See also **Creating Structured Linked Data to Generate Scholarly Profiles: A Pilot Project using Wikidata and Scholia** (Lemus-Rojas and Odell, 2018).

Development

fnielsen / scholia

Unwatch 9 Star 41 Fork 9

Code Issues 330 Pull requests 0 Projects 16 Wiki Insights Settings

Filters Labels Milestones [New issue](#)

330 Open 108 Closed

Author Labels Projects Milestones Assignee Sort

- #517 "Nearby locations as topics in works" display same location multiple times **aspects** **bug**
- #516 From "missing" pages, link to TABernacle **SPARQL** **editing** **tools+workflows**
- #515 **In topic aspect, add panel for statements related to the topic that are referenced to scholarly publications** **SPARQL** **panels**
- #514 **In topic aspect, add panel on relevant events** **P921-main-subject** **SPARQL** **maps** **panels** **timelines**
- #513 **Today's publications** **SPARQL** **aspects** **panels**
- #512 **In event aspect, add panel for organizations involved** **SPARQL** **aspects** **panels**
- #511 **"| wdt:P5804" is missing in the co-author graph in event aspect**
- #510 **In co-authorship graph for events, add non-attending co-authors** **P50-author**
- #509 **In queries about works, use P1476 (title) instead of label to save time** **P1476-title** **Wikidata-labels**
- #507 **Improve the performance of the "Recently published works" in venue aspect** **SPARQL** **aspects** **enhancement** **performance**
- #506 **Related paper by co-citation analysis in the work aspect** **aspects** **enhancement**

Development takes place on GitHub under GPL at <https://github.com/fnielsen/scholia/>.

Three developers: Egon Willighagen (almost all cheminformatics aspects, biological pathways, etc., see also (Willighagen et al., 2018b)) and Daniel Mietchen.

Provided a Python development environment, you can download and run Scholia on your own computer.

Thanks!

References

- Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2016). [Enriching Word Vectors with Subword Information](#).
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. N. (2018). [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](#).
- Gabrilovich, E. and Markovitch, S. (2006). [Overcoming the brittleness bottleneck using Wikipedia: enhancing text categorization with encyclopedic knowledge](#). *Proceedings of the Twenty-First AAAI Conference on Artificial Intelligence*, 2:1301–1306.
- Lemus-Rojas, M. and Odell, J. D. (2018). [Creating Structured Linked Data to Generate Scholarly Profiles: A Pilot Project using Wikidata and Scholia](#). *Journal of Librarianship and Scholarly Communication*, 6. DOI: [10.7710/2162-3309.2272](#).
- Masukume, G. (2014). [Insights into abdominal pregnancy](#). 1. DOI: [10.15347/WJM/2014.012](#).
- Mehdi, M., Okoli, C., Mesgari, M., Nielsen, F. Å., and Lanamäki, A. (2017). Excavating the mother lode of human-generated text: A systematic review of research that uses the Wikipedia corpus. *Information Processing & Management*, 53:505–529. DOI: [10.1016/J.IPM.2016.07.003](#).
- Melchior, L., Kivisild, T., Lynnerup, N., and Dissing, J. (2008). Evidence of authentic DNA from Danish Viking Age skeletons untouched by humans for 1,000 years. *PLOS ONE*, 3:e2214. DOI: [10.1371/JOURNAL.PONE.0002214](#).
- Mietchen, D., Wodak, S., Wasik, S., Szostak, N., and Dessimoz, C. (2018). Submit a Topic Page to PLOS Computational Biology and Wikipedia. *PLOS Computational Biology*, 14:e1006137. DOI: [10.1371/JOURNAL.PCBI.1006137](#).
- Nielsen, F. Å. (2017). [Wembedder: Wikidata entity embedding web service](#). DOI: [10.5281/ZENODO.1009127](#).

Nielsen, F. Å. and Hansen, L. K. (2017). [Open semantic analysis: The case of word level semantics in Danish](#). *Human Language Technologies as a Challenge for Computer Science and Linguistics*, pages 415–419.

Nielsen, F. Å., Mietchen, D., and Willighagen, E. (2018). [Geospatial data and Scholia](#). *Proceedings of the 3rd International Workshop on Geospatial Linked Data and the 2nd Workshop on Querying the Web of Data*. DOI: [10.5281/ZENODO.1202256](#).

Stoeger, A. S., Mietchen, D., Oh, S., de Silva, S., Herbst, C. T., Kwon, S., and Fitch, W. T. (2012). An Asian elephant imitates human speech. *Current Biology*, 22:2144–2148. DOI: [10.1016/J.CUB.2012.09.022](#).

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., and Dessimoz, C. (2013). Approximate Bayesian computation. *PLOS Computational Biology*, 9:e1002803. DOI: [10.1371/JOURNAL.PCBI.1002803](#).

West, A. G., Chang, J., Venkatasubramanian, K., Sokolsky, O., and Lee, I. (2011). Link spamming Wikipedia for profit. *Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference*. DOI: [10.1145/2030376.2030394](#).

Willighagen, E., Jahn, N., and Nielsen, F. Å. (2018a). The EU NanoSafety Cluster as Linked Data visualized with Scholia. DOI: [10.6084/M9.FIGSHARE.6727931](#).

Willighagen, E., Slenter, D., Mietchen, D., Evelo, C. T., and Nielsen, F. Å. (2018b). [Wikidata and Scholia as a hub linking chemical knowledge](#). *11th International Conference on Chemical Structures. Program & Abstracts*, page 146. DOI: [10.6084/M9.FIGSHARE.6356027.V1](#).