Wikidata

The free and open knowledge base

Wikidata for Research: Enabling Open Science and Resource Discovery

Ewan McAndrew – @emcandre

ewan.mcandrew@ed.ac.uk

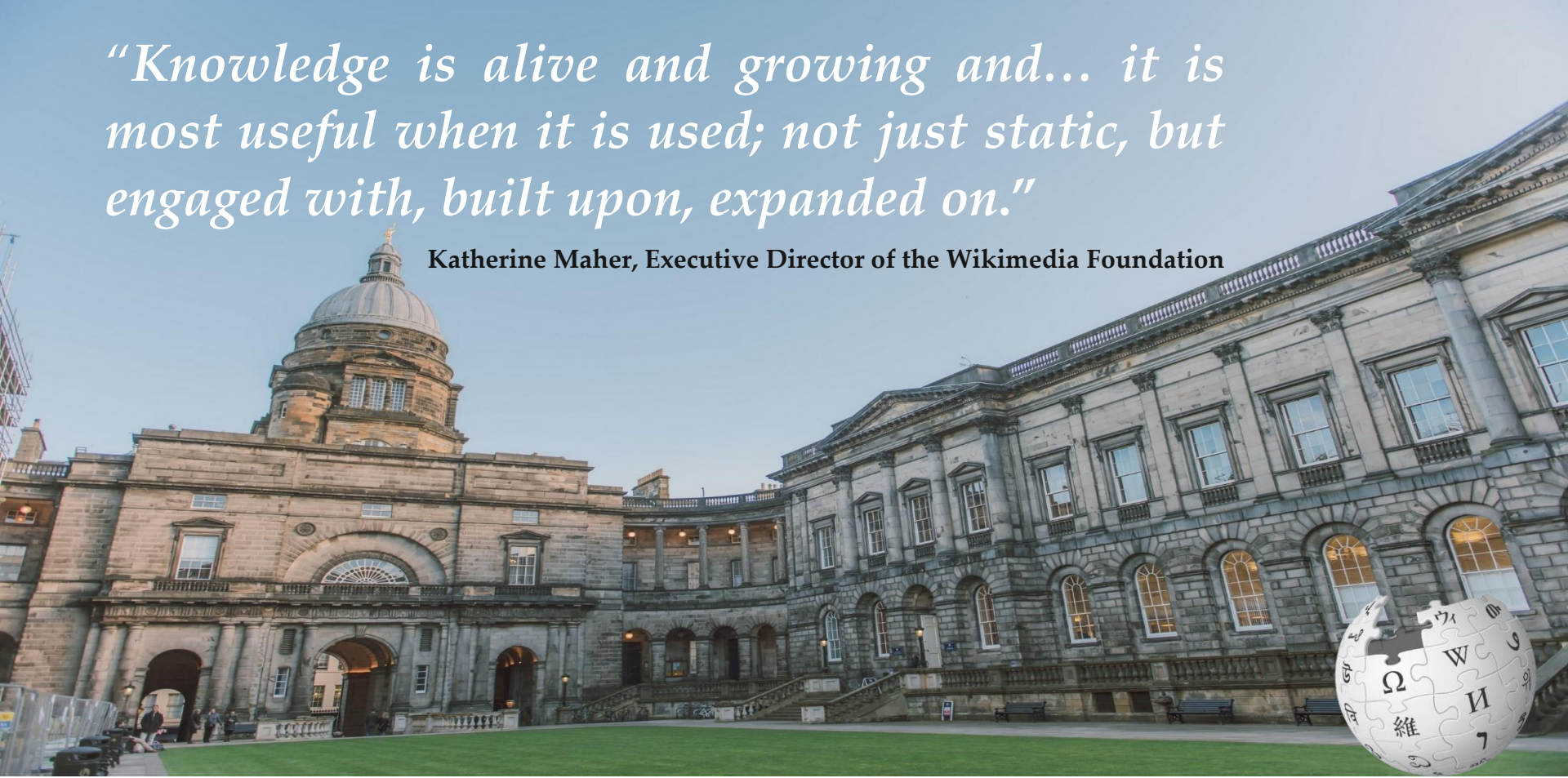
Slides: http://tinyurl.com/SummerOfData

Wikimedia's family of open knowledge projects

https://wikimediafoundation.org/wiki/Our_projects

"Knowledge is alive and growing and… it is most useful when it is used; not just static, but engaged with, built upon, expanded on."

Katherine Maher, Executive Director of the Wikimedia Foundation



WIKIPEDIA
The Free Encyclopedia

Robert Louis Stevenson

From Wikipedia, the free encyclopedia

Robert Louis Balfour Stevenson (13 November 1850 – 3 December 1894) was a Scottish novelist, poet, essayist, and travel writer. His most famous works are *Treasure Island*, *Kidnapped*, *Strange Case of Dr Jekyll and Mr Hyde* and *A Child's Garden of Verses*.

A literary celebrity during his lifetime, Stevenson now ranks among the 26 most translated authors in the world.[1] His works have been admired by many other writers, including Jorge Luis Borges, Bertolt Brecht, Marcel Proust, Arthur Conan Doyle, Henry James, Cesare Pavese, Ernest Hemingway, Rudyard Kipling, Jack London, Vladimir Nabokov,[2] J. M. Barrie,[3] and G. K. Chesterton, who said of him that he "seemed to pick the right word up on the point of his pen, like a man playing spillikins."[4]

Works [edit | edit source]

Novels [edit | edit source]

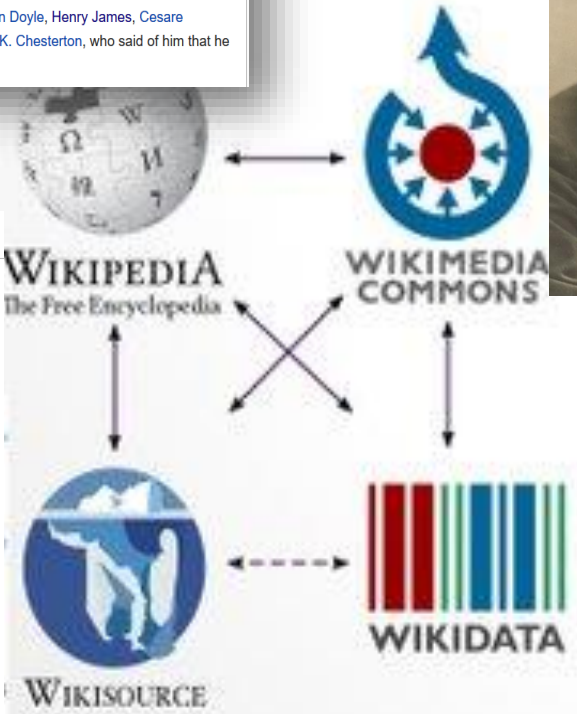
- Treasure Island (1883)
- Prince Otto (1885)
- Strange Case of Dr Jekyll and Mr Hyde (1886)
  - The Annotated Strange Case of Dr Jekyll and Mr Hyde
- Kidnapped (*Adventures of David Balfour*, 1886)
- The Black Arrow (1888)
- The Master of Ballantrae (1889) (External scan)
- Catriona (*Adventures of David Balfour*, 1893)
- Weir of Hermiston (1896, unfinished)
- St. Ives (1898, unfinished)

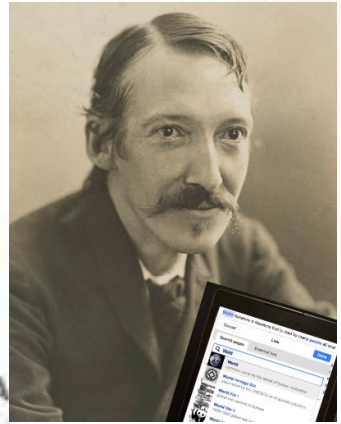
with Lloyd Osbourne [edit | edit source]

- The Wrong Box (1889) (External scan)
- The Wrecker (1892)
- The Ebb-Tide (1894) (External scan)

Short stories [edit | edit source]

- A Lodging for the Night (1877)
- The Sire de Maletroit's Door (1878)





Public Domain pic of Robert Louis Stevenson https://commons.wikimedia.org/wiki/File:Robert_Louis_Stevenson_by_Henry_Walter_Barnett.jpg

ROBERT LOUIS STEVENSON (Q1512)

Spouse (P26): Fanny Stevenson Q3066501
Place of birth (P19): Edinburgh Q23436
Place of death (P20): Vailima, Samoa Q548806
Father (P22): Thomas Stevenson Q325068
Educated at (P69): Edinburgh Uni Q160302
Occupation (P106): Writer Q38180
Poet Q49757
Novelist Q6625963
Essayist Q11774202

What is Wikidata?

Established 2012.
All data is CC-0 licensed.

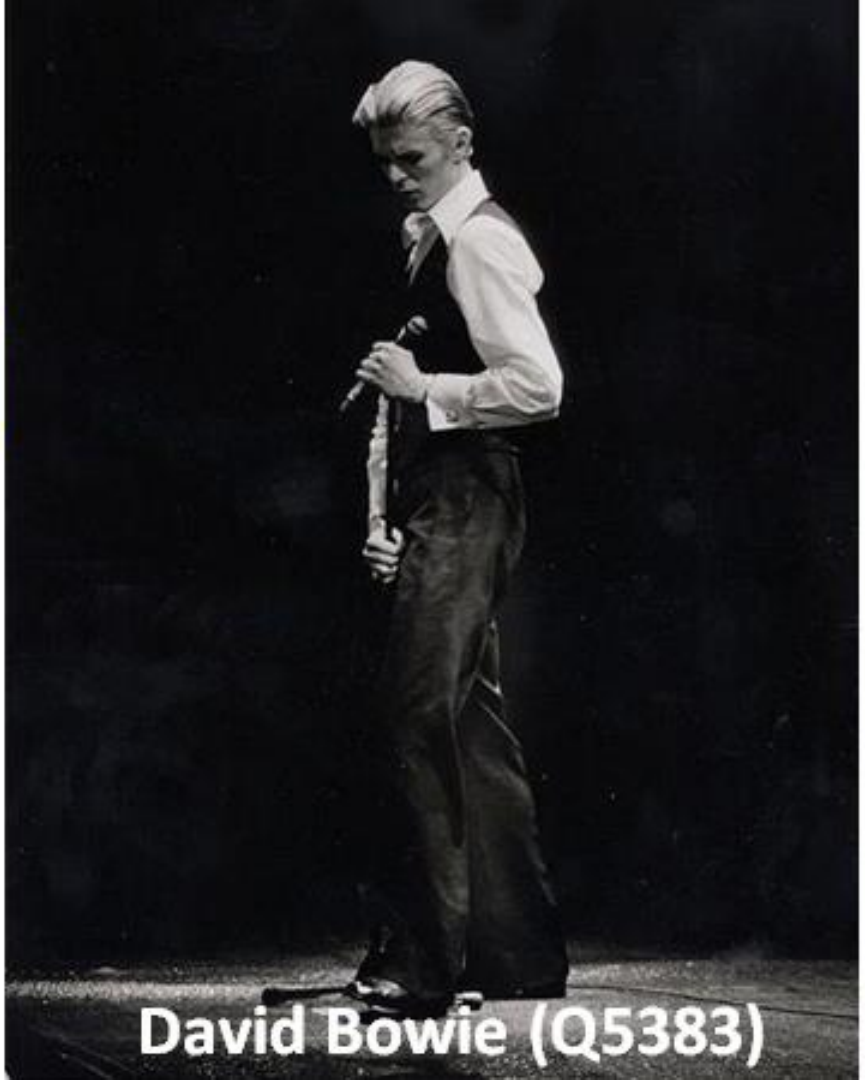
Wikidata is a free linked database of structured secondary data that can be read and edited by both humans and machines.

Wikidata acts as central storage for the structured data of its Wikimedia sister projects including Wikipedia, Wikivoyage, Wikisource, and others.

- Bibliographic
- Biographic
- Biomedical
- Geographic
- Taxonomic
- Authority file
- And more besides

date of birth (P569)

8 January 1947



English language label for Q34
Sweden (Q34)
Unique Q no. for item
Value
population (P1082)
9,747,355
according to: Statistics Sweden
method: estimation
point in time: 31 Dec 2014
Reference
Qualifier
Property
picture by Hector Melo A.

A provenance-preserving Answer Engine

+

The sum of all **human knowledge**

The sum of all **data and sources** backing human knowledge

Wikidata key stats – data and community

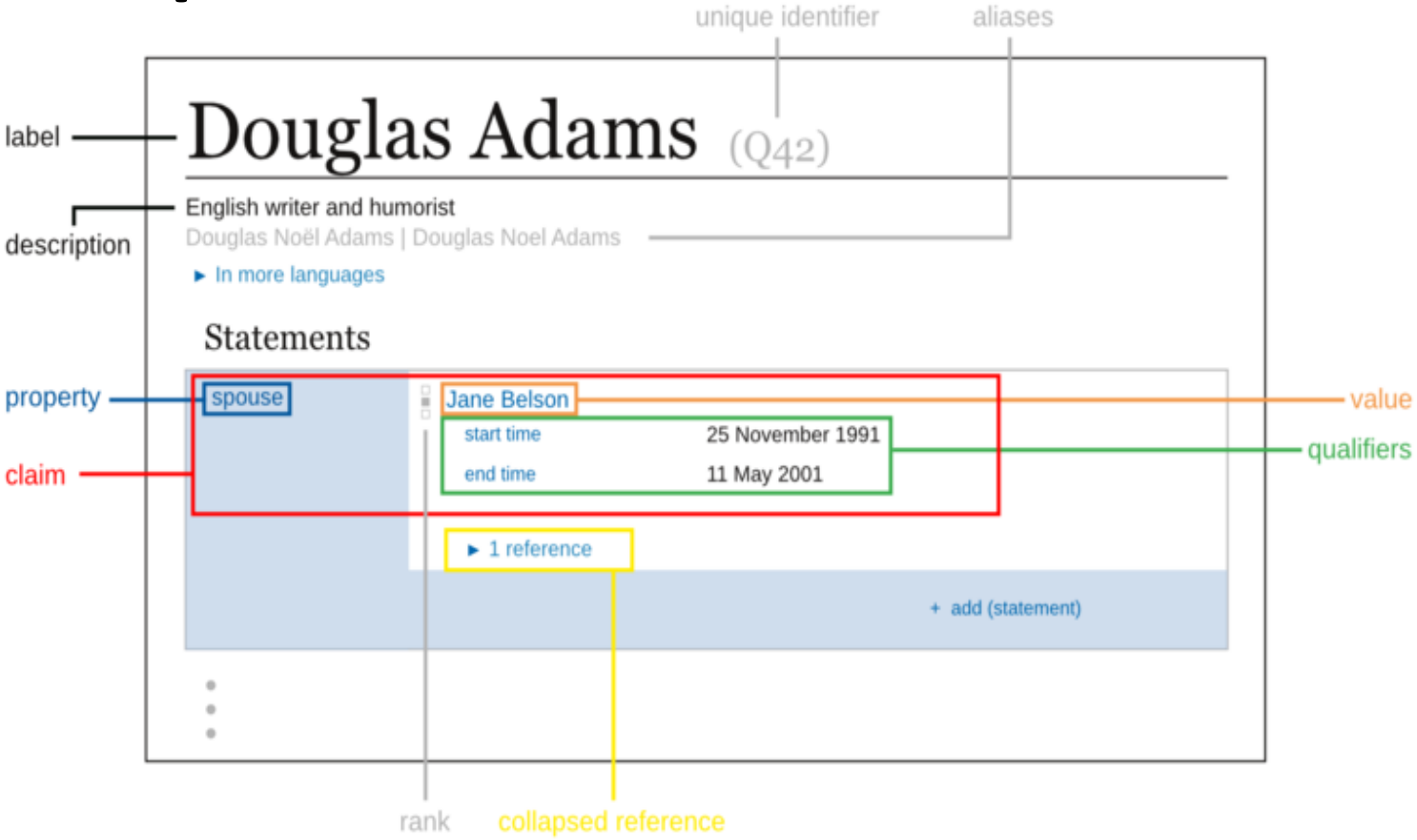
- 33 million items
- 195 million statements
- 541 million edits
- 17,792 active editors

- 14 million more items since August 2016
- 95 million more statements since August 2016.
- 1200 more active editors since August 2016.

"Data is beautiful. Data is information."

Official Wikidata stats More stats

Example Wikidata item & statement



wikidata.org/wiki/Q42

Adding data to Wikidata

Firstly consult Wikidata's Data Import Hub.

1. **Manual editing** by user. (Video tutorial 1) (Video tutorial 2).

2. **Mass editing** using tools e.g.

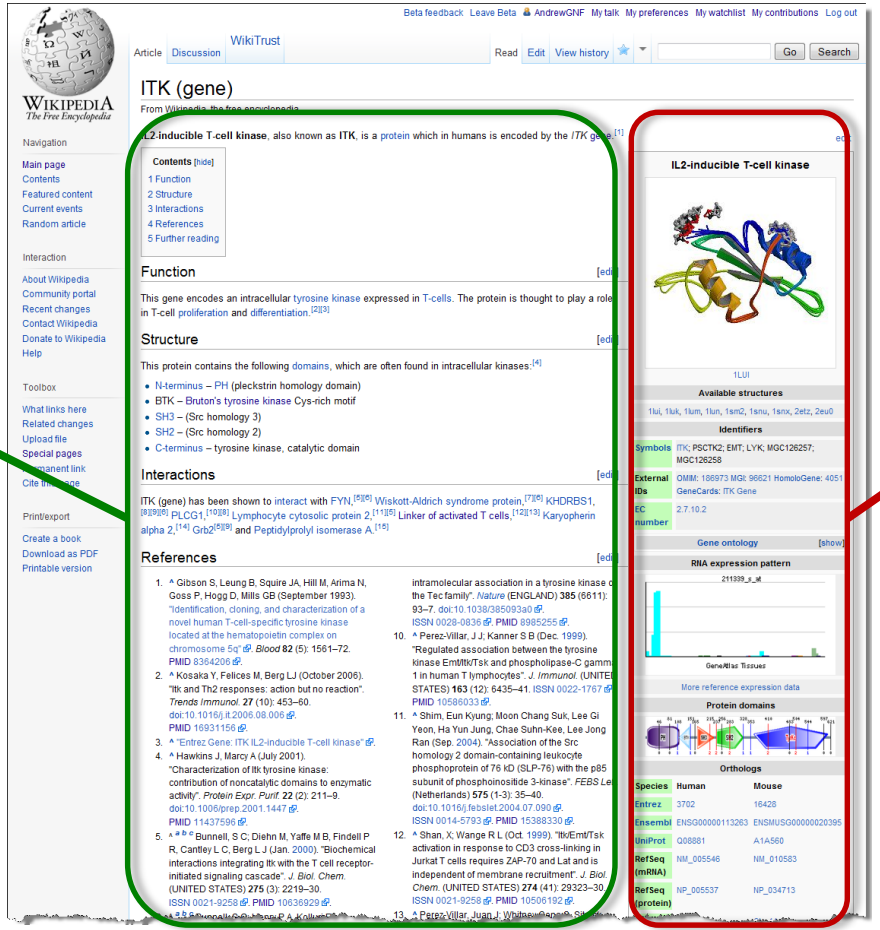
a) The Mix n Match tool allows you to match an external data set to Wikidata items.

b) Importing data from Google Spreadsheets into the Quickstatements tool. (Video tutorial 1) (Video tutorial 2). (Exemplar sheet). The syntax you need is explained in QuickStatements v.1 and you can use the Wikidata plug-in for Google Sheets to make formating the data in Google Sheets easier.

3. **Bot editing** (making edits over the API) e.g. ProteinBoxBot

See the Bots request section of the Data Import Guide.

Enabling Open Science **e.g.** The Gene Wiki project circa 2008



Summarized knowledge via crowdsourcing

Data imported from structured databases

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4795929/

Huss, PLoS Biol, 2008

Reelin

From Wikipedia, the free encyclopedia

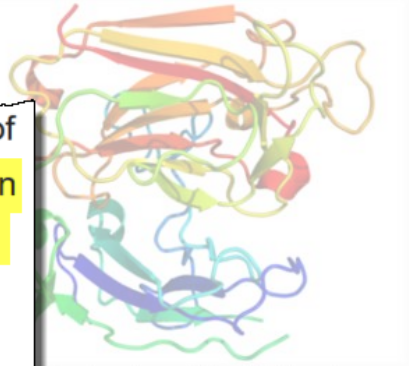
Reelin is a large secreted extracellular matrix glycoprotein that helps regulate processes of neuronal migration and positioning in the developing brain by controlling cell–cell interactions. Besides this important role in early development, reelin continues to work in the adult brain. It modulates synaptic plasticity by [2][3] It also stimulates dendrite[4] migration of neuroblasts generat zones. It is found not only in the tissues.

Reelin has been suggested to b expression of the protein has be bipolar disorder, but the cause of this observation remains uncertain as studies show that psychotropic medication itself affects reelin expression. Moreover, epigenetic hypotheses aimed at explaining the changed levels of reelin expression[6] are controversial.[7][8] Total lack of reelin causes a form of lissencephaly. Reelin may also play a role in Alzheimer's disease, temporal lobe epilepsy and autism.

Reelin's name comes from the abnormal reeling gait of *reeler* mice,[9] which were later found to have a deficiency of this brain protein and were homozygous for mutation of the RELN gene. The

Reelin has been suggested to be implicated in pathogenesis of several brain diseases. The expression of the protein has been found to be significantly lower in schizophrenia and psychotic bipolar disorder, but the cause of this observation remains uncertain as studies show that psychotropic medication itself

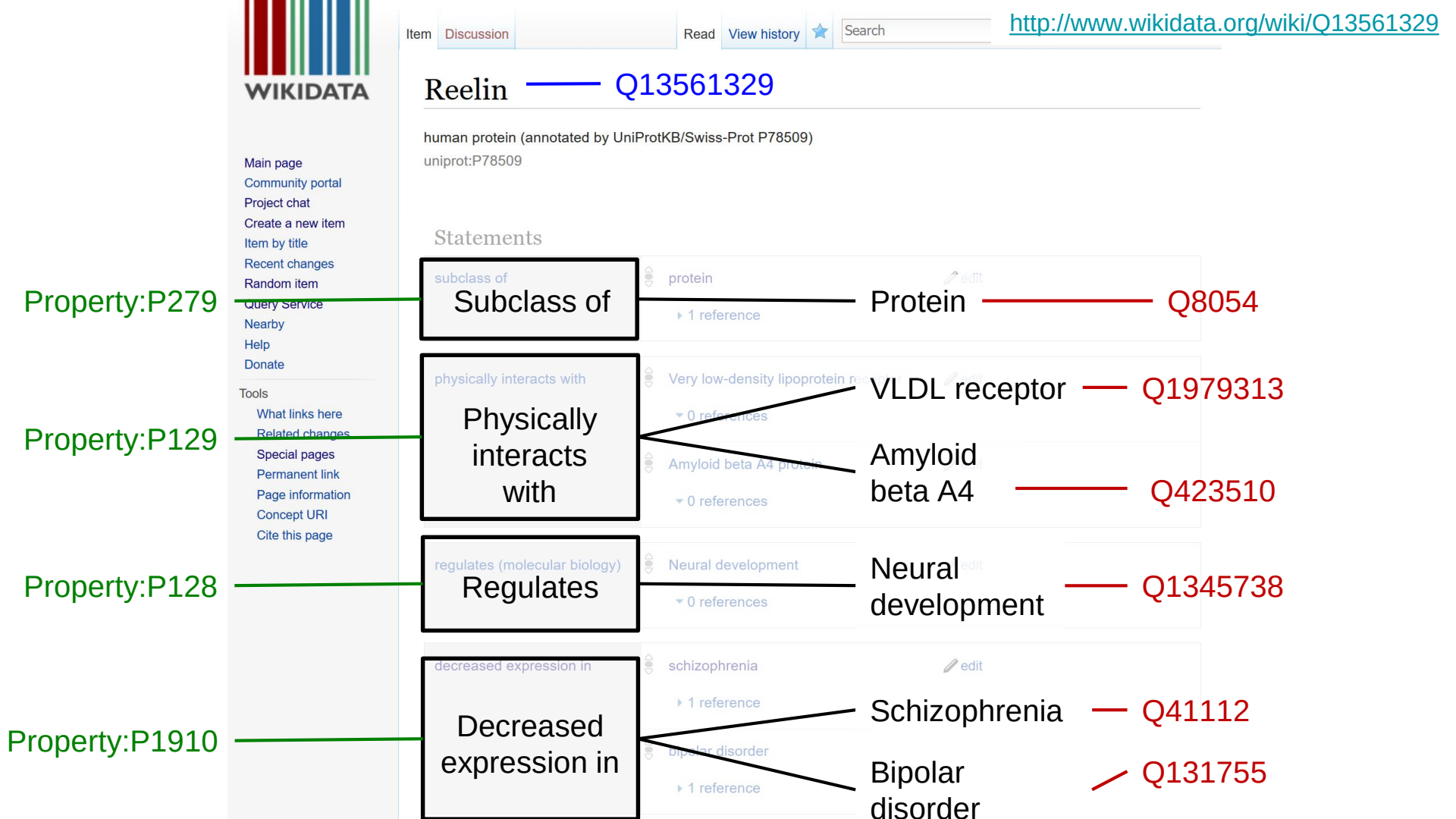
Reelin

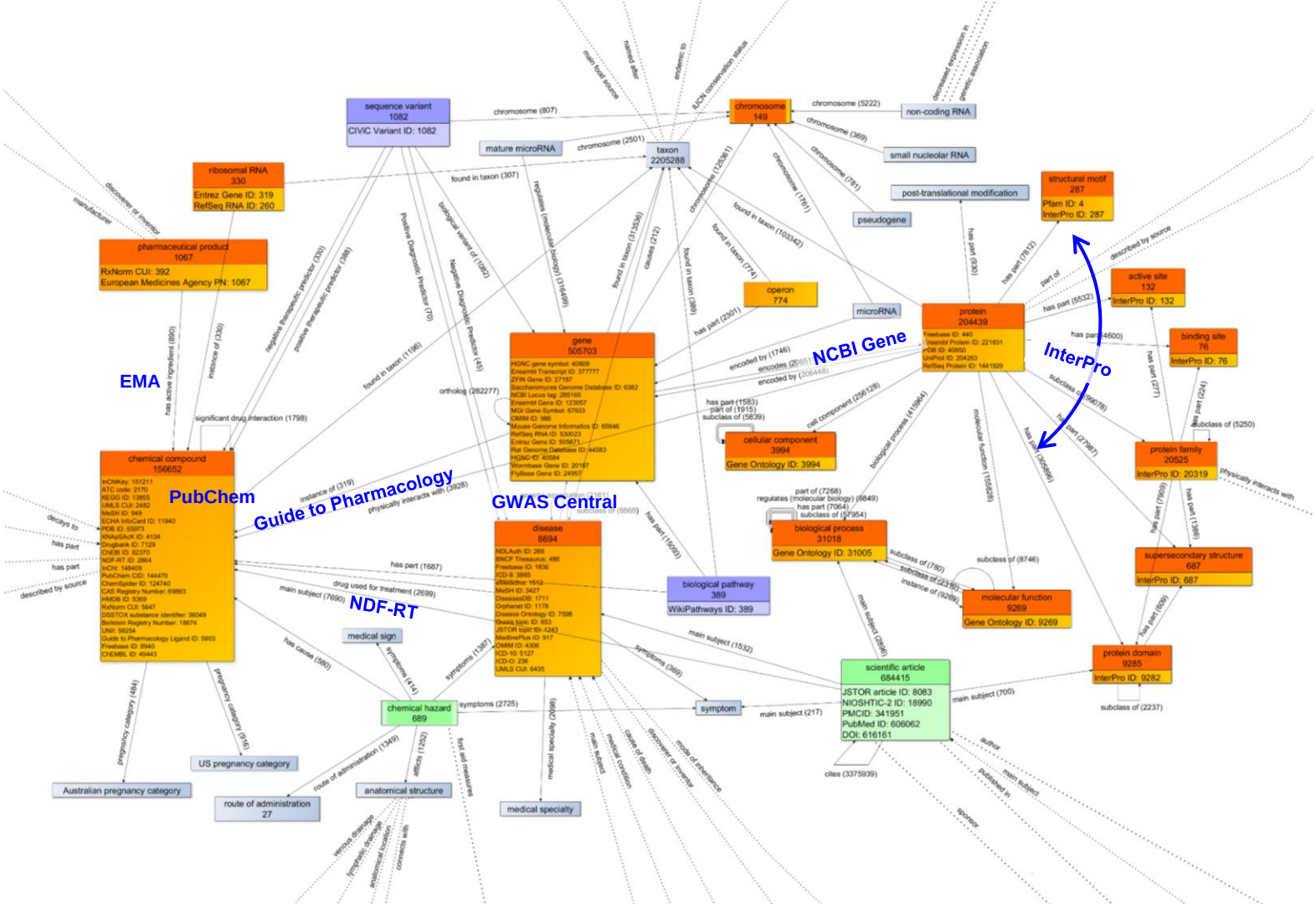


hic structure of the third reelin repeat domain.[1]

Available structures	
PDB	Ortholog search: PDBe, RCSB
	List of PDB id codes [show]

Identifiers	
Symbols	RELN ; LIS2; PRO1598; RL
External	OMIM: 600514 MGI: 103022


WIKIDATA
Item
Discussion
Read
View history
Search
http://www.wikidata.org/wiki/Q13561329
Reelin
Q13561329
human protein (annotated by UniProtKB/Swiss-Prot P78509)
uniprot:P78509
Main page
Community portal
Project chat
Create a new item
Item by title
Recent changes
Random item
Query Service
Nearby
Help
Donate
Tools
What links here
Related changes
Special pages
Permanent link
Page information
Concept URI
Cite this page
Statements
Property:P279
subclass of
Subclass of
protein
Protein
Q8054
1 reference
Property:P129
physically interacts with
Physically interacts with
Very low-density lipoprotein
VLDL receptor
Q1979313
0 references
Amyloid beta A4 protein
Amyloid beta A4
Q423510
0 references
Property:P128
regulates (molecular biology)
Regulates
Neural development
Neural development
Q1345738
0 references
Property:P1910
decreased expression in
Decreased expression in
schizophrenia
edit
Schizophrenia
Q41112
1 reference
bipolar disorder
Bipolar disorder
Q131755
1 reference


sequence variant
1082
CIViC Variant ID: 1082
chromosome
149
chromosome (807)
chromosome (5222)
non-coding RNA
chromosome (369)
small nucleolar RNA
mature microRNA
chromosome (2501)
taxon
2205288
ribosomal RNA
330
Entrez Gene ID: 319
RefSeq RNA ID: 260
found in taxon (307)
pharmaceutical product
1067
RxNorm CUI: 392
European Medicines Agency PN: 1067
EMA
operon
774
gene
505703
HGNC gene symbol: 40809
Ensembl Transcript ID: 377777
ZFIN Gene ID: 27187
Saccharomyces Genome Database ID: 6382
NCBI Locus tag: 285166
Ensembl Gene ID: 123057
MGI Gene Symbol: 67933
OMIM ID: 986
Mouse Genome Informatics ID: 65946
RefSeq RNA ID: 530023
Entrez Gene ID: 505671
Rat Genome Database ID: 44583
HGNC ID: 40584
Wormbase Gene ID: 20167
FlyBase Gene ID: 24957
NCBI Gene
microRNA
protein
204439
Freebase ID: 440
Ensembl Protein ID: 221831
PDB ID: 40650
UniProt ID: 204263
RefSeq Protein ID: 1441929
InterPro
structural motif
287
Pfam ID: 4
InterPro ID: 287
post-translational modification
pseudogene
active site
132
InterPro ID: 132
binding site
76
InterPro ID: 76
chemical compound
156652
InChIKey: 151211
ATC code: 2170
KEGG ID: 13855
UMLS CUI: 2482
MeSH ID: 949
ECHA InfoCard ID: 11940
PDB ID: 55973
KNApSAcK ID: 4134
Drugbank ID: 7129
ChEBI ID: 82370
NDF-RT ID: 3864
InChI: 148409
PubChem CID: 144470
ChemSpider ID: 124740
CAS Registry Number: 69663
HMDB ID: 5369
ReNetm CUI: 5647
DSSTOX substance identifier: 36049
Beilstein Registry Number: 18674
UNII: 58254
Guide to Pharmacology Ligand ID: 5903
Freebase ID: 8940
ChEMBL ID: 49443
PubChem
Guide to Pharmacology
cellular component
3994
Gene Ontology ID: 3994
protein family
20525
InterPro ID: 20319
disease
8694
NDLAuth ID: 266
NCIT Thesaurus: 486
Freebase ID: 1836
ICD-9: 3865
MeSHDer: 1512
MeSH ID: 3427
DiseasesDB: 1711
Orphanet ID: 1178
Disease Ontology ID: 7598
Gwoq topic ID: 653
JSTOR topic ID: 5243
MedlinePlus ID: 917
OMIM ID: 4306
ICD-10: 5127
ICD-O: 236
UMLS CUI: 6435
GWAS Central
biological process
31018
Gene Ontology ID: 31005
supersecondary structure
687
InterPro ID: 687
biological pathway
389
WikiPathways ID: 389
molecular function
9289
Gene Ontology ID: 9269
NDF-RT
medical sign
chemical hazard
689
symptom
scientific article
684415
JSTOR article ID: 8083
NIOSHTIC-2 ID: 18990
PMCID: 341951
PubMed ID: 606062
DOI: 616161
protein domain
9285
InterPro ID: 9282
US pregnancy category
Australian pregnancy category
route of administration
27
anatomical structure
medical specialty
ortholog (282277)
encoded by (1746)
encodes (2665)
encoded by (208648)
has part (1583)
part of (1915)
subclass of (5839)
cell component (256128)
biological process (415964)
molecular function (155828)
part of (7268)
regulates (molecular biology) (6849)
has part (7064)
subclass of (57954)
has part (1687)
has part (15093)
main subject (7690)
drug used for treatment (2699)
symptoms (1367)
symptoms (414)
has cause (580)
main subject (1532)
symptoms (369)
symptoms (2725)
main subject (217)
main subject (2696)
main subject (700)
subclass of (2237)
subclass of (8746)
subclass of (760)
instance of (9269)
has part (7909)
has part (1396)
has part (809)
pregnancy category (484)
pregnancy category (916)
route of administration (1349)
affects (1252)
fluid and measures
medical specialty (2098)
main subject
medical condition
cause of death
discoverer or inventor
mode of inheritance
author
published in
sponsor
cites (3375939)
has part (7812)
part of
has part (5532)
has part (4600)
has part (271)
has part (224)
subclass of (5250)
has part (27987)
has part (4056)
physically interacts with
has part (7909)
has part (1396)
has part (809)
subclass of (2237)
has part (930)
has part (2001)
has part (1583)
part of (1915)
subclass of (5839)
has part (1687)
has part (15093)
has part (7064)
subclass of (57954)
has part (4056)
has part (27987)
has part (809)
has part (1396)
has part (7909)
has part (224)
has part (271)
has part (4600)
has part (5532)
has part (7812)
has part (930)
has part (2001)
has part (1583)
has part (1687)
has part (15093)
has part (7064)
has part (4056)
has part (27987)
has part (809)
has part (1396)
has part (7909)
has part (224)
has part (271)
has part (4600)
has part (5532)
has part (7812)
has part (930)
has part (2001)

Getting data out of Wikidata

- **API**

  For getting data about individual Wikidata items (or groups of up to 50)

- **SPARQL Endpoint**

  Run advanced queries and get back data for up to around 200k items

- **Data Dump**

  Download all available data for large scale local processing of any size

Read more about Wikidata data access →

SPARQL endpoint for querying Wikidata

SPARQL is a programming language for querying ‘linked data’ stored on the web.

Learning SPARQL will allow you to query Wikidata, but also the countless other data sources offering a SPARQL query service.

Timeline of people educated at the University of St. Andrews (View as Wikidata query)

- **People born on this day →**
- **List of countries ordered by the number of their cities with female mayor →**
- **Children of Genghis Khan →**
- **Map of female writers around the world by birthplace**

Useful links

Wikidata Query Service Beta – Official query service

SPARQL Query Examples

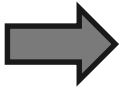
WDQ – Third party query service which is simple to learn

Video tutorial 1 – Getting started with SPARQL queries.

Video tutorial 2 – SPARQL query workshop.

Simple data retrieval – Try these example queries

“Retrieve genes with genetic association with asthma” (According to GWAS).



39 genes

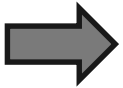
gene	geneLabel		gene	geneLabel		gene	geneLabel		gene	geneLabel
Q5013317	COL22A1		Q18027370	IGSF3		Q18053559	CDHR3		Q14903974	SMAD3
Q14912759	SLC22A5		Q18045382	HPSE2		Q18045669	ATG3		Q18033889	IL1RL1
Q14914243	PSAP		Q18048437	IL33		Q18035037	RAD50		Q17917202	ERBB4
Q14907990	SLC30A8		Q18051900	PYHIN1		Q18036984	FBXL7		Q18027836	IL6R
Q18025002	GAB1		Q17709208	ACO1		Q18033919	XPR1		Q18030185	NOTCH4
Q18035589	C6orf10		Q18027822	IL2RB		Q15326496	RORA		Q18030409	PDE4D
Q18054256	GSDMA		Q18030364	PBX2		Q18042132	GSDMB		Q18045645	IKZF4
Q18058487	C5orf56		Q18037773	ABI3BP		Q18029145	MKLN1		Q18039979	KLHL5
Q18030785	PRKG1		Q18039623	CTNNA3		Q18036729	RAP1GAP2		Q18026947	HLA-DQA1
Q18033424	IL18R1		Q18046350	ZNF665		Q14878303	IL13			

```
SELECT DISTINCT ?gene ?geneLabel where {
  ?gene wdt:P2293 wd:Q35869 .  # gene has genetic association to "asthma"
  ?gene wdt:P31 wd:Q7187 .     # gene is subclass of "gene"
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}
```

http://bit.ly/bosc2017_wikidata

Data integration

"Retrieve genes with GWAS association with asthma **and gene product is localized to membrane**"



22 genes

gene	geneLabel		gene	geneLabel		gene	geneLabel		gene	geneLabel
Q14912759	SLC22A5		Q18027370	IGSF3		Q18035037	RAD50		Q18027836	IL6R
Q14914243	PSAP		Q18033424	IL18R1		Q18033919	XPR1		Q18030409	PDE4D
Q14907990	SLC30A8		Q18045382	HPSE2		Q18042132	GSDMB		Q18030185	NOTCH4
Q18035589	C6orf10		Q18027822	IL2RB		Q18036729	RAP1GAP2		Q18026947	HLA-DQA1

```
SELECT DISTINCT ?gene ?geneLabel where {
  ?gene wdt:P2293 wd:Q35869 .  # gene has genetic association to "asthma"

  ?gene wdt:P31 wd:Q7187 .     # gene is subclass of "gene"

  ?gene wdt:P688 ?protein .                # gene encodes a protein
  ?protein wdt:P681 ?cc .                  # protein has a cellular component
  ?cc wdt:P279*|wdt:P361* wd:Q14349455 .   # cell component is 'part of' or 'subclass of' membrane

  SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}
```

http://bit.ly/bosc2017_wikidata

Leveraging the Disease Ontology structure

"Retrieve genes with GWAS association with **any respiratory disease** and gene product is localized to membrane (non-IEA)"

31 genes / 8 diseases

diseaseGALabel	gene_counts	geneList
asthma	15	SMAD3, RAP1GAP2, IL18R1, HPSE2, SLC30A8, SLC22A5, PSAP, ERBB4, HLA-DQA1, IGSF3, IL2RB, IL6R, NOTCH4, PDE4D, RAD50
chronic obstructive pulmonary disease	5	HLA-C, SFTPD, ANXA5, ANXA11, ATP2C2
lung cancer	3	TGM5, VTI1A, PHACTR2
interstitial lung disease	2	DSP, ATP11A
non-small-cell lung carcinoma	2	NALCN, DLST
nasopharynx carcinoma	2	ITGA9, TNFRSF19
adenocarcinoma of the lung	1	BTNL2
pulmonary emphysema	1	BICD1

```
SELECT ?diseaseGALabel (count (DISTINCT ?
(group_concat(DISTINCT ?geneLabel; separa
  ?gene wdt:P2293 ?diseaseGA .          # g
  ?diseaseGA wdt:P279* wd:Q3286546 .  # to a type of respiratory system disease

  ?gene wdt:P31 wd:Q7187 ; wdt:P688 ?protein ;    # gene is subclass of "gene" and encodes protein
        rdfs:label ?geneLabel .
  FILTER (lang(?geneLabel) = "en")
  ?protein p:P681 ?s .                            # protein's cell component statement
    ?s ps:P681 ?cp .                              # get statement value
```

http://bit.ly/bosc2017_wikidata

Opportunistic integration

4 diseases / 6 chemical hazards

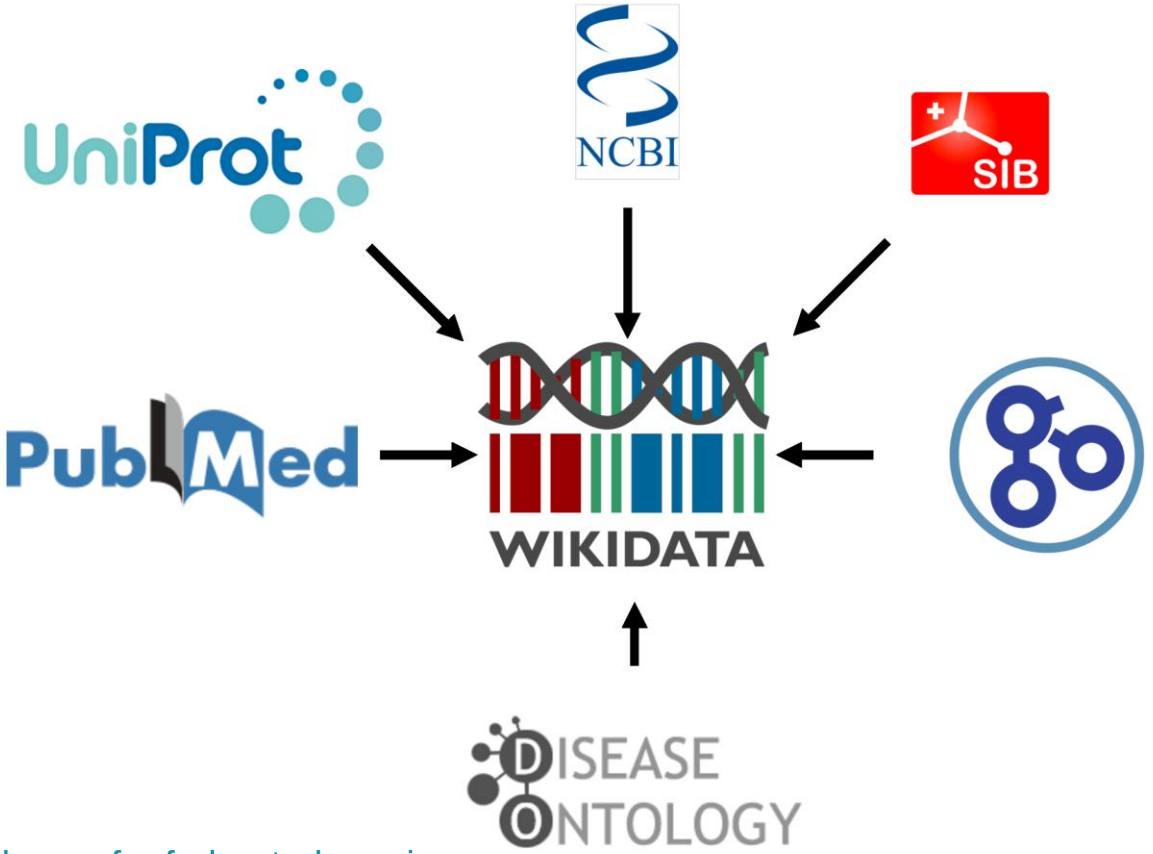
"Retrieve genes with GWAS association with any respiratory disease and gene product is localized to membrane (non-IEA) **and show causative chemical hazards"**

diseaseGALabel	exposureLabel
lung cancer	arsenic pentoxide exposure
lung cancer	HN1 exposure
lung cancer	mechlorethamine exposure
lung cancer	HN3 exposure
asthma	Phenacyl chloride exposure
pulmonary emphysema	phosgene exposure

```
12
13   ?exposure wdt:P1542 ?diseaseGA .  # something causes disease
14   ?exposure wdt:P279 wd:Q21167512 . # and that something is a chemical hazard
15
16   SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
17 }
```

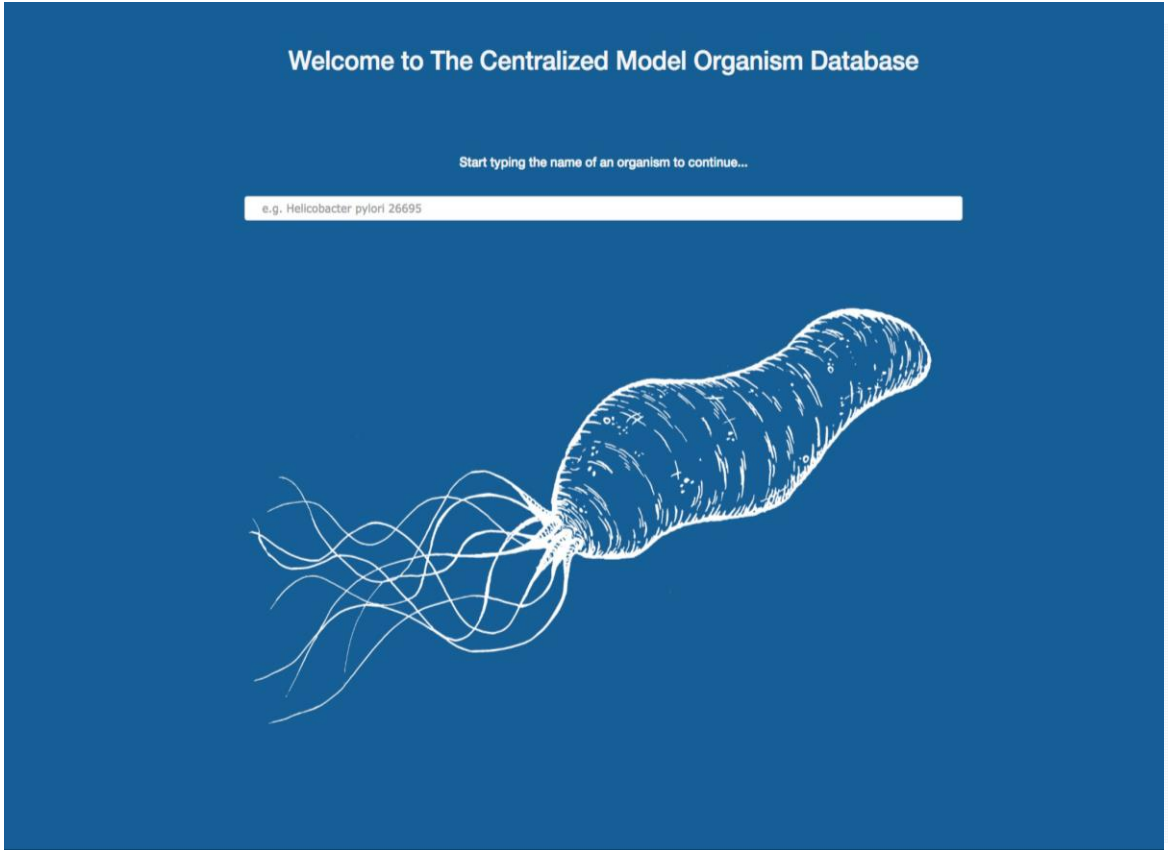
http://bit.ly/bosc2017_wikidata

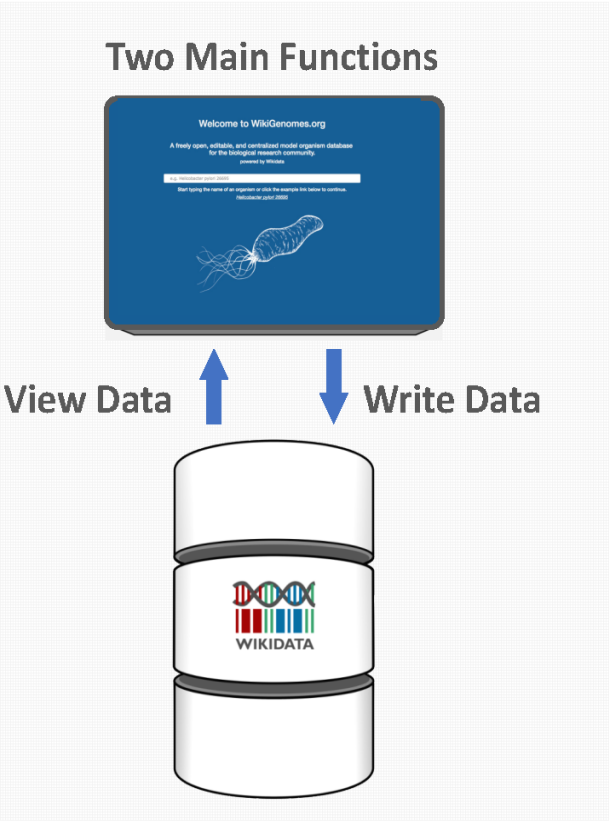
Centralizing and Linking the Data



Federated SPARQL queries
Currently allowed external databases for federated queries.

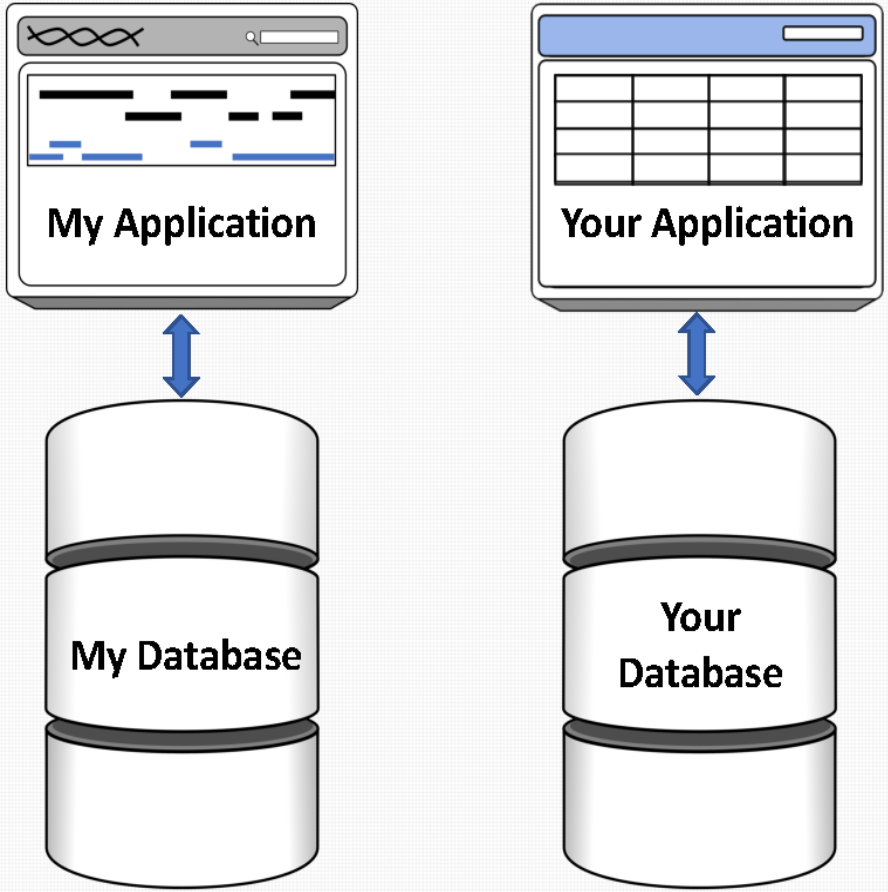
http://wikigenomes.org/





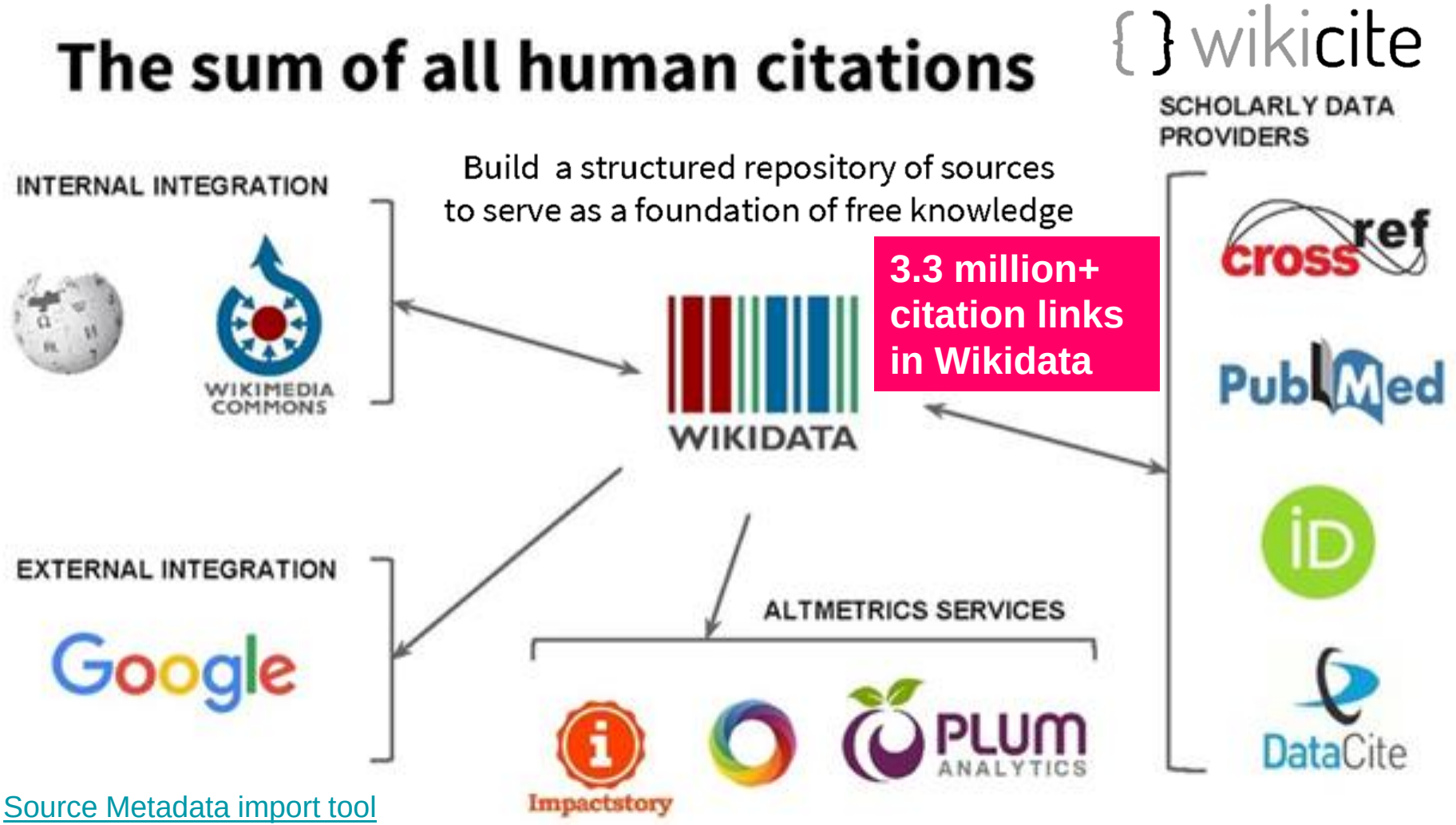
WikiGenomes: an open Web application for community consumption and curation of gene annotation data in Wikidata. – Article on Biorxiv.org

The Way it's Always Been

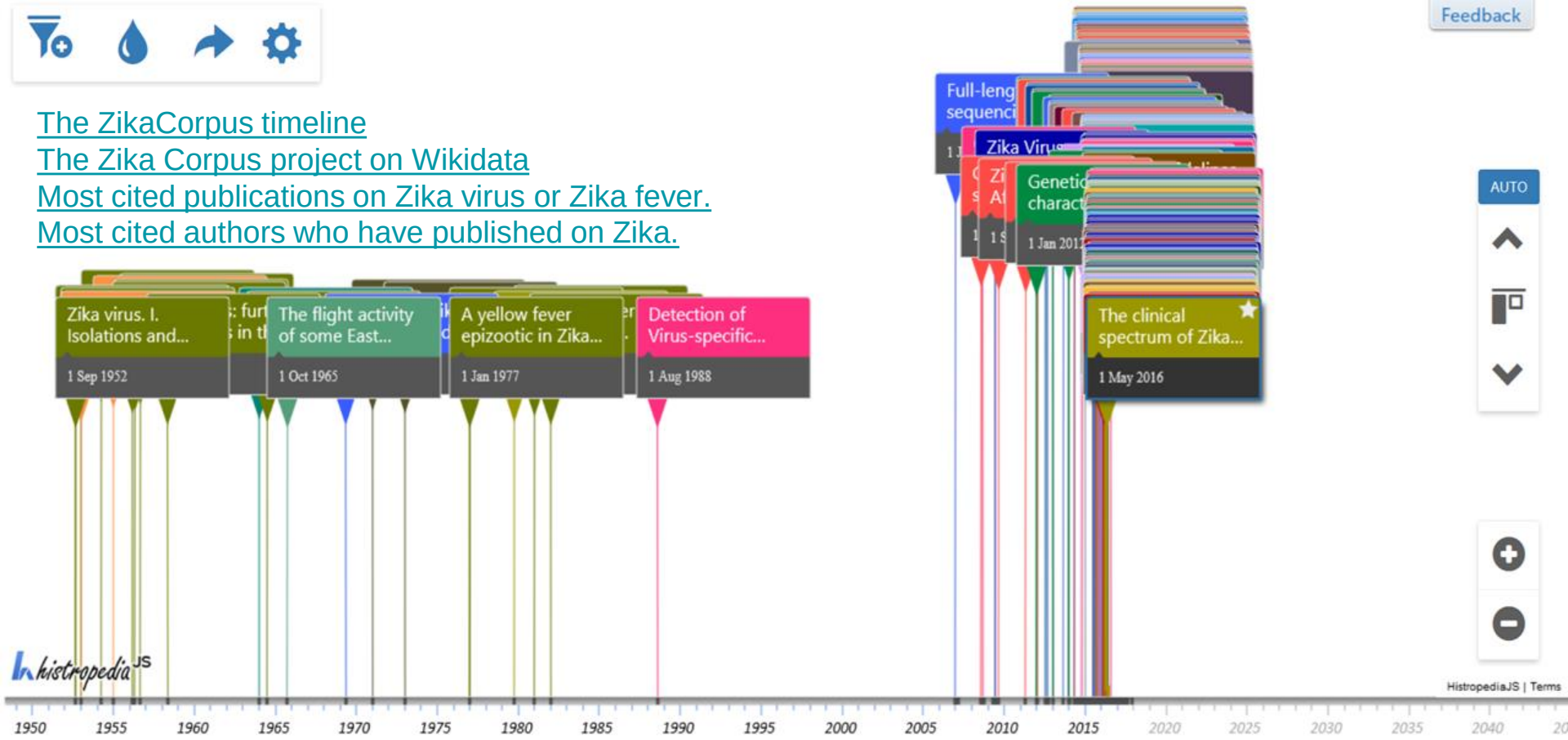


The sum of all human citations

{ } wikicite



Source Metadata import tool



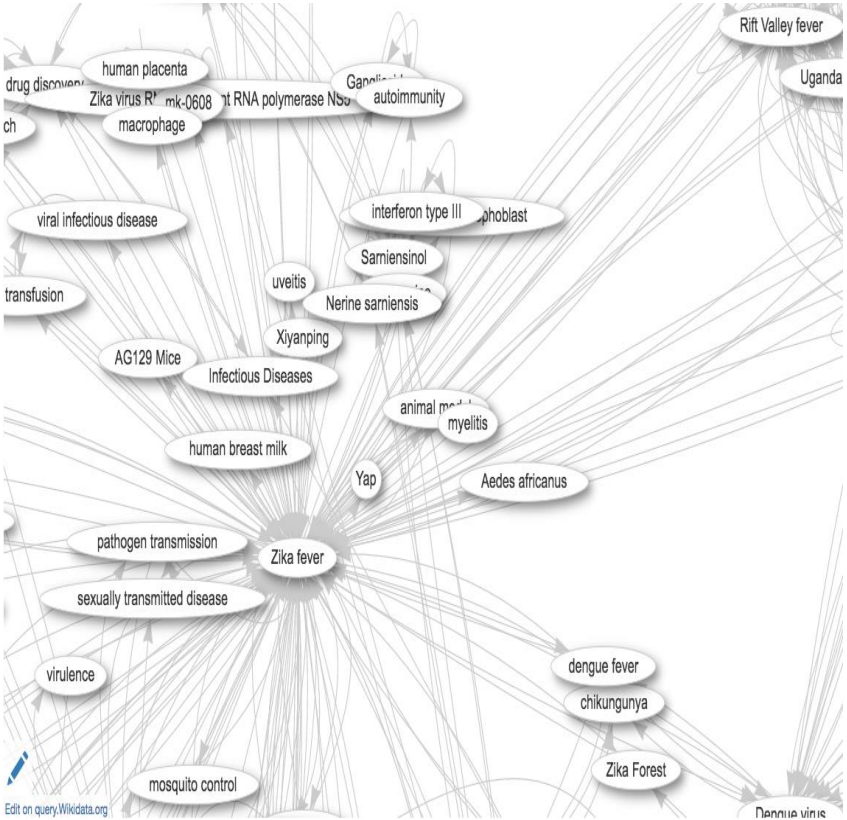
The Zika Corpus – Histropedia timeline of scientific articles by publication date, colour-coded by publication and filtered by author and original language of work. In February 2016, the World Health Organization declared a public health emergency over the Zika virus outbreak

Scholia - 2.3 million scientific articles in Wikidata

The Scholia Web service creates on-the-fly scholarly profiles for researchers, organizations, journals, publishers, individual scholarly works, and for research topics.

Among several display formats available are lists of publications for individual researchers and organizations, publications per year, employment timelines, co-author networks and citation graphs.

Example, Blog article + Video presentation
Paper on arxiv.org by Finn Årup Nielsen.



Wikidata: enabling discovery

How can I find reference materials about Jane Austen?

Project Gutenberg, Librivox, Find A Grave, WikiQuote, Kindred Britain (family tree), 103 different language Wikipedias, The Huygens database of Won

- “Jane Austen”
- “Austen, Jane”
- “jane-austen”
- “Аустен, Джен”
- ~~“Jane G. Austin”~~

Authority control	VIAF: 102333412 · LCCN: n79032879 · ISNI: 0000 0001 2283 635X · GND: 118505173 · SELIBR: 207420 · SUDOC: 02669719X · BNF: cb118896036 · ULAN: 500249665 · NLA: 35010277 · NDL: 00431922 · NKC: jn19990000321 · SBN: IT\ICCU\CFIV\017136 · DBNL: aust003 · PTBNP: 8531 · NTA: 068721722 · BAV: ADV10179859 · NUKAT: n93080119 · IBDB: 6642 · ISFDB: 31095 · LibriVox: 155 · Project Gutenberg: 68 · Find a Grave: 44 · BNE: XX1124986 · NLR: RUNLRAUTH770143336 · Open Library: OL21594A · IMDB: nm0000807 · Freebase: /m/040dv · ODNB: 101000904 · GEC: 0006187 · English Wikisource: 941 · WorldCat

Jane Austen (Q36322)

Enabling discoverability - Doctoral Thesis Metadata

Oxford Research Archive has **3237** Oxford doctoral theses on open access for anyone to download and read.

ORA are sharing their doctoral thesis metadata with Wikidata (title, author, subject, publisher, links to WP)

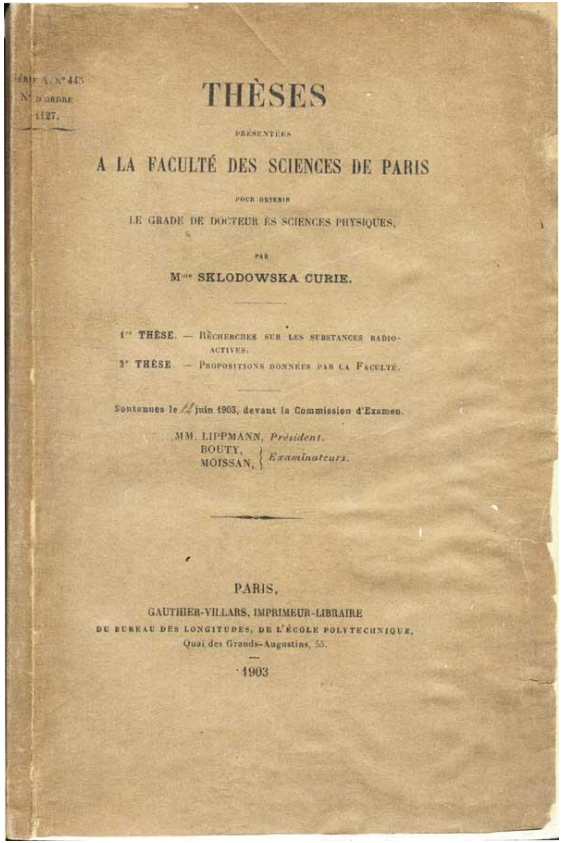
Query showing all doctoral theses on Wikidata.

New property: P4101 - *Dissertation submitted to*

How Wikidata links the Oxford theses - query result

And the query itself.

See also: WikiProject:Source Metadata



THÈSES

PRÉSENTÉES

A LA FACULTÉ DES SCIENCES DE PARIS

POUR OBTENIR

LE GRADE DE DOCTEUR ÈS SCIENCES PHYSIQUES,

PAR

Mme SKLODOWSKA CURIE.

1re THÈSE. — RECHERCHES SUR LES SUBSTANCES RADIO-ACTIVES.

2e THÈSE. — PROPOSITIONS DONNÉES PAR LA FACULTÉ.

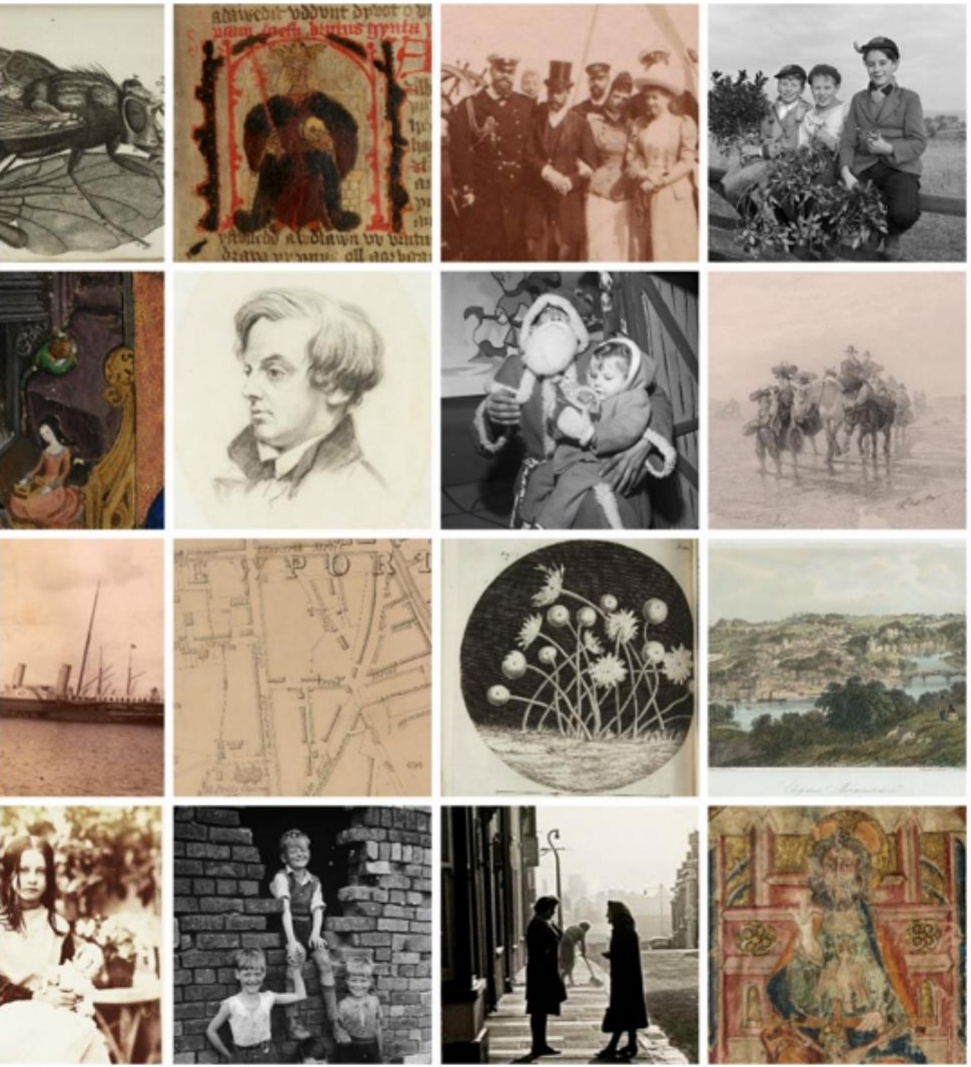
Soutenues le 14 juin 1903, devant la Commission d'Examen.

MM. LIPPMANN, *Président.*
BOUTY, MOISSAN, *Examinateurs.*

PARIS,

GAUTHIER-VILLARS, IMPRIMEUR-LIBRAIRE
DU BUREAU DES LONGITUDES, DE L'ÉCOLE POLYTECHNIQUE,
Quai des Grands-Augustins, 55.

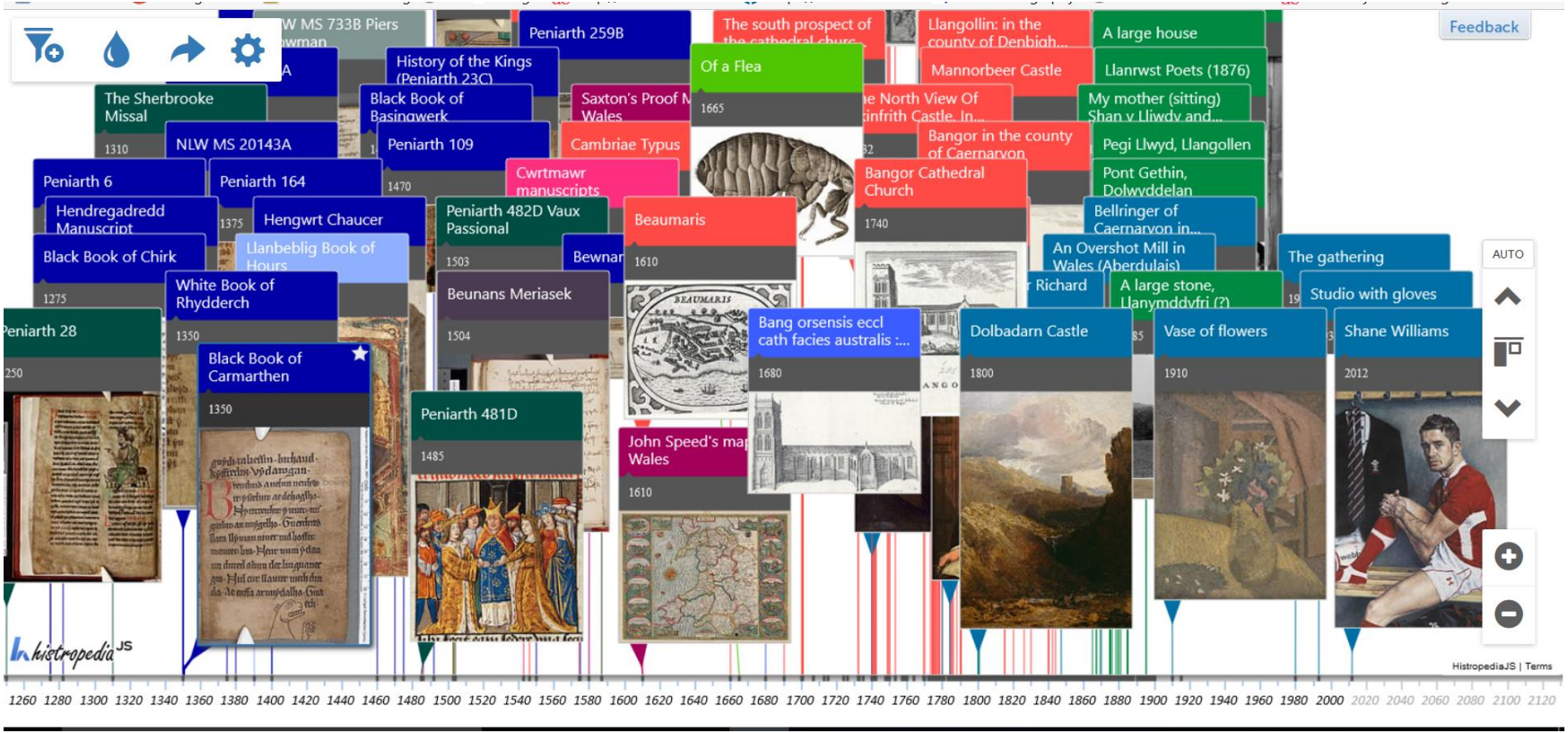
1903



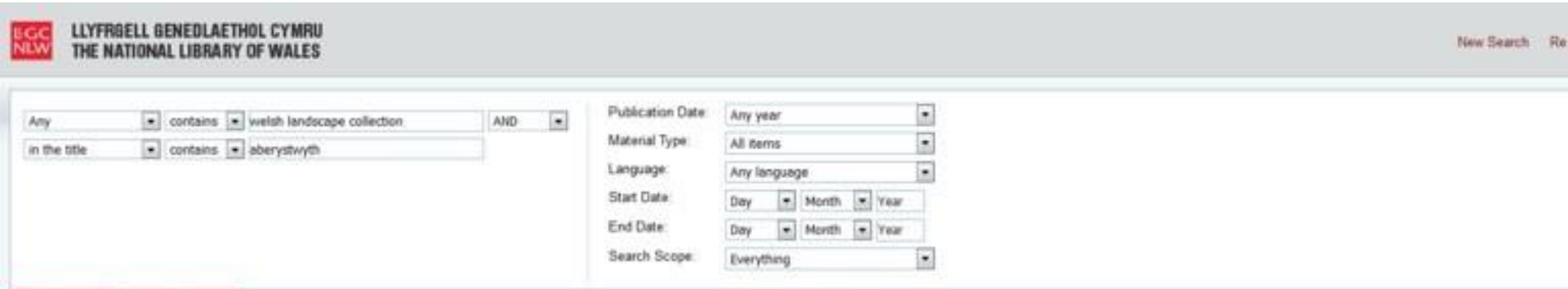
265 Million Views on Wikipedia

National Library of Wales ID – (P2966)

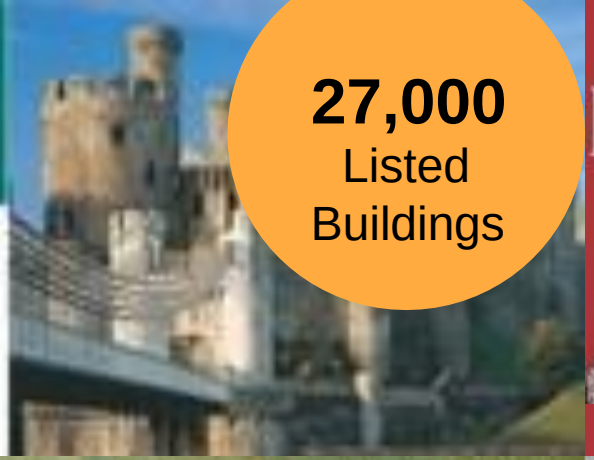
http://tinyurl.com/NLWtimeline



Accelerate the *discoverability*, *reusability*, and *societal impact* of open access



Cadw



27,000
Listed Buildings

Dictionary of Welsh Biography

4500
People

500
Merchant ships

4500
Welsh Landscapes

Edith Eleanor (Q24027483)

merchant ship

Other properties

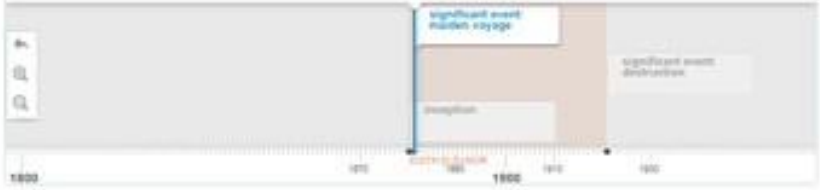
location of final assembly	Aberystwyth town in Ceredigion, Wales
described by source	Aberystwyth shipping records collection of shipping records
manufacturer	Warrell and Co. ship builders
length	80.4 (80.3–80.5)
width	21.3 (21.2–21.4)
mass	105 ± 1
instance of	merchant vessel boat or ship that transports cargo or carries passengers for hire
	topsail schooner ship type
catalog code	85261 catalog : Aberystwyth shipping records [NL]
port of registry	Aberystwyth town in Ceredigion, Wales
inception	1880
significant event	maiden voyage first journey made by a craft point in time : 1881 [NL]
	destruction concept of damage to an object, system or an idea point in time : 1921 [NL] location : Bristol Channel [NL] has cause : shipwrecking [NL]

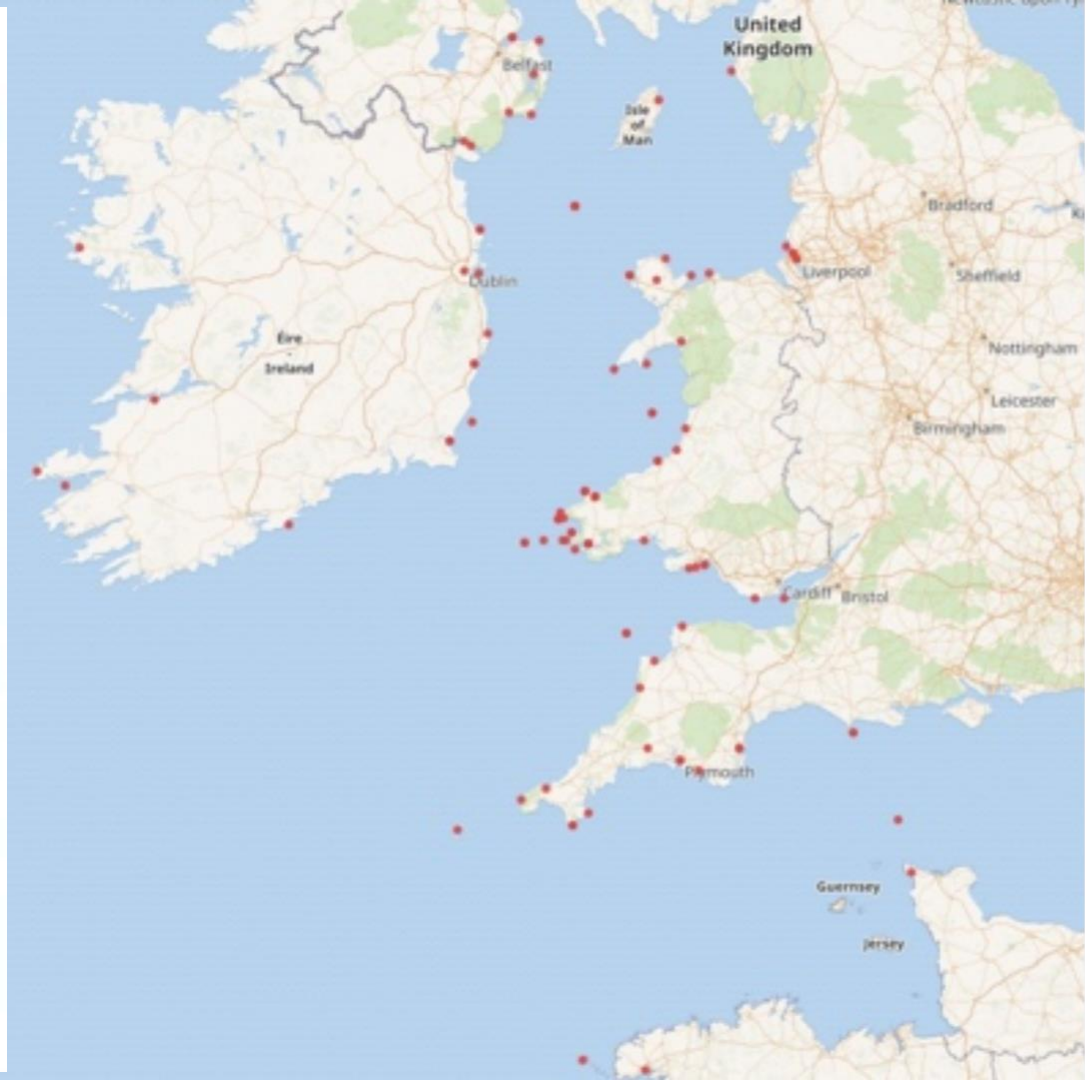
Timeline

1881

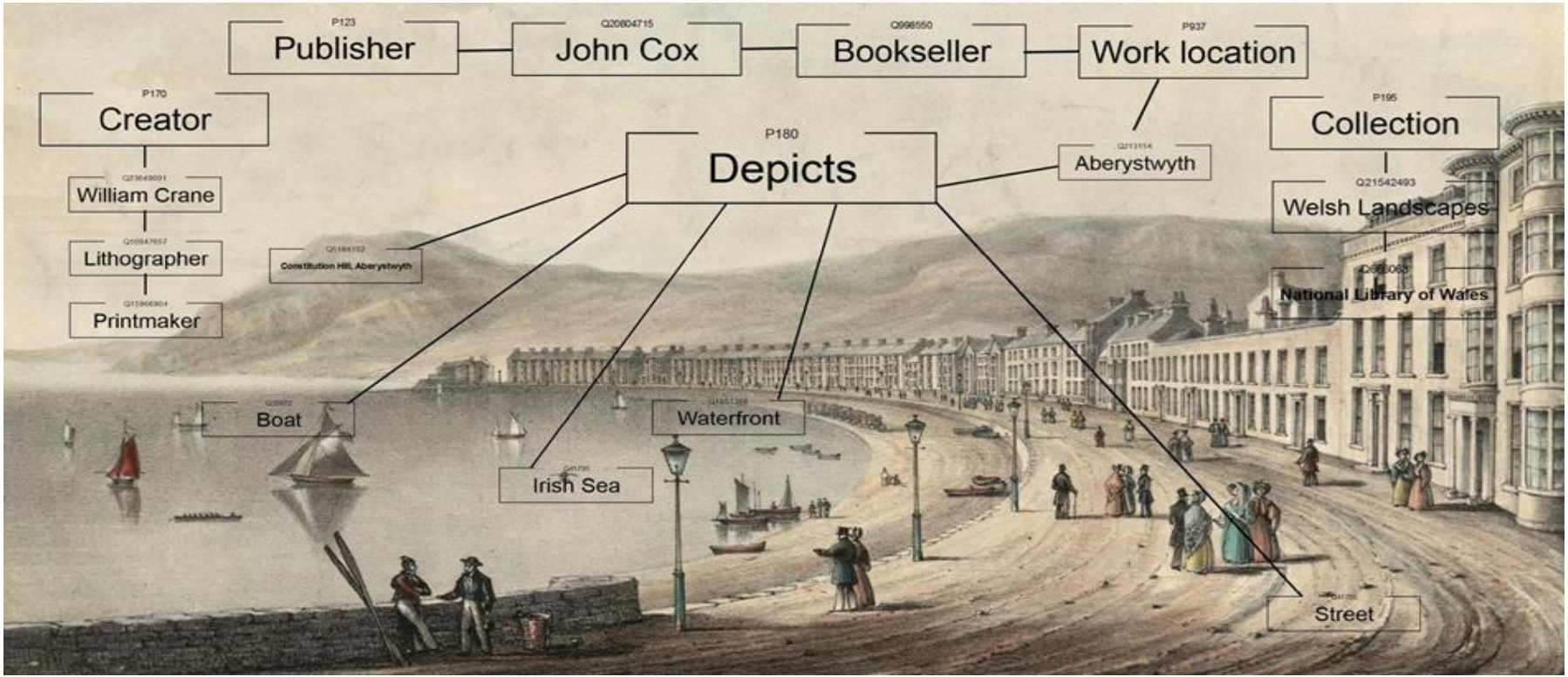
significant event: maiden voyage

first journey made by a craft





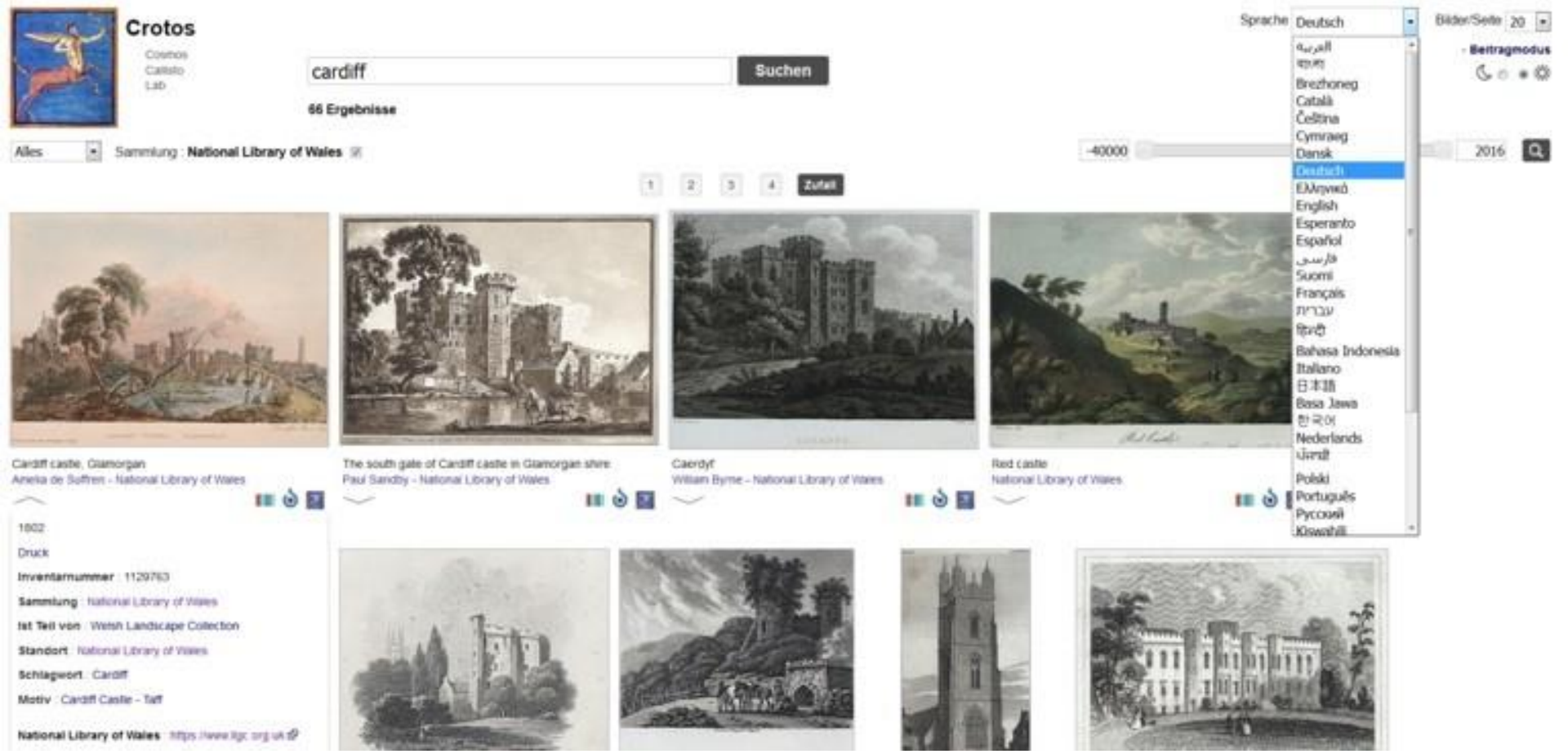
National Library of Wales



Timeline of NLW collection works

Link to Crotos

Sum of all paintings project Lists.

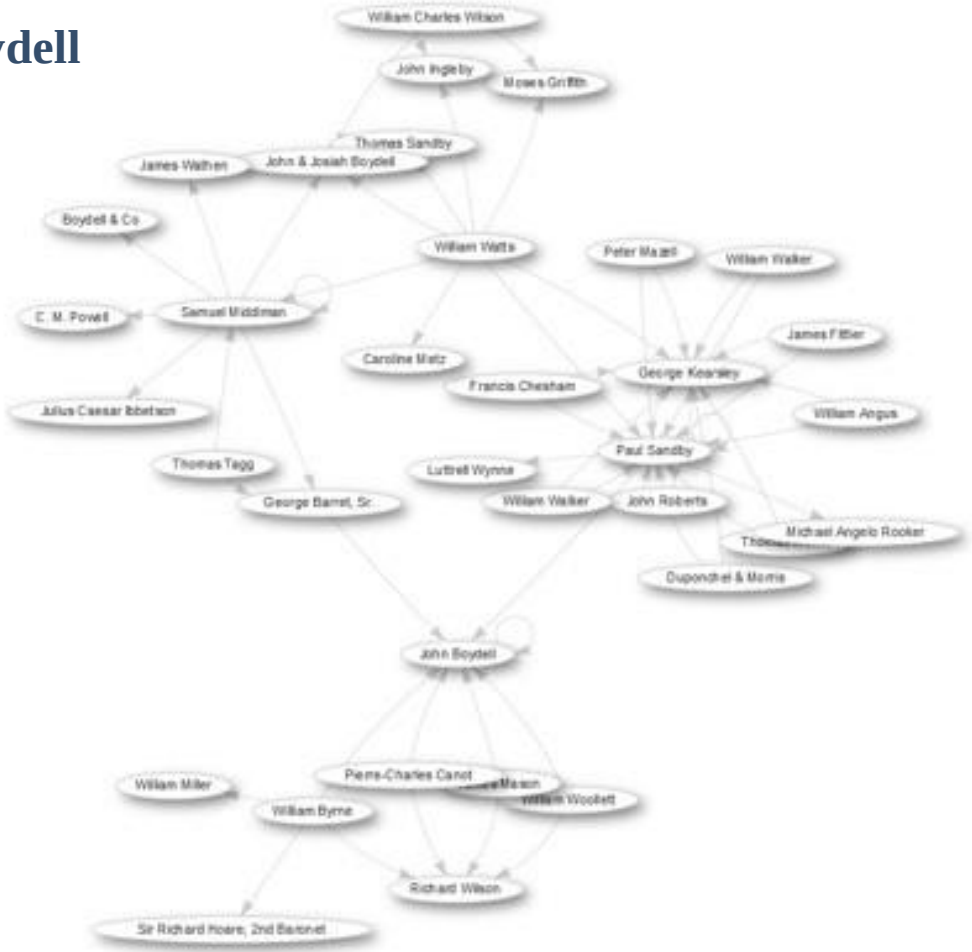


[1. Images with boats]

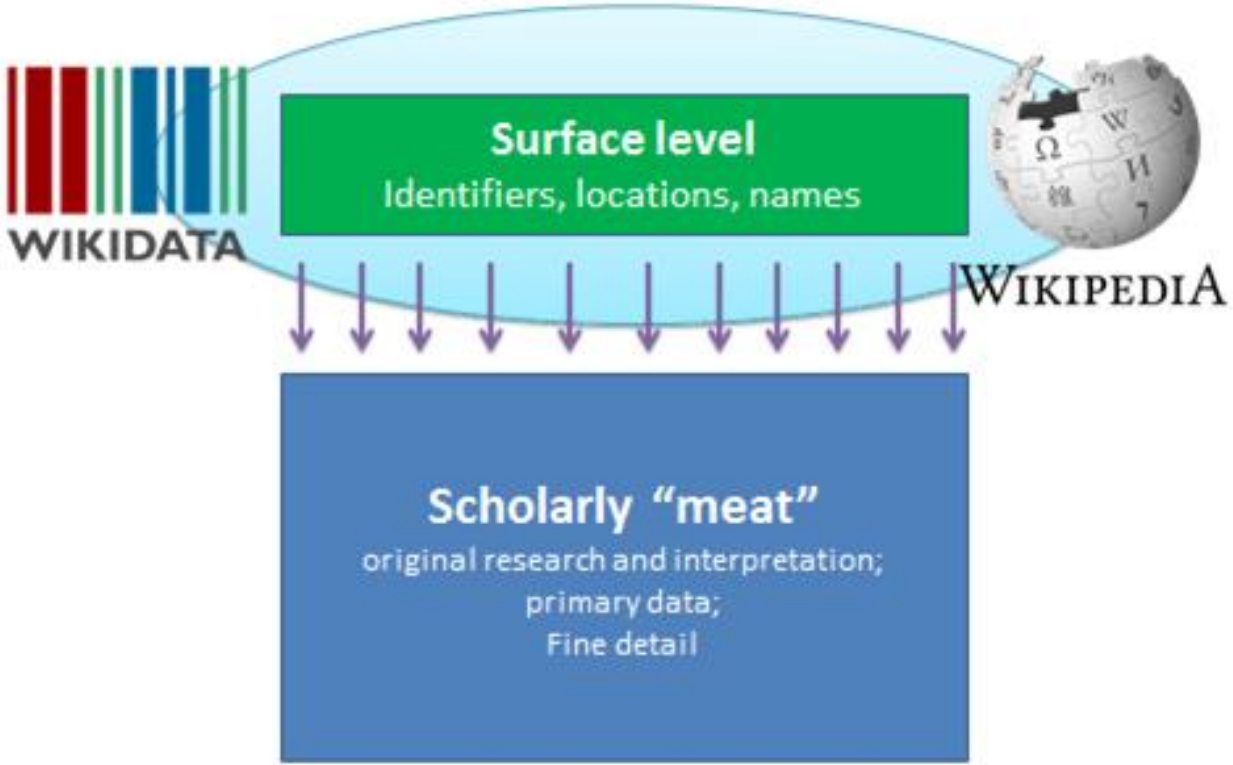
[2. Images of Carnarvon Castle]

[3. Filtering to just the images featuring Eagle Tower at Carnarvon Castle]

Links between John Boydell and other Printers/publishers & engravers



Why contribute to another repository?
Enrich both repositories by combining datasets.



Wikidata

The free and open knowledge base

Thanks for listening!
Any questions?

ewan.mcandrew@ed.ac.uk

Federated queries

Run queries that combine data from Wikidata and other selected data sources on the web.

List of 3rd party services supported for federated queries inc. Europeana, Bibliothèque nationale de France, Biblioteca Virtual Miguel de Cervantes and UK Office for National Statistics among others.

Simple example federated query:

Works by Lope de Vega, retrieved from the BVMC digital library

1. Lope de Vega's unique BVMC id is determined from Wikidata
2. This id is then used to retrieve works by Lope de Vega on the BVMC digital library

Links and further reading

- https://www.wikidata.org/wiki/Wikidata:Data_Import_Guide
- https://www.wikidata.org/wiki/Wikidata:Database_download
- https://www.wikidata.org/wiki/Property:P2966 – National Library of Wales ID.
- https://www.wikidata.org/w/api.php
- https://tools.wmflabs.org/reasonator/ - Wikidata made ‘pretty’
- https://tools.wmflabs.org/reasonator/?q=Q42 – Reasonator page for Douglas Adams (Q42) by way of example.
- https://www.mediawiki.org/wiki/Extension:ArticlePlaceholder – another way of placeholding articles using structured data from wikidata to populate information in the meantime until an article can be created.
- Wikidata: Current trends and priorities (May 2017 presentation with current stats)
- Wikidata video presentations on Media Hopper.

Developer links

#wikidata on chat.freenode.net

wikidata-l@lists.wikimedia.org

Wikidata – The New Rosetta Stone (article).

Google closes Freebase (article).

Google's sketchy attempt to control the world's knowledge (article).

api @ **wikidata.org/w/api.php**

sandbox @ **wikidata.org/wiki/Special:ApiSandbox**

The Wikidata Game: https://tools.wmflabs.org/wikidata-game/distributed/

PHP

Wikibase API Library: github.com/addwiki/wikibase-api

SPARQL abstraction: github.com/Benestar/asparagus

Python

Wiki bot Framework: mediawiki.org/wiki/Manual:Pywikibot/Wikidata

C# .NET

Wikibase API Library: github.com/Benestar/wikibase.net

Practical - Editing a query

Step 1: Load the sample query: http://tinyurl.com/ycxw4eyw

Step 2: Modify the query to find a different set of results, by:

- Changing values
- Changing properties
- Removing lines

Step 3: Share your query on Twitter and/or add to etherpad!

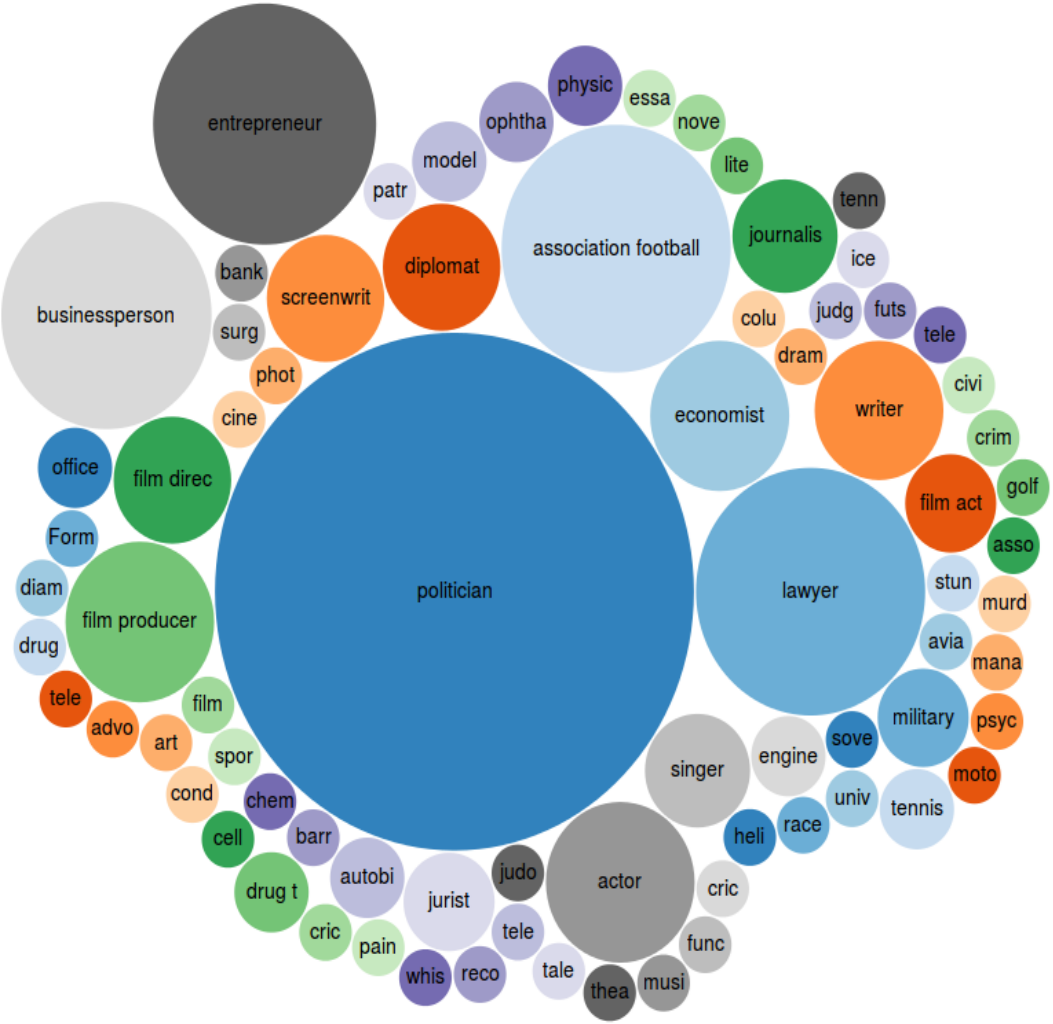
Panama Papers

P106: occupation

P793: significant event

Q23702848: Panama Papers

https://en.wikipedia.org/wiki/User:Fnielsen/Autolists/Panama_Papers

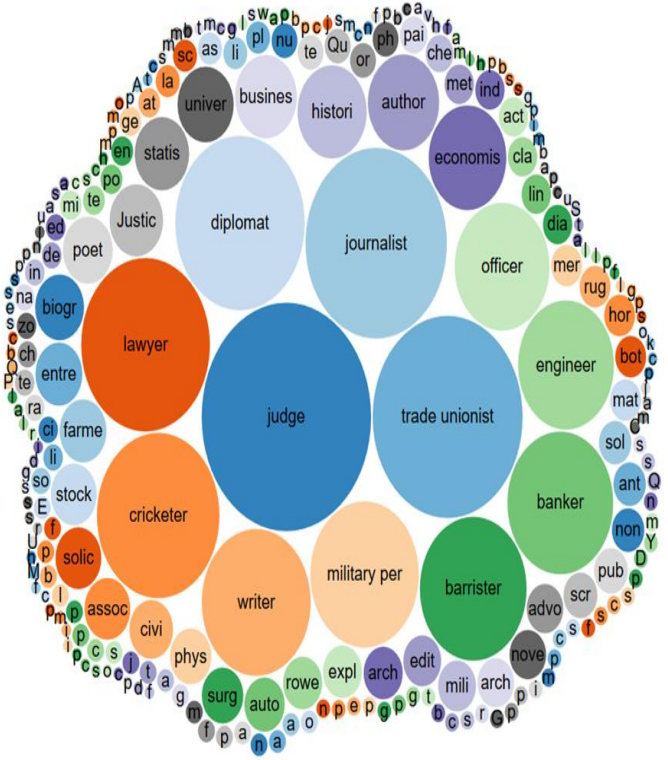


MPs’ occupations and place of education.

Link to Wikidata query - occupation.

Link to Wikidata query - education.

Image of Ken Clarke by Chris McAndrew (CC-BY)



Other notable examples of use cases

YLE - The Finnish Broadcasting Company, Yle, has since April 1st 2016 tagged online news and feature articles with concepts from Wikidata.

Inventaire - Create an inventory of your books with Wikidata at inventaire.io

WikiGenomes - A freely open, editable, and centralized model organism database for the biological research community. Paper on WikiGenomes at Biorxiv.org

Quora - Links to Quora topics will be available through the Wikidata entities and also from Quora topic pages to Wikidata entities.

Crotos - search and display engine for visual artworks powered by **Wikidata**.

And much more besides.

https://wikimediafoundation.org/wiki/Our_projects

English Wikipedia *"only includes only includes 30% of the items entered in the other 287 languages."**

Language	Language (local)	Wiki	Articles	Total	Edits	Admins	Users	Active Users	Images	Depth
English	English	en	5,205,057	39,859,108	841,187,469	1,297	28,749,525	112,301	851,064	935.45
Swedish	svenska	sv	3,361,100	6,752,452	36,314,210	71	500,503	2,223	0	5.48
Cebuano	Cebuano	ceb	2,668,059	4,872,925	9,794,491	5	28,432	107	66	1.37
German	Deutsch	de	1,958,190	5,667,148	161,040,878	236	2,446,855	18,341	131,917	101.94
Dutch	Nederlands	nl	1,874,463	3,633,438	48,119,572	49	778,155	3,901	19	11.66
French	français	fr	1,772,894	8,183,494	130,180,351	167	2,558,409	14,397	49,111	207.99

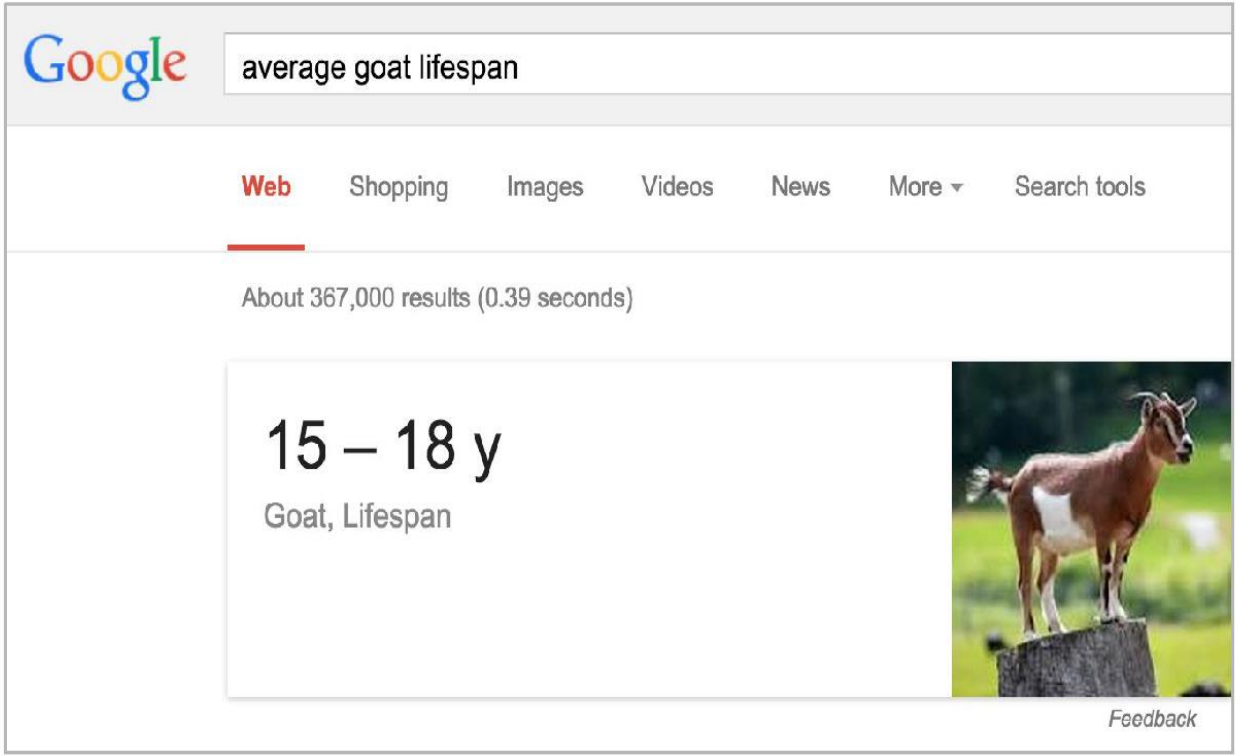
Grand total [edit | edit source]

Articles	Total	Edits	Admins	Users	Active Users	Images
40,742,921	151,823,044	2,214,808,661	3,998	62,976,506	263,618	2,410,128

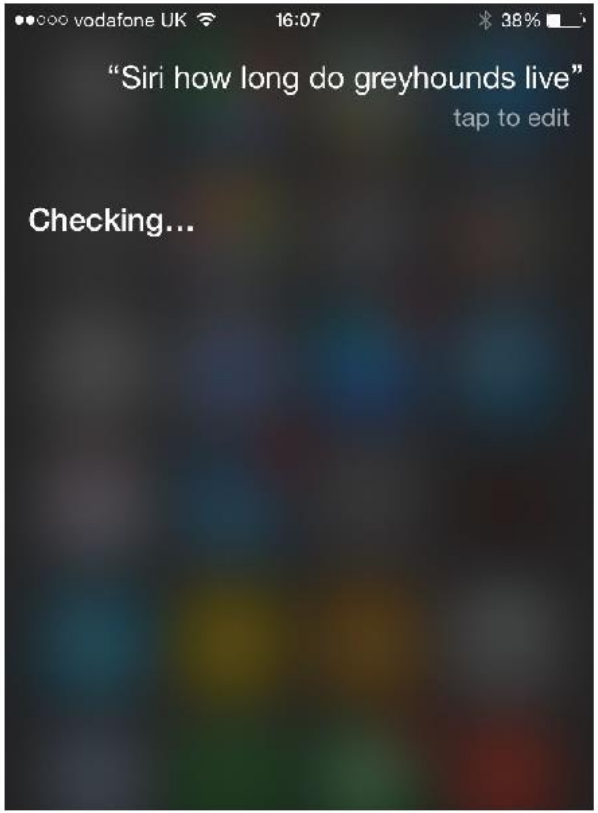
Vietnamese	Tiếng Việt	vi	1,146,809	3,211,773	24,059,192	23	501,251	1,400	21,222	24.29
Japanese	日本語	ja	1,022,325	2,966,811	61,403,824	49	1,109,224	11,787	83,431	74.88
Portuguese	português	pt	928,143	4,135,001	47,141,295	70	1,793,088	6,793	38,558	136.1

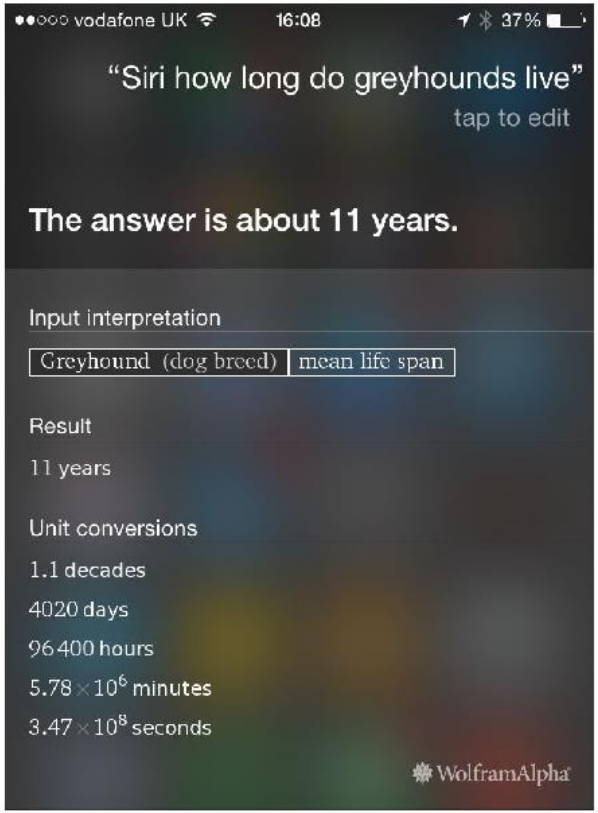
"The promise of linked open data seems to have finally arrived."

Okay Google….. what's the average lifespan of a goat?

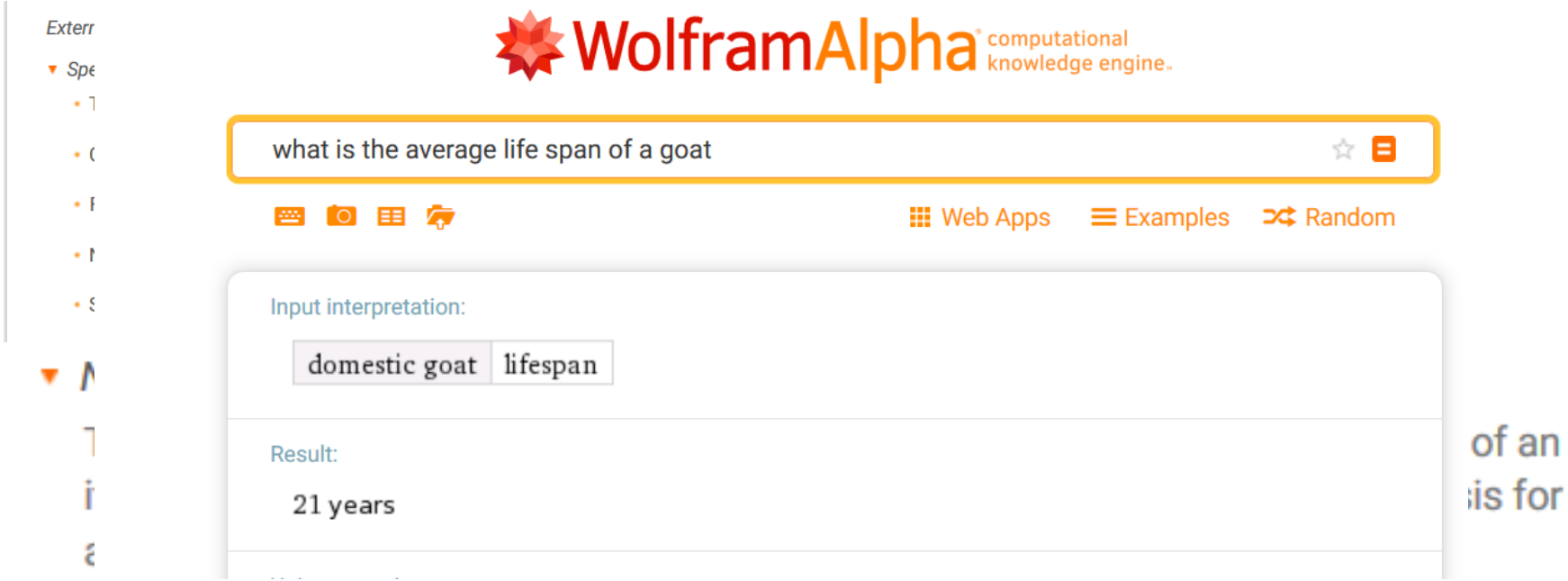


Seldom does the no.1 search engine display the provenance of where the 'suggested answers' it provides come from, undermining "*people's ability to verify information and, ultimately, to develop well-informed opinions.*"(Dewey, 2016)



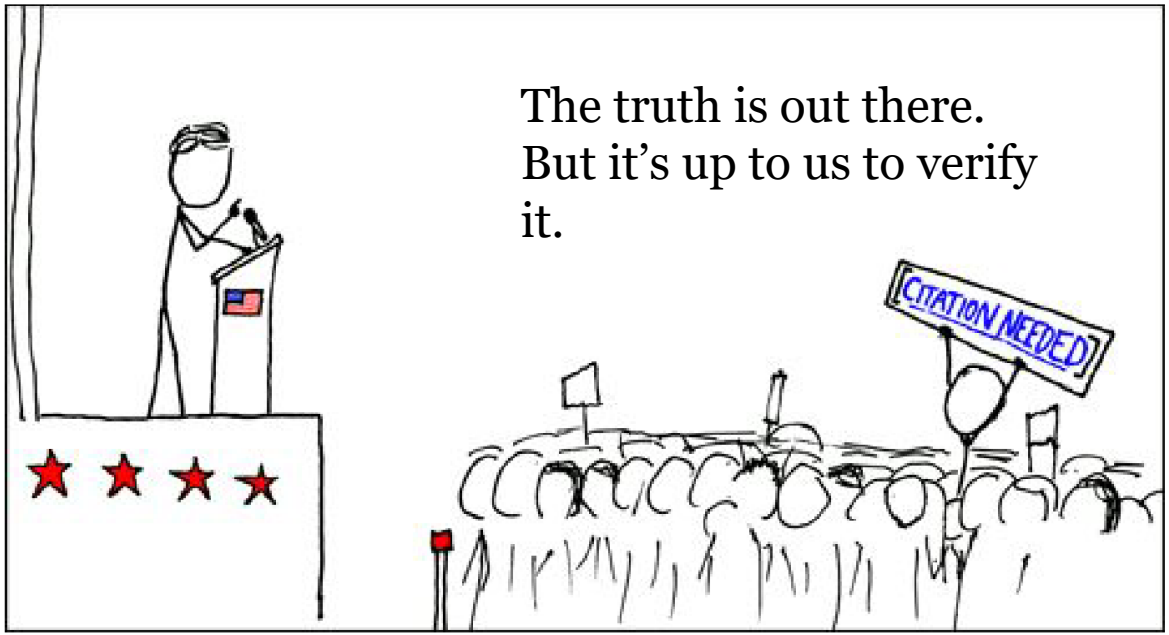


“Search is the way we now live.” (Darnton in Hillis, Petit & Jarrett, 2012)



Requests by researchers for detailed information on the sources for individual Wolfram|Alpha results can be directed here. Requests will handled on the basis of priority, importance and available resources.

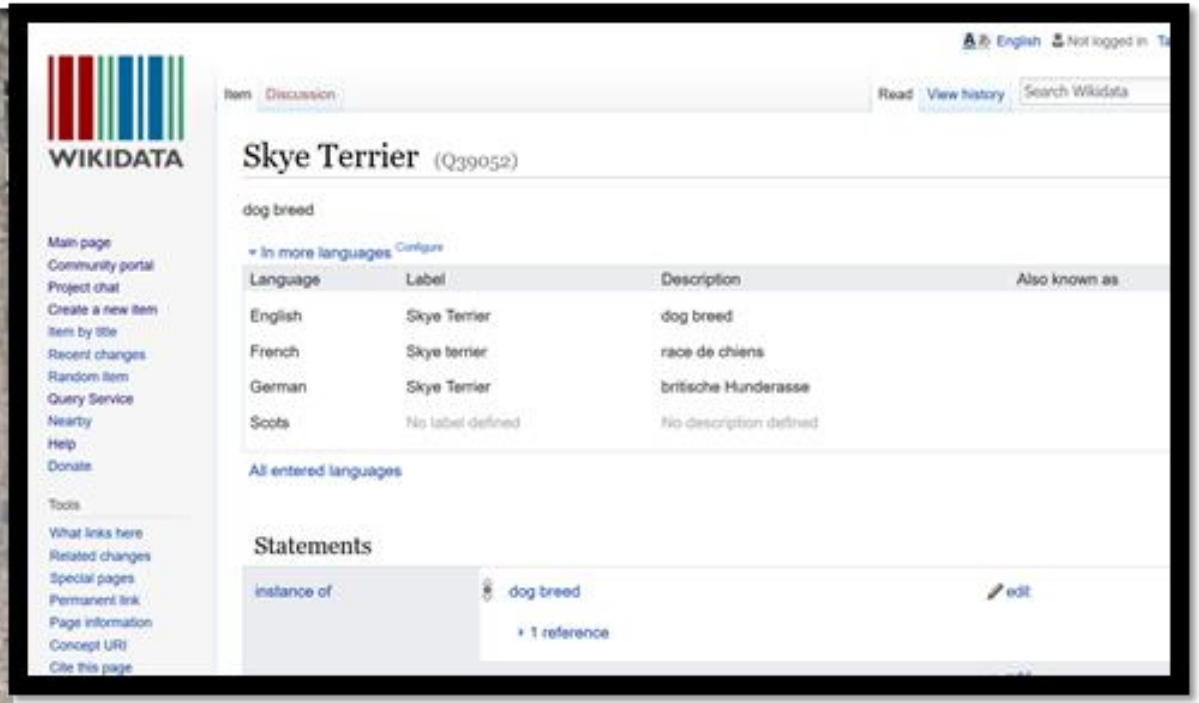
Fact checking sites are springing up in greater numbers but convenience and cognitive ease have proven to be the drivers for the virtual generation.



Randall Munroe, *Wikipedian protester* http://tinyurl.com/p3rodlb [CC BY]

A bibliographic database to serve the sum of all knowledge?

{ } wikicite



WIKIDATA

Item Discussion Read View history Search Wikidata

Skye Terrier (Q39052)

dog breed

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

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Concept URI
Cite this page

In more languages Configure

Language	Label	Description	Also known as
English	Skye Terrier	dog breed	
French	Skye terrier	race de chiens	
German	Skye Terrier	britische Hunderasse	
Scots	No label defined	No description defined	

All entered languages

Statements

instance of — dog breed — edit

1 reference

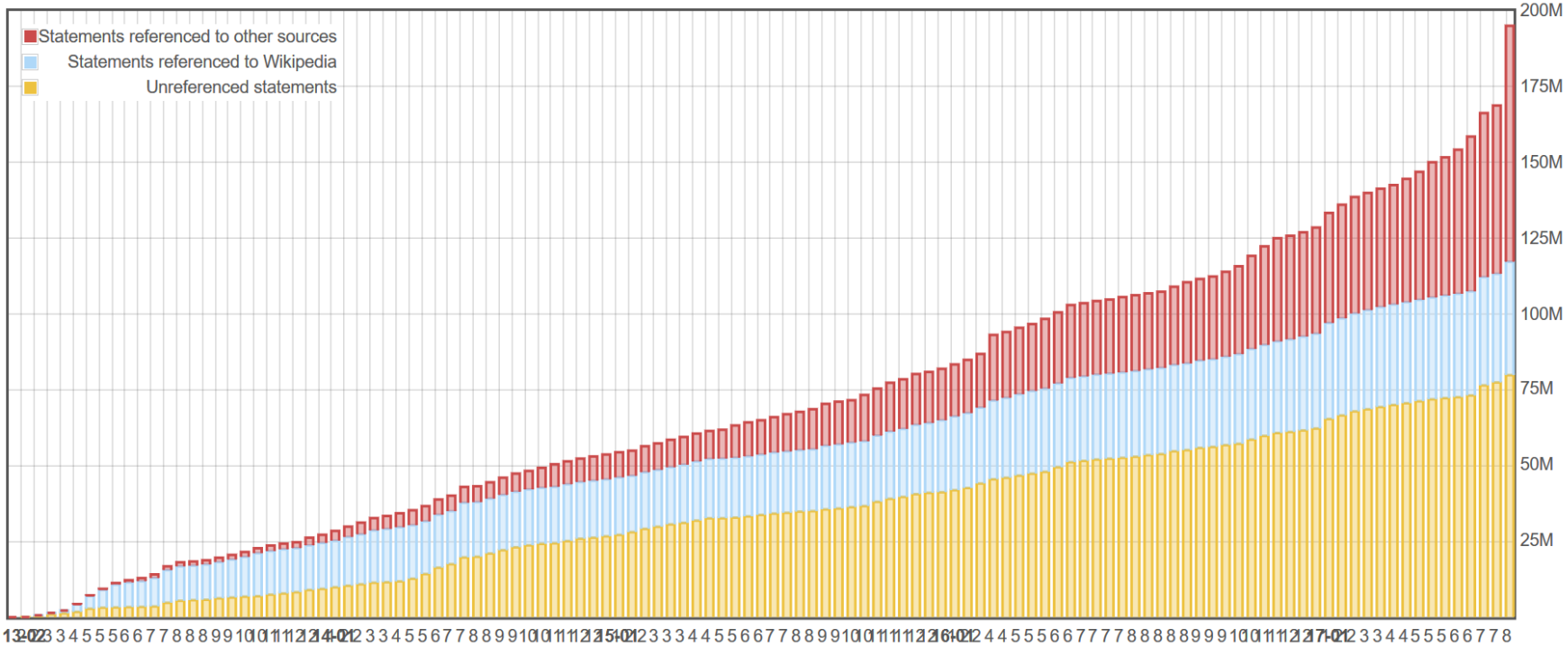
Greyfriar's Bobby (1855-1872)

Sat by his master's grave for 14 years.

Died aged 16.

Kennel club say Skye Terrier's have a life expectancy of "over 12 years".

References in Wikidata



https://tools.wmflabs.org/wikidata-todo/stats.php
https://www.wikidata.org/wiki/Wikidata_talk:WikiProject_Source_MetaData#Sources_used_as_references_on_Wikidata

References in Wikipedia

```
The molecular origins of insulin go at least as far back as
the simplest unicellular [[eukaryotes]].<ref
name='LeRoith'>{{cite journal | vauthors = LeRoith D, Shiloach
J, Heffron R, Rubinovitz C, Tanenbaum R, Roth J | title =
Insulin-related material in microbes: similarities and
differences from mammalian insulins | journal = Can. J.
Biochem. Cell Biol. | volume = 63 | issue = 8 | pages = 839–49
| year = 1985 | pmid = 3933801 | doi = 10.1139/o85-106
}}</ref> Apart from animals, insulin-like proteins are also
known to exist in Fungi and Protista kingdoms.
```

Rammstein

Grŭp Neue Deutsche Härte yw **Rammstein**. Sefydlwyd y band yn Berlin yn 1994. Mae Rammstein wedi cyhoeddi cerddoriaeth ar label recordio Republic Records, Motor Music, Universal Music Group.

Cynnwys [cuddio]

1 Aelodau
2 Disgyddiaeth
3 albwm
4 sengl
5 Misc
6 Dolen allanol
7 Cyfeiriadau

Rammstein

Gwybodaeth gefndirol

Tarddiad	yn Berlin
Cerddoriaeth	Grŭp Neue Deutsche Härte
Blynyddoedd	1994
Label(i) recordio	Republic Records, Motor Music, Universal Music Group

Aelodau [golygu | golygu cod y dudalen]

- Finnish Defence Forces
- Paul Landers
- Christian Lorenz
- Till Lindemann
- Richard Z. Krusper
- Christoph Schneider
- Oliver Riedel

Disgyddiaeth [golygu | golygu cod y dudalen]

Rhesh Wicidata

Dosedderwch y rhesh nawr | WD3 | Chwilwch am ddelweddau

albwm [golygu | golygu cod y dudalen]

enw	dyddiad cyhoeddi	label recordio
Herzeleid	1995-09-25 1995-11-24	*Motor Music*
Sehnsucht	1997-08-25	*Motor Music* *Slash Records*
Mutter	2001-04-02	*Motor Music*
Rosenrot	2005-10-28	*Universal Music Group*
Liebe ist für alle da	2009-10-16	*Universal Records* *Vagrant Records*
XXI	2015-12-04	
Rammstein		

Activity – Creating ‘worklists’ using the Listeria tool to search Wikidata.

Listeria subpages – Creating Wikidata ‘worklists’ on Wikipedia.

Basic Listeria code

===List of Scottish female writers on Wikidata===
__TOC__

{{Wikidata list
|sparql =**SELECT distinct ?item WHERE { ?item wdt:P106 wd:Q36180 . ?item wdt:P27 wd:Q22 . ?item wdt:P21 wd:Q6581072 .}**
|section=P69
|sort=label
|columns=label:article,P18,description,P19,P569,P800,P19,P21:gender,P373
|thumb=128
|min_section=1
}}

{{Wikidata list end}}

NB: Please remember to add the {{Wikidata list end}} to the code at the bottom of your listeria query.

Check the code in the Wikidata Query Service
https://query.wikidata.org/

How Do We Get People to Participate?

- This has proven difficult in past efforts
- We are getting specific research communities involved in the development
  - Chlamydiverse.org is under development

Summary/Next Steps

- Application development pattern for using a public, open and shared database
- Technical layer that allows users to integrate data into the underlying schema
- Data that comes from WikiGenomes becomes available to any application that uses Wikidata as its backend. (e.g. chlamydiverse.org, Repurpose.us)

- Modularize the framework to be repurposed for any list of taxids
- And standby for WikiGenomes 2.0!!! Coming soon

VS

Auto-extracted from Wikipedia	Added by community
CC BY-SA v3.0 Licence Attribution is required	CC0 Licence Attribution not required
Separate language versions of data	Single dataset, accessible in any language

Dbpedia is the way in which Wikipedia is translated into static RDF triples dumps.

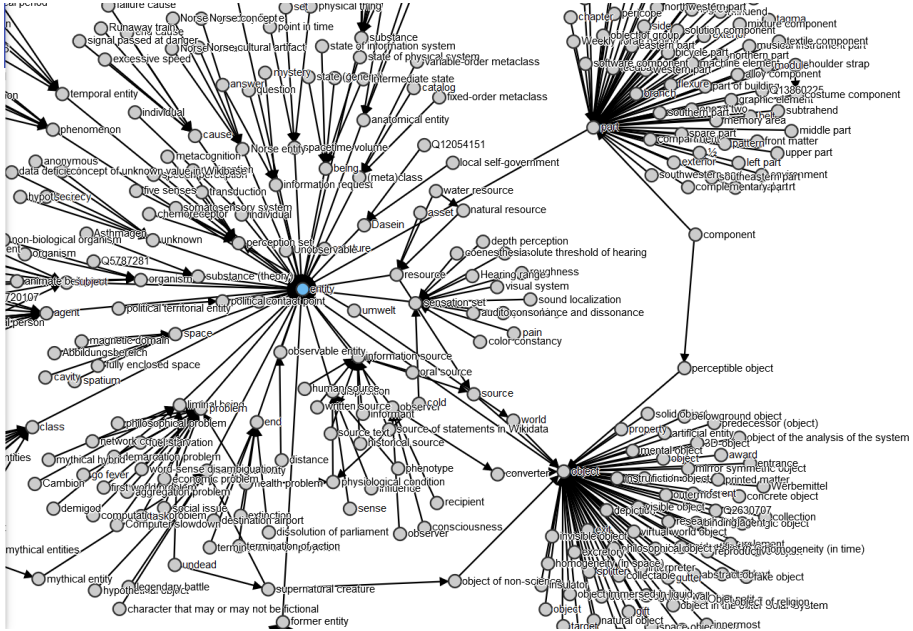
Creating new properties on Wikidata

●To create a new predicate ("property" in Wikidata), you have to add a request here - https://www.wikidata.org/wiki/Wikidata:Property_proposal - select the type of property from the list of options (e.g "Authority control" for adding an ID property). Once on one of the category pages, there's a box with button just after the contents).

●Once proposed, other community members will either support or oppose, and some questions may be asked/discussed. After around 2 weeks to a month, if there's been no unresolved objections then the property will be created by an admin.

●The notability of the data is still best discussed initially by posting a message into Project Chat - notability of a *proposed property* is also discussed on the property proposal page

The schema of Wikidata



- The hierarchy of Wikidata is being created by the community (using "instance of" and "subclass of" properties to create a tree of "what things are"). It's generally not very consistent at the moment, and is not fixed to any existing method of classifying things.

There are tools for examining the class tree though - e.g.

https://tools.wmflabs.org/sqid/#/ and

https://angryloki.github.io/wikidata-graph-builder/?property=P279&item=Q35120&limit=2&mode=reverse