



Mobilise action: Wikidata as a linked-data hub for Biodiversity data

Andra Waagmeester¹

1) Micelio, Antwerp, Belgium | Email: andra@micelio.be, Twitter: @andrawaag



The Gene Wiki project, circa 2008

Summarized
knowledge via
crowdsourcing

IL2-inducible T-cell kinase

Function

This gene encodes an intracellular tyrosine kinase expressed in T-cells. The protein is thought to play a role in T-cell proliferation and differentiation.^{[2][3]}

Structure

This protein contains the following domains, which are often found in intracellular kinases:^[4]

- N-terminus – PH (pleckstrin homology domain)
- BTK – Bruton's tyrosine kinase Cys-rich motif
- SH3 – (Src homology 3)
- SH2 – (Src homology 2)
- C-terminus – tyrosine kinase, catalytic domain

Interactions

ITK (gene) has been shown to interact with FYN,^{[5][6]} Wiskott-Aldrich syndrome protein,^{[7][8]} KHRBS1,^{[8][9][10]} PLCG1,^{[10][11]} Lymphocyte cytosolic protein 2,^{[11][12]} Linker of activated T cells,^{[12][13]} Karyophem alpha 2,^[14] Grb2^{[5][9]} and Peptidyl/prolyl isomerase A.^[15]

References

- Gibson S, Leung B, Squire JA, Hill M, Arima N, Goss P, Hogg D, Mills GB (September 1993). "Identification, cloning, and characterization of a novel human T-cell-specific tyrosine kinase located at the hematopoietin complex on chromosome 5q". *Blood* **82** (5): 1561–72. PMID 8354205.
- Kosaka Y, Felices M, Berg LJ (October 2006). "Itk and Th2 responses: action but no reaction". *Trends Immunol.* **27** (10): 453–60. doi:10.1016/j.it.2006.08.008. PMID 16931156.
- "Entrez Gene: ITK (IL2-inducible T-cell kinase)".
- Hawkins J, Marcy A (July 2001). "Characterization of Itk tyrosine kinase: contribution of noncatalytic domains to enzymatic activity". *Protein Expr. Purif.* **22** (2): 211–9. doi:10.1006/prep.2001.1447. PMID 11437596.
- Bunnell, S.C.; Diehn M; Yaffe M B; Findell P R; Cantley L C; Berg L J (Jan. 2000). "Biochemical interactions integrating Itk with the T cell receptor-initiated signaling cascade". *J. Biol. Chem. (UNITED STATES)* **275** (3): 2219–30. ISSN 0021-9258. PMID 10639929.
- Intramolecular association in a tyrosine kinase of the Tec family. *Nature (ENGLAND)* **385** (6611): 93–7. doi:10.1038/385093a0. ISSN 0028-0836. PMID 8985255.
- Perez-Villar, J. J; Kanner S B (Dec. 1999). "Regulated association between the tyrosine kinase Emh1/Tsk and phospholipase-C gamma 1 in human T lymphocytes". *J. Immunol. (UNITED STATES)* **163** (12): 6435–41. ISSN 0022-1767. PMID 10580033.
- Shim, Eun Kyung; Moon Chang Suk; Lee Gi Yeon; Ha Yun Jung; Chae Suhm-Ke; Lee Jong Ran (Sep. 2004). "Association of the Src homology 2 domain-containing leukocyte phosphoprotein of 76 kD (SLP-76) with the p85 subunit of phosphoinositide 3-kinase". *FEBS Lett. (Netherlands)* **575** (1–3): 35–40. doi:10.1016/j.febslet.2004.07.090. ISSN 0168-7739. PMID 15388330.
- Shan, X; Wang R L (Oct. 1999). "Itk/Emt/Task activation in response to CD3 cross-linking in Jurkat T cells requires ZAP-70 and Lat and is independent of membrane recruitment". *J. Biol. Chem. (UNITED STATES)* **274** (41): 29323–30. ISSN 0021-9258. PMID 10501192.
- Perez-Villar, Juan J; Whitehead-Boake, Simon

Data imported
from structured
databases

Reelin

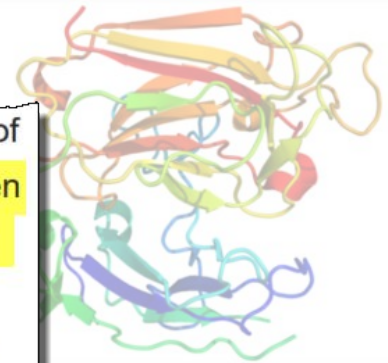
From Wikipedia, the free encyclopedia

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](#) generating [zones](#). It is found not only in the tissues.

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 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



3D ribbon structure of the third reelin repeat domain.^[1]

Available structures

PDB Ortholog search: [PDBe](#) [RCSB](#)

List of PDB id codes [\[show\]](#)

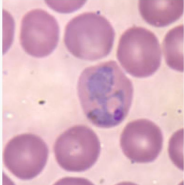
Identifiers

Symbols [RELN](#) ; LIS2; PRO1598; RL

External [OMIM: 600514](#) [MGI: 103022](#)

Wikipedia: Maintained independently by >300 language communities

Dutch

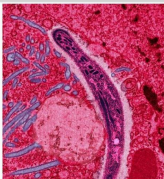
nl.m.wikipedia.org
Malaria

Rode bloedcel geïnfecteerd met <i>P. vivax</i>
Coderingen
ICD-10 B50 ↗
ICD-9 084 ↗
OMIM 248310 ↗

Greek

el.m.wikipedia.org
Ελονοσία
Ταξινόμηση και εξωτερικές πηγές

Ταξινόμηση B50 ↗
ICD-10 084 ↗
ICD-9 248310 ↗
OMIM 7728 ↗
DiseasesDB 000621 ↗
MedlinePlus med/1385 ↗ emerg/305 ↗
eMedicine ped/1357 ↗

English

en.m.wikipedia.org
Malaria

A Plasmodium from the saliva of a female mosquito moving across a mosquito cell
Classification and external resources
Specialty Infectious disease
ICD-10 B50 ↗ -B54 ↗
ICD-9-CM 084 ↗



Dutch

nl.m.wikipedia.org
Curacao
Hoofdstad Oranjestad
Regeringsvorm Constitutionele monarchie
Staatshoofd Koning Willem-Alexander Fredis Riefunjo (gouverneur)
Regeringsleider Mike Eman (Arubaanse Volkspartij)
Religie Katholiek 82%, protestant 8%

103,400 ^[2] (197th)

• Εκτίμηση 2014

Greek

el.m.wikipedia.org
Πολίτευμα Συνταγματική Μοναρχία
Μονάρχης Γουλιέλμος-Κυβερνήτης Αλέξανδρος
Πρωθυπουργός Φρένις Ρεφουνιόλ Μάικ Έμαν
Πλήρης αυτονομία από το Βασίλειο των Κάτω Χωρών Συντάγμα 1η Ιανουαρίου 1986
Έκταση • Συνολό Ακτογραμμή 180 km² (213η) 68,5 km
Πληθυσμός • Εκτίμηση 2014 107.394 ^[1] (196η) • Απογραφή 2000 103.065 • Πυκνότητα 556,4 κατ./km² (21η)
Α.Ε.Π. (PPP) • Ολικό (2005) 2,258 δισ. \$ ^[2]

107.394^[1] (196η)

English

en.m.wikipedia.org
Forma di governance Democrazia parlamentare
Rei Monarkia costituzionale
Governador Willem-Alexander
Prome Minister Fredis Riefunjo
Pais den Reino di Hulanda Mike Eman
Status aparte 1 januari di 1986
Area • Total 193 km² (n/a)

101.484 (2010)^[2]

110.663 (2014)^[3]

(614,8/km² (2014))

Wikidata is to data as Wikipedia is to text

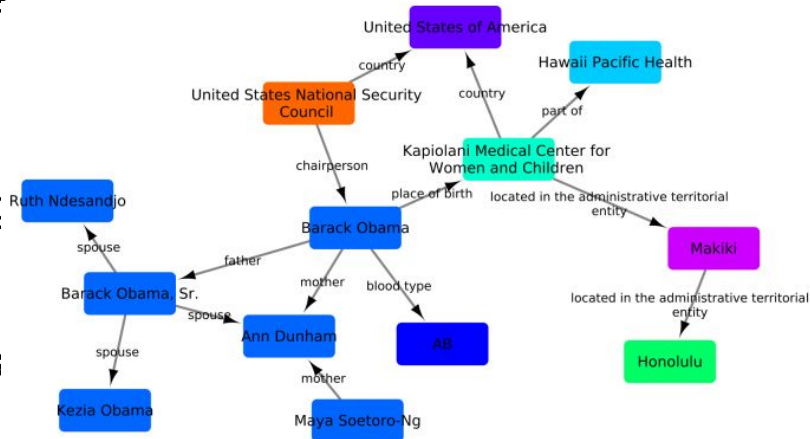
Wikidata is a collaboratively edited knowledge base operated by the Wikimedia Foundation



- Completely free, even for commercial usage (CC0)
- Anybody can contribute
- Covers all domains of knowledge
- Extensive item history, talk pages, projects, users
- Integration with the semantic web
- High performance query engine (SPARQL)

- Stable! Long term support not dictated by funding cycles
- Actively developed
- Already has large number of active users, editors contributors!

A giant graph of knowledge!



Reelin - Wikidata

Secure | https://www.wikidata.org/wiki/Q13561329

☆

📄 📌 🕒 🔗 📺 🗨️ 🗺️ 🗳️ 🗨️ 🗨️

English Andrawaag 21 99+ Talk Preferences Beta Watchlist Contributions Log out

Item Discussion

Read View history ⭐ More

Search Wikidata 🔍



[Main page](#)[Community portal](#)[Project chat](#)[Create a new item](#)[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](#)[Import interwiki](#)

Reelin (Q13561329)

mammalian protein found in Homo sapiens [edit](#)

RELN | reelin | uniprot:P78509

[In more languages](#)

Statements

instance of	 protein 1 reference + add value
subclass of	 protein 1 reference  Reelin 1 reference + add value
image	 2DDU.png 1 reference + add value

Retinoic acid receptor alpha (Q254943)

mammalian protein found in Homo sapiens
Nuclear receptor subfamily 1 group B member 1 | RARA

Statements

molecular function

molecular function (P680)

represents gene ontology function annotations

Wikipedia (7 entries)

ar

مستقبل حمض الريتينويك ألفا

en

Retinoic acid receptor alpha

es

Receptor de ácido retinoico alfa

sh

Receptor retinoinske kiseline alfa

sr

Receptor retinoinske kiseline alfa

uk

RARA

zh

视黄酸受体α

retinoic acid binding (Q14901431)

Interacting selectively and non-covalently with retinoic acid, 3,7 GO:0001972

Statements

subclass of

retinoid binding

1 reference

retinoic acid binding

determination method

1 reference

retrieved

stated in

3 January 2017

A human retinoic acid receptor which belongs to the family of nuclear receptors

UniProt-GOA

British Heart Foundation

http://www.ebi.ac.uk/QuickGO/GA nnotation?protein=P10276

IDA

transcription corepressor activity

determination method

IDA

1 reference

IDA (Q23174122)

Gene Ontology evidence code

Inferred from Direct Assay

Statements

instance of

Gene Ontology Evidence code

manual assertion

A human retinoic acid receptor which belongs to the family of nuclear receptors (Q24339631)

Statements

instance of

scientific article

Identifiers

PubMed ID

2825025

British Heart Foundation (Q4970039)

Statements

instance of

organization

official website

http://www.bhf.org.uk/

Identifiers

GRID ID

grid.452924.c



Revision history of "Retinoic acid receptor alpha" (Q254943)

[View logs for this item](#)

Search for revisions

From year (and earlier):

From month (and earlier):

Tag filter:

Diff selection: Mark the radio boxes of the revisions to compare and hit enter or the button at the bottom.

Legend: **(cur)** = difference with latest revision, **(prev)** = difference with preceding revision, **m** = minor edit.

Select: [All](#), [None](#), [Invert](#)

- ☐ [\(cur | prev\)](#) ☐ 20:13, 21 March 2017 ProteinBoxBot (talk | contribs) .. (454,236 bytes) (-440) .. (Updated item: replace thumbnail gene atlas image with fs) (undo)
- ☐ [\(cur | prev\)](#) ☐ 08:00, 28 January 2017 Edoderoobot (talk | contribs) .. (454,676 bytes) (+67) .. (Updated item: #proteine) (undo) (restore)
- ☐ [\(cur | prev\)](#) ☐ 12:06, 4 January 2017 ProteinBoxBot (talk | contribs) .. (454,609 bytes) (+165,607) .. (Updated item: update GO terms) (undo) (restore)
- ☐ [\(cur | prev\)](#) ☐ 03:57, 3 January 2017 ProteinBoxBot (talk | contribs) .. (289,002 bytes) (+1,584) .. (Updated item) (undo) (restore)
- ☐ [\(cur | prev\)](#) ☐ 09:07, 17 September 2016 Okkn (talk | contribs) .. (287,418 bytes) (-2) .. (Changed claim: subclass of (P279): Retinoic acid receptor (Q2838685)) (undo | thank) (restore)
- ☐ [\(cur | prev\)](#) ☐ 15:18, 16 September 2016 ProteinBoxBot (talk | contribs) .. (287,420 bytes) (-292) .. (Updated item) (undo) (restore)
- ☐ [\(cur | prev\)](#) ☐ 12:03, 17 August 2016 ProteinBoxBot (talk | contribs) .. (287,712 bytes) (0) .. (Updated item) (undo) (restore)
- ☐ [\(cur | prev\)](#) ☐ 04:50, 8 August 2016 ProteinBoxBot (talk | contribs) .. (287,712 bytes) (+11,503) .. (Updated item) (undo) (restore)

Getting data in..

License: CC0

- All structured data from the main, Property, Lexeme, and EntitySchema namespaces is available under the Creative Commons CC0 License

License	Add to Wikidata	Add to Commons	Add to Wikipedia
CC0	+	+	+
CC-BY	+	+	+
CC-BY-NC	-	-	-
CC-BY-SA	-	+	+
CC-BY-ND	-	-	-
CC-BY-NC-SA	-	-	-
CC-BY-NC-ND	-	-	-

Eligibility for inclusion

1. It contains at least one valid sitelink to a page on Wikipedia, Wikivoyage, Wikisource, Wikiquote, Wikinews, Wikibooks, Wikidata, Wikispecies, Wikiversity, or Wikimedia Commons.
2. It refers to an instance of a clearly identifiable conceptual or material entity. The entity must be notable, in the sense that it can be described using serious and publicly available references.
3. It fulfills some structural need, for example: it is needed to make statements made in other items more useful.

Infrastructure

Resource

Content

Where?



Text

<https://<lang>.wikipedia.org>
<https://releases.wikimedia.org/mediawiki/>



Media

<https://commons.wikimedia.org>



Data

<https://www.wikidata.org>
<https://www.wikibase.org>

Wikibase and WBStack



github.com/wmde/wikibase-docker

[Read the guide](#)

 [wmde / wikibase-docker](#) 

[Code](#) [Pull requests 6](#) [Actions](#) [Security](#) [Insights](#)

 Docker images and example compose file for Wikibase and surrounding services

wbstack.com

[WbStack](#) Alpha Hi andra@micel.io [DASHBOARD](#) [ACCOUNT](#) [LOGOUT](#)

Your open data..

..starting with a Wikibase stack

Blog posts:

- [Infrastructure overview](#)
- [November review](#)
- [October introduction](#)



What is Wikibase?

[Wikibase](#) is an open source software suite for running a collaborative knowledge base. One installation of it is [Wikidata](#).

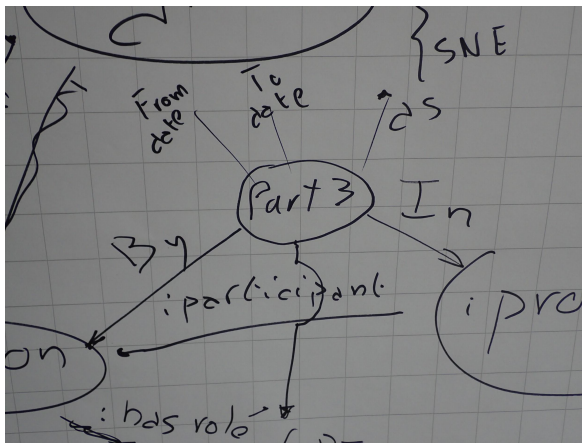
— learningwikibase.com CC-BY 4.0

[Community](#)
[MediaWiki](#)
[Wikibase](#)

[Organization](#)
[About Us](#)
[Privacy](#)

Community engagement and model discussion

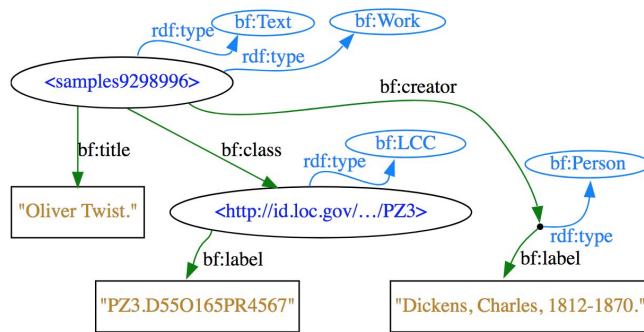
Data
Integration
Extension for
Grants
Ontology



Formally capture and describe model and community consensus

Model development

- Legacy review – develop punch lists for existing data issues that needs fixing
- Documentation – terse, human-readable representation helping contributors and maintainers quickly grok the model
- Client pre-submission – submitters test their data before submission to make sure they're saying what they want to say and that the receiving schema can accommodate all of their data
- Server pre-ingestion – submission process checks data as it comes in and either rejects or warns about non-conformant data



```
Data (Turtle)
<samples9298996>
  rdf:type bf:Text ;
  rdf:type bf:Work ;
  bf:title "Oliver Twist." ;
  bf:class <id.loc.gov/.../PZ3> ;
  bf:creator [
    rdf:type bf:Person ;
    bf:label "Dickens, Charles, 1812-1870." ;
  ] .

<id.loc.gov/.../PZ3>
  rdf:type bf:LCC ;
  bf:label "PZ3.D55O165PR4567" .
```

← → ↺ 🏠

🔒 <https://www.wikidata.org/wiki/EntitySchema:E37>

⌵ 📄 ⋮ 📁 ⭐

⬇️ 📏 📖 📷 📺 📡 📶 ⚙️

Donate

Print/export

Download as PDF

Tools

What links here

Related changes

Special pages

Permanent link

Page

Cite

pt	gene humano	edit
----	-------------	----------------------

E108: genome_assembly
IMPORT <<https://www.wikidata.org/wiki/Special:EntitySchemaText/E108>>
PREFIX E108: <<https://www.wikidata.org/wiki/Special:EntitySchemaText/E108#>>

E109: human_chromosome
IMPORT <<https://www.wikidata.org/wiki/Special:EntitySchemaText/E109>>

p:P31 @<#P31_instance_of_gene> ;

PREFIX wdt: <<http://www.wikidata.org/prop/direct/>>
PREFIX p: <<http://www.wikidata.org/prop/>>

<#P31_instance_of_gene> {
 ps:P31 @<#gene_types> ; # Instance of [P31] gene types
 prov:wasDerivedFrom @<#ncbi-gene-reference> OR @<#ensembl-gene-reference>
}

start = @<#wikidata-human-gene>

(
 p:P644 @<#P644_genomic_start> ; # Its genomic start location
 p:P645 @<#P645_genomic_end> ; # Its genomic end location
)* ; # Zero or more start and end locations.

Value statements contain either actual values, or pointers to other Wikidata items.
Identifier statements capture
external identifiers, erroneous statements are those that are errors.

check entities against this Schema [🔗](#) | [edit](#) |

Enter an entity to check e.g.Q42 [Check](#)

ShEx2 — Simple Online Validator

```
# Shape Expression for Human genes in Wikidata
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX p: <http://www.wikidata.org/prop/>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX pq: <http://www.wikidata.org/prop/qualifier/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX prv: <http://www.wikidata.org/prop/reference/value/>
PREFIX pr: <http://www.wikidata.org/prop/reference/>
PREFIX ps: <http://www.wikidata.org/prop/statement/>

BASE <http://www.wikidata.org/entity/>

start = @<#wikidata-human-gene>

# Query with results
# SELECT * WHERE {?item wdt:P31 wd:Q7187 ; wdt:P703 wd:Q15978631 .} LIMIT 10

# Indicates which shape to use to start iterating over the graph if none is
provided.

# wikidata-human gene is the main shape for a human gene data model in Wikidata.
Each line between the brackets
# represents the structure than can be enforced to validate human gene annotations
in Wikidata
```

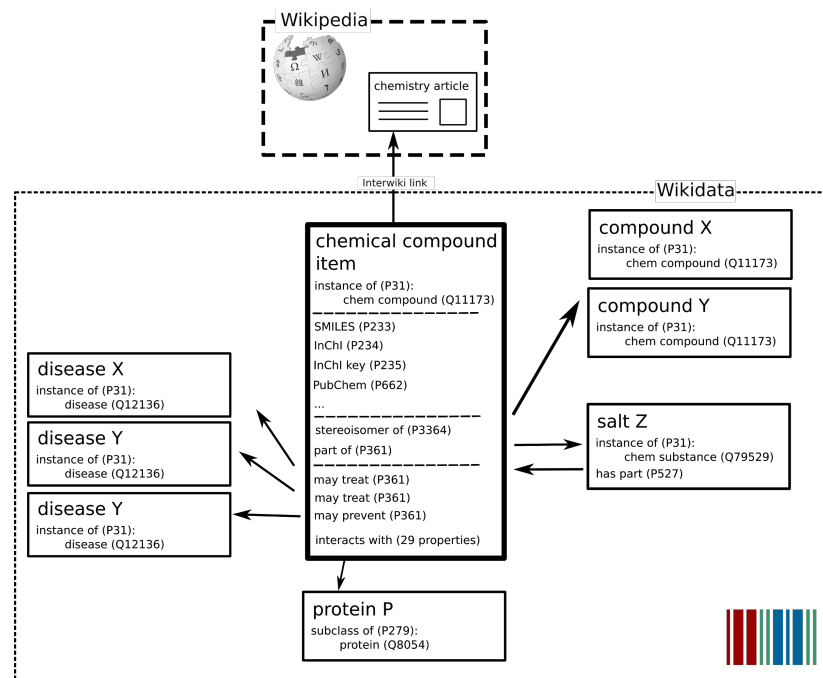
abort (ctrl-enter)

Query **Entities to check**

<http://www.wikidata.org/entity/Q414043>	@	START	-	✓
<http://www.wikidata.org/entity/Q415594>	@	START	-	✓
<http://www.wikidata.org/entity/Q416426>	@	START	-	✓
<http://www.wikidata.org/entity/Q417169>	@	START	-	✓
<http://www.wikidata.org/entity/Q417743>	@	START	-	✓
<http://www.wikidata.org/entity/Q418553>	@	START	-	✓

Seeding with data

- Model structure of items (genes, drugs, diseases, .. etc) & relationships between items
- Import data from many sources and ontologies
- Linked to many identifiers from external databases
- Architecture for maintaining data from external sources



A Wikidata Python module integrating the MediaWiki API and the Wikidata SPARQL endpoint

397 commits

2 branches

1 release

7 contributors

MIT

Branch: master

New pull request

Find file

Clone or download

 sebotic fixed an omission where new items don't get created when domain not s... Latest commit 2f5d2fd 22 hours ago

 doc Wikidata to Wikipedia mapping prototype for diseases added. 2 years ago

 wikidataintegrator fixed an omission where new items don't get created when domain not s... 22 hours ago

 Jenkins

Jenkins

Running

New Item

People

Build History

Edit View

Delete View

Manage Jenkins

My Views

Credentials

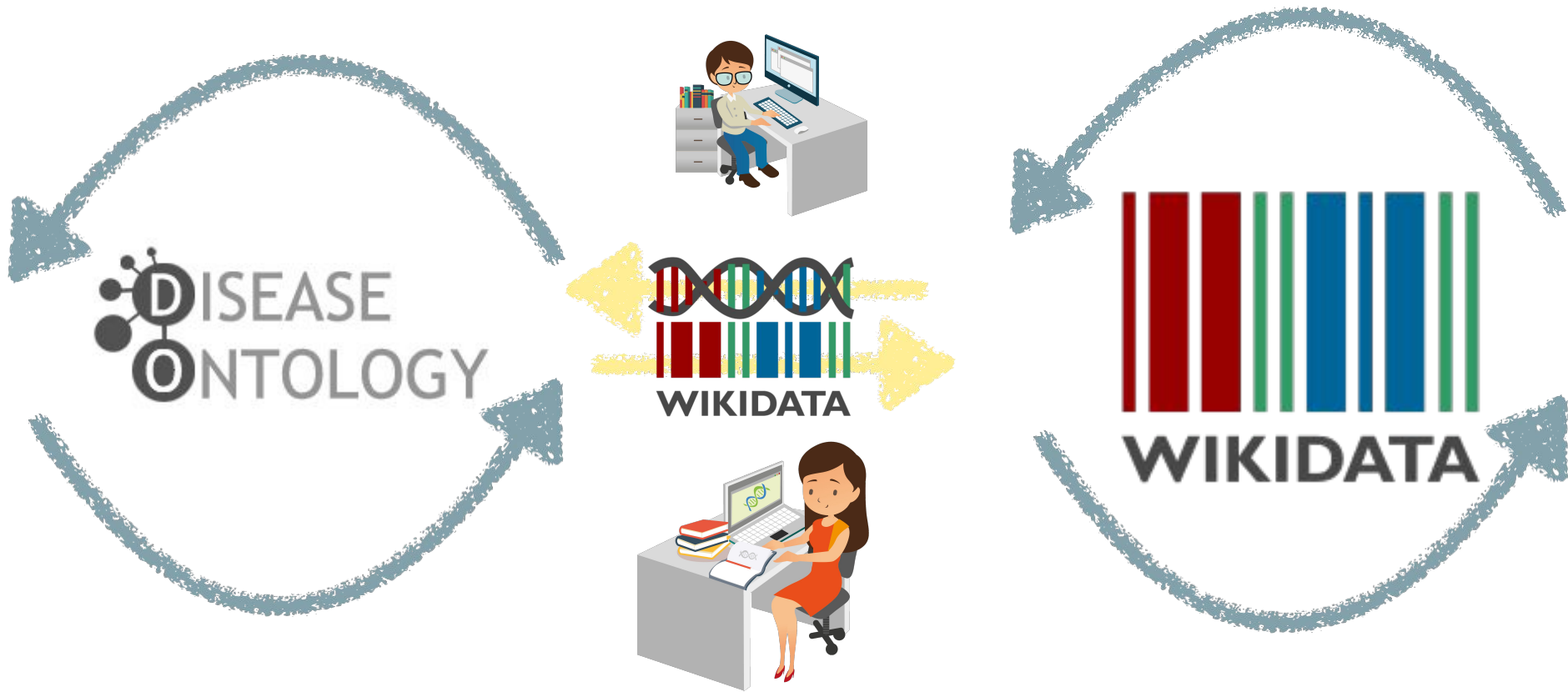
Build Queue

Running Bots

AllRunning+

S	Name	Last Success ↑	Last Failure
	ProteinBot_homo_sapiens	1 day 21 hr - #12	N/A
	GOBot_bigmern	2 days 15 hr - #15	9 days 15 hr - #14
	GeneBot_Homo_sapiens	2 days 19 hr - #25	2 days 20 hr - #24
	Disease_Ontology	2 days 23 hr - #11	4 days 13 hr - #8
	GeneDiseaseBot	2 days 23 hr - #9	1 mo 6 days - #2

Feedback loop



Use..

Simple data retrieval

“Retrieve genes with
GWAS association
with asthma”



39 genes

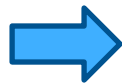
gene	geneLabel		gene	geneLabel		gene	geneLabel		gene	geneLabel
Q5013317	COL22A1		Q18027370	IGSF3		Q18053559	CDHR3		Q14903974	SMAD3
Q14912759	SLC22A5		Q18045382	HPSE2		Q18045669	ATG3		Q18033889	IL1RL1
Q14914243	PSAP		Q18048437	IL33		Q18035037	RAD50		Q17917202	ERBB4
Q14907990	SLC30A8		Q18051900	PYHIN1		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			

```

1 SELECT DISTINCT ?gene ?geneLabel where {
2   ?gene wdt:P2293 wd:Q35869 .    # gene has genetic association to "asthma"
3   ?gene wdt:P31 wd:Q7187 .       # gene is subclass of "gene"
4   SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
5 }
```

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

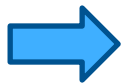
```

1 SELECT DISTINCT ?gene ?geneLabel where {
2   ?gene wdt:P2293 wd:Q35869 . # gene has genetic association to "asthma"
3
4   ?gene wdt:P31 wd:Q7187 . # gene is subclass of "gene"
5
6   ?gene wdt:P688 ?protein . # gene encodes a protein
7   ?protein wdt:P681 ?cc . # protein has a cellular component
8   ?cc wdt:P279*|wdt:P361* wd:Q14349455 . # cell component is 'part of' or 'subclass of' membrane
9
10  SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
11 }
```

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

```

1 SELECT ?diseaseGALabel (count (DISTINCT ?geneLabel) as ?gene_counts)
2 (group_concat(DISTINCT ?geneLabel; separator="," as ?geneList)
3   ?gene wdt:P2293 ?diseaseGA . # gene is subclass of "gene" and encodes protein
4   ?diseaseGA wdt:P279* wd:Q3286546 . # to a type of respiratory system disease
5
6   ?gene wdt:P31 wd:Q7187 ; wdt:P688 ?protein ; # gene is subclass of "gene" and encodes protein
7     rdfs:label ?geneLabel .
8   FILTER (lang(?geneLabel) = "en")
9   ?protein p:P681 ?s . # protein's cell component statement
10  ?s ps:P681 ?cp . # get statement value
11  FILTER NOT EXISTS (

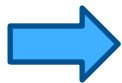
```

http://bit.ly/bosc2017_wikidata

... and show associated pathways

16 genes / 59 pathways

“Retrieve genes with GWAS association with any respiratory disease and gene product is localized to membrane (non-IEA), show causative chemical hazards and **show pathways where they have a role.**”



gene	pathway
SMAD3	Androgen receptor signaling pathway
SMAD3	TGF-beta Receptor Signaling
SMAD3	mechlorethamine exposure
HLA-C	Allograft Rejection
SFTPD	Regulation of toll-like receptor signaling pathway
....	

```

11  .cp wdt:P279 | wdt:P301 wd:Q4915012 .
12
13  ?pathway wdt:P31 wd:Q4915012 ;           # instance of a biological pathway
14      wdt:P527 ?gene .
15
16  SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
17  }
```

http://bit.ly/bosc2017_wikidata


```
PREFIX wp:      <http://vocabularies.wikipathways.org/wp#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dc:      <http://purl.org/dc/elements/1.1/>
```

```
SELECT DISTINCT ?metabolite1Label ?metabolite2Label ?mass1 ?mass2 WITH {
```

```
  SELECT ?metabolite1 ?metabolite2 WHERE {
    ?pathwayItem wdt:P2410 "WP706";
                wdt:P2888 ?pwIri.
```

Wikidata

```
  SERVICE <http://sparql.wikipathways.org/> {
    ?pathway dc:identifier ?pwIri.
    ?interaction rdf:type wp:Interaction;
                  wp:participants ?wpmb1, ?wpmb2;
                  dcterms:isPartOf ?pathway.

    FILTER (?wpmb1 != ?wpmb2)
    ?wpmb1 wp:bdbWikidata ?metabolite1.
    ?wpmb2 wp:bdbWikidata ?metabolite2.
  }
```

Wikipathways

```
} AS %metabolites WHERE {
```

```
  INCLUDE %metabolites.
  ?metabolite1 wdt:P2067 ?mass1.
  ?metabolite2 wdt:P2067 ?mass2.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
```

Wikidata

[Try me....](#)

From a remote SPARQL endpoint to Wikidata



SPARQL Downloads

Documentation/Help Contact

Your query

Add common prefixes

```
20 SELECT DISTINCT ?wd_item ?physically_interacts_with ?interactswithLabel ?type ?iri ?uniprot ?text WHERE {
21   {SELECT * WHERE { ?iri a up:Protein ;
22     up:organism taxon:9606 ;
23     up:annotation ?annotation .
24     ?annotation a up:Natural_Variant_Annotation ;
25     rdfs:comment ?text .
26     FILTER (CONTAINS(?text, 'loss of function'))
27   }}
28   SERVICE <https://query.wikidata.org/bigdata/namespace/wdq/sparql> {
29     VALUES ?use {wd:Q427492}
30     ?wd_item wdt:P352 ?uniprot ;
31       wdt:P129 ?physically_interacts_with ;
32       wdt:P2888 ?iri ;
33       wdt:P703 wd:Q15978631 .
34     ?wd_item p:P129 ?phys_interacts_with_node .
35     ?phys_interacts_with_node ps:P129 ?physically_interacts_with ;
36       pq:P366 ?use .
37     ?physically_interacts_with wdt:P31 ?type ;
38       rdfs:label ?interactswithLabel .
39     FILTER (lang(?interactswithLabel) = "en")
40   }}
```

UniProt

Wikidata

Submit Query

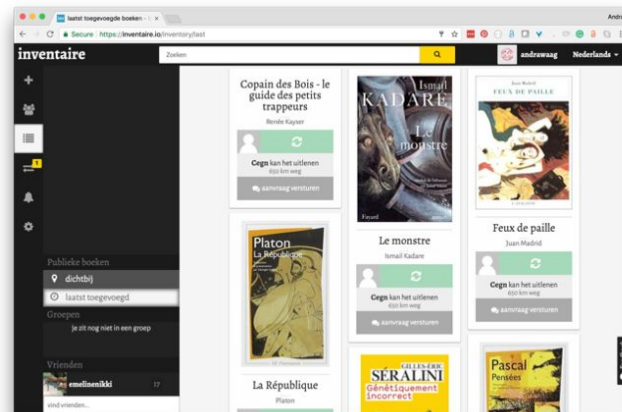
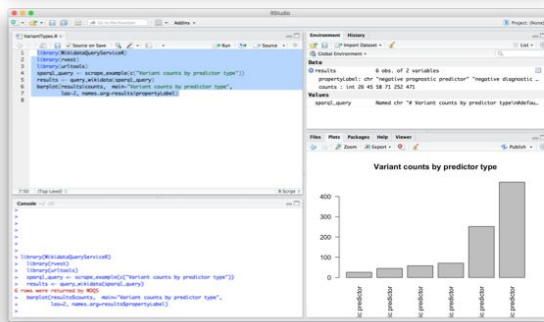
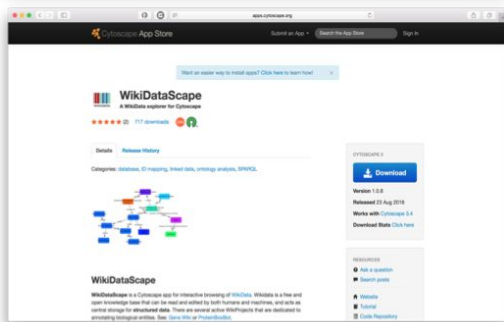
Cancel

Tools using Wikidata

<http://www.wikigenomes.org> Wikipedia and Wikidata
google plugin



<http://inventaire.io>



Cytoscape

R plugins

Introducing Wikiproject iNaturalist



Acknowledgements

Wikidata as a FAIR knowledge graph for the life sciences

 Andra Waagmeester,  Gregory Stupp,  Sebastian Burgstaller-Muehlbacher,  Benjamin M Good,  Malachi Griffith,  Obi Griffith,  Kristina Hanspers,  Henning Hermjakob,  Kevin Hybiske,  Sarah M. Keating,  Magnus Manske,  Michael Mayers,  Elvira Mitraka,  Alexander R. Pico,  Timothy Putman,  Anders Riutta,  Núria Queralt-Rosinach,  Lynn M. Schriml,  Denise Slenter,  Ginger Tsueng,  Roger Tu,  Egon Willighagen,  Chunlei Wu,  Andrew I Su

doi: <https://doi.org/10.1101/799684>

Thousands of Wikidatans



By Helpameout - Own work, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=20337311>

Funding

- National Institutes of General Medical Sciences (R01GM089820)
- NIH Common Fund programs for Big Data to Knowledge (U54GM114833)

Acknowledgements



Lynn Schriml



Benjamin
Good

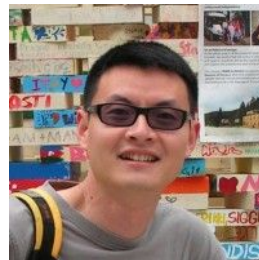
Tim Putman



Andrew Su



Sebastian
Burgstaller



Thousands of WikiDatans
Chunlei Wu

- Elvira Mitraka
(Disease Ontology, U
Baltimore)
- Gang Fu, Evan Bolton
(NIH, PubChem)
- Wikimedia Foundation



Gregory Stupp

By Helpameout - Own work, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=20337311>

Funding

- National Institutes of General Medical Sciences (R01GM089820)
- NIH Common Fund programs for Big Data to Knowledge (U54GM114833)