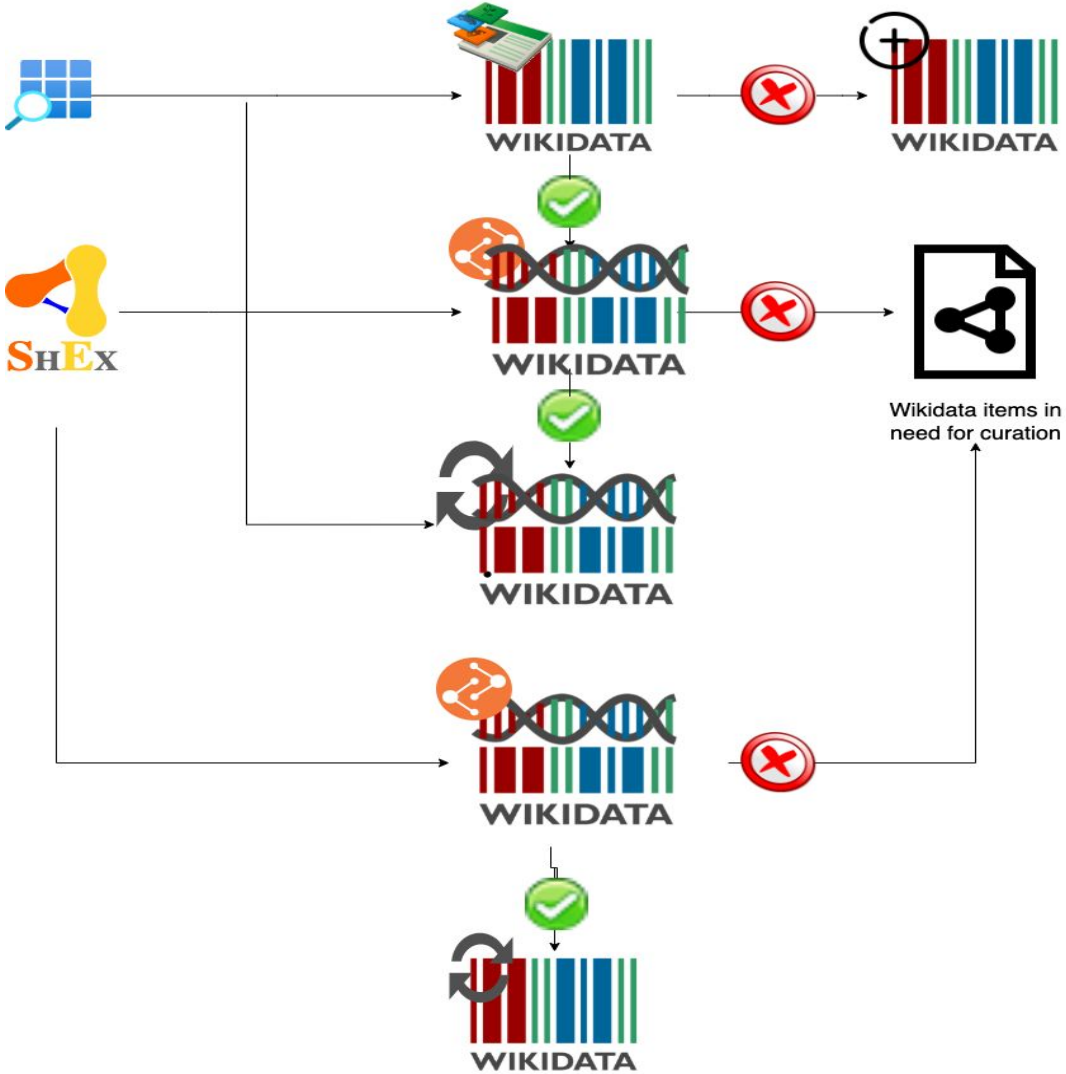
Gene Wiki: Reusing FAIR data on Wikidata

Andra Waagmeester,
Micelio, Antwerp, Belgium | Email: andra@micelio.be, Twitter: @andrawaag


WIKIDATA
WIKIDATA
ShEx
WIKIDATA
Wikidata items in need for curation
WIKIDATA
WIKIDATA

FAIR …

Findable

Accessable

Interoperable

R reusable

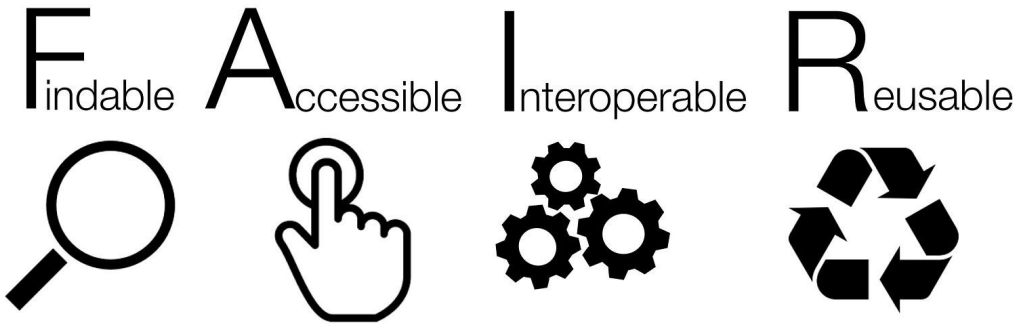
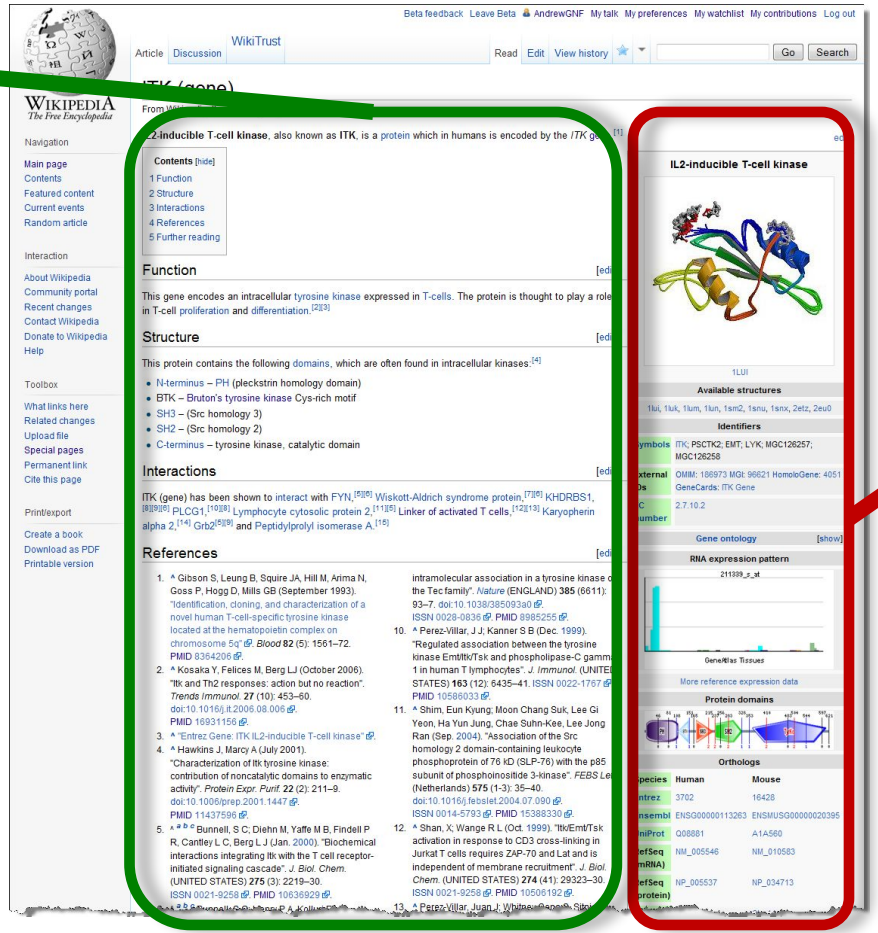


Image source: https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg

The Gene Wiki project, circa 2008

Summarized knowledge via crowdsourcing



Data imported from structured databases

Huss, PLoS Biol, 2008

Reelin

From Wikipedia, the free encyclopedia

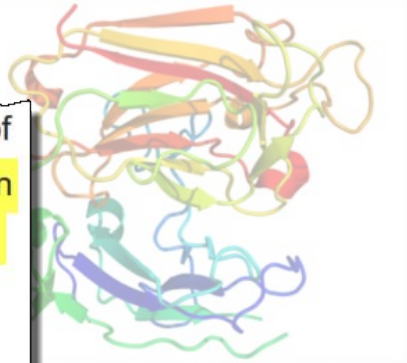
Reelin is a large secreted extracellular matrix glycoprotein that helps regulate processes of neuronal migration and positioning in the developing brain by controlling cell–cell interactions. Besides this important role in early development, reelin continues to work in the adult brain. It modulates synaptic 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 be 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]

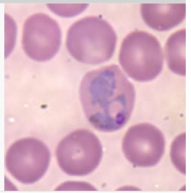
Identifiers	
Symbols	RELN ; LIS2; PRO1598; RL
External	OMIM: 600514 MGI: 103022

Wikipedia: Maintained independently by >300 language communities

Dutch

nl.m.wikipedia.org

Malaria



Rode bloedcel geïnfecteerd met *P. vivax*

Coderingen

ICD-10	B50
ICD-9	084
OMIM	248310

Greek

el.m.wikipedia.org

Ελονοσία

Ταξινόμηση και εξωτερικές πηγές

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

ICD-10 B50

ICD-10 B50-B54

Ταξινόμηση ICD-10 B50

Dutch

nl.m.wikipedia.org

	Papiamento
Hoofdstad	Oranjestad
Regeringsvorm	Constitutionele monarchie
Staatshoofd	Koning Willem-Alexander Fredis Refunjol (gouverneur)
Regeringsleider	Mike Eman (Arubaanse Volkspartij)
Religie	Katholiek 82%, protestant 8%

103,400 [2] (197th)

Greek

el.m.wikipedia.org

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

• Εκτίμηση 2014 107.394[1] (196η)

English

pap.m.wikipedia.org

Forma di gobernacion	Democracia p Monarkia con
- Rei	Willem-Alexan
- Gobernador	Fredis Refunjo
- Prome Minister	Mike Eman
Pais den Reino di Hulanda	
Status aparte	1 januari di 19
Area	
- Total	193 km² (n/a)

101.484 (2010)[2]

110.663 (2014)[3]

(614,8/km² (2014))

Reelin - Wikipedia

Andra

Secure | https://en.wikipedia.org/wiki/Reelin

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

Article Talk | Read Edit View history More | Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

Reelin

From Wikipedia, the free encyclopedia

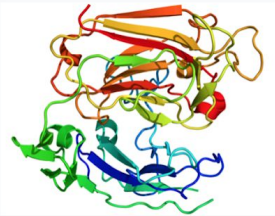
Reelin (RELN)[5] is a large secreted extracellular matrix glycoprotein that helps regulate processes of neuronal migration and positioning in the developing brain by controlling cell-cell interactions. Besides this important role in early development, reelin continues to work in the adult brain. It modulates synaptic plasticity by enhancing the induction and maintenance of long-term potentiation.[6][7] It also stimulates dendrite[8] and dendritic spine[9] development and regulates the continuing migration of neuroblasts generated in adult neurogenesis sites like subventricular and subgranular zones. It is found not only in the brain, but also in the spinal cord, blood, and other body organs and tissues.[*citation needed*]

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,[10] 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[11] are controversial.[12][13] Total lack of reelin causes a form of lissencephaly. Reelin may also play a role in Alzheimer's disease, temporal lobe epilepsy and autism.[*citation needed*]

Reelin's name comes from the abnormal reeling gait of *reeler* mice,[14] which were later found to have a deficiency of this brain protein and were homozygous for mutation of the RELN gene. The primary phenotype associated with loss of reelin function is a failure of neuronal positioning throughout the developing central nervous system (CNS). The mice heterozygous for the reelin gene, while having little neuroanatomical defects, display the endophenotypic traits linked to psychotic disorders.[15]

Contents [hide]

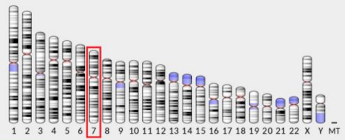
RELN



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

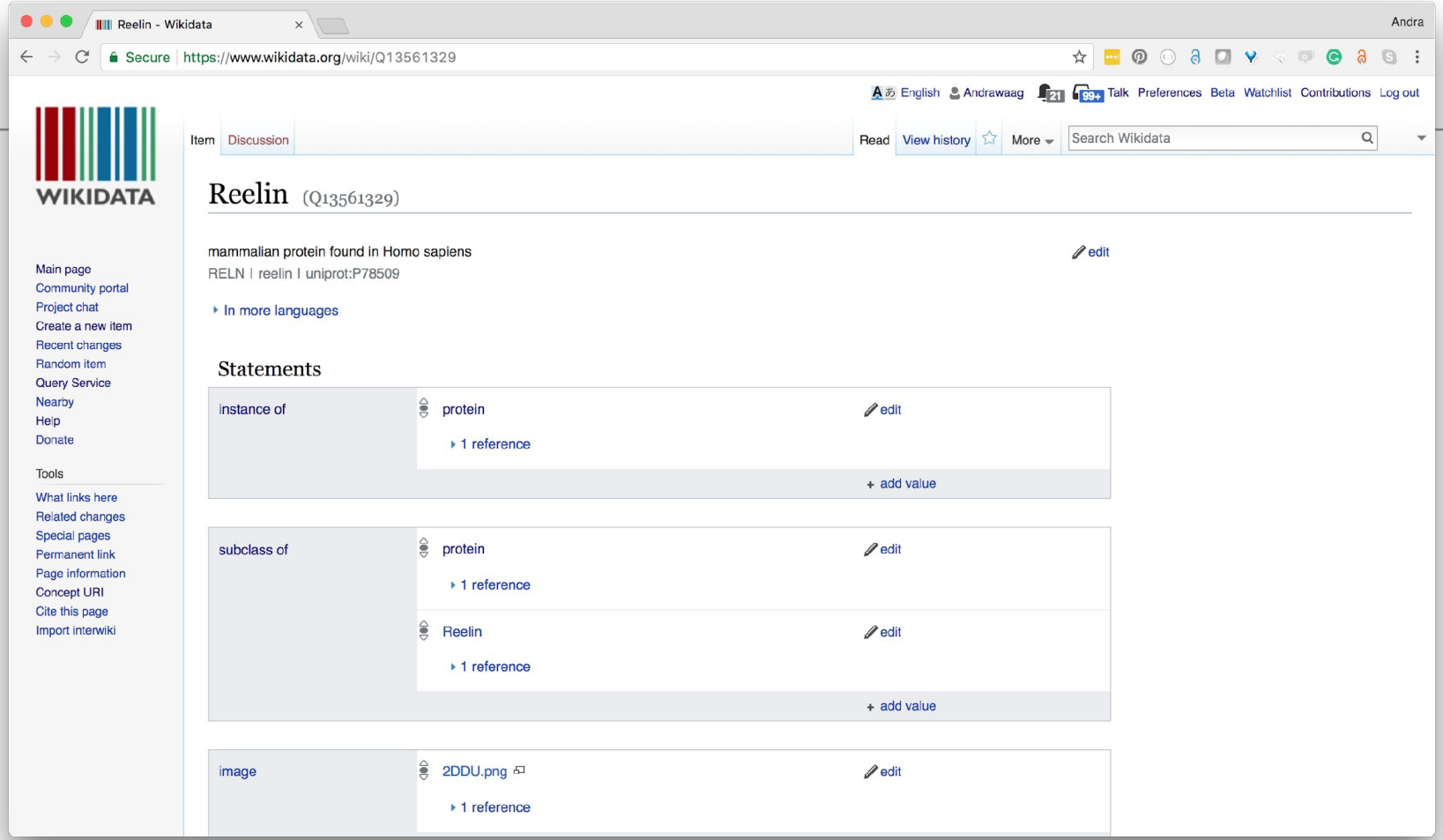
Identifiers	
Aliases	RELN, LIS2, PRO1598, RL, reelin, ETL7
External IDs	OMIM: 600514 MGI: 103022 HomoloGene: 3699 GeneCards: RELN

Gene location (Human) [hide]



Chr. Chromosome 7 (human)[1]

https://en.wikipedia.org/wiki/Reelin#Psychotropic_medication


Reelin - Wikidata
Andra
Secure https://www.wikidata.org/wiki/Q13561329
English Andrawaag 21 99+ Talk Preferences Beta Watchlist Contributions Log out
WIKIDATA
Item Discussion
Read View history More
Search Wikidata
Reelin (Q13561329)
mammalian protein found in Homo sapiens
edit
RELN | reelin | uniprot:P78509
In more languages
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
Statements
instance of
protein
edit
1 reference
add value
subclass of
protein
edit
1 reference
Reelin
edit
1 reference
add value
image
2DDU.png
edit
1 reference

Retinoic acid receptor alpha (Q254943)

mammalian protein found in Homo sapiens

Nuclear receptor subfamily 1 group B member 1 | RARA

Statements

molecular function

retinoic acid binding — edit

determination method — IDA

1 reference

retrieved — 3 January 2017

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

UniProt-GOA

curator — British Heart Foundation

reference URL — http://www.ebi.ac.uk/QuickGO/GAnnotation?protein=P10276

determination method — IDA

+ add reference

transcription corepressor activity — edit

determination method — IDA

1 reference

molecular function (P680)

represents gene ontology function annotations

retinoic acid binding (Q14901431)

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

GO:0001972

Statements

subclass of — retinoid binding

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

Wikipedia (7 entries) edit

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

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

View logs for this item

Search for revisions

From year (and earlier): 2017 From month (and earlier): all Tag filter: Show

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

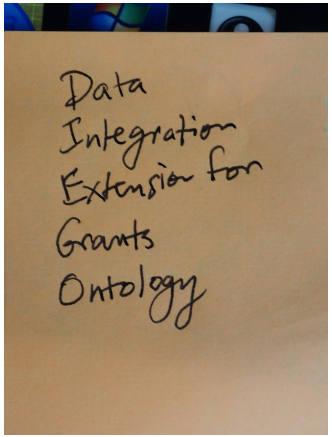
Legend: **(cur)** = difference with latest revision, **(prev)** = difference with preceding revision, **m** = minor edit.

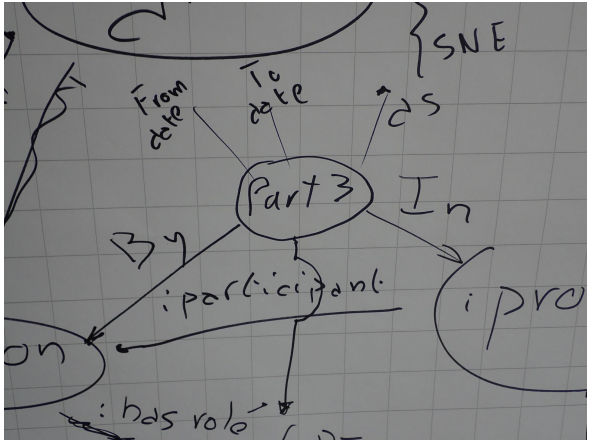
Compare selected revisions

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, 9 August 2016 ProteinBoxBot (talk | contribs) . . (287,712 bytes) **(+11,503)** . . *(Updated item)* (undo) (restore)

Community engagement and model discussion

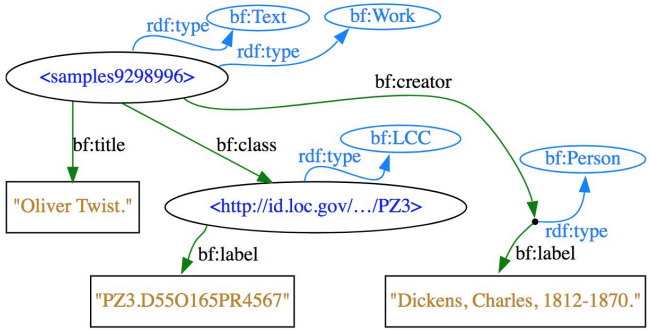




Formally capture and describe model and community consensus

Model development

- Legacy review – develop punch lists for existing data issues that needs fixing
- Documentation – terse, human-readable representation helping contributers 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

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>
```

check entities against this Schema | edit |

Enter an entity to check e.g.Q42 Check

```
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.
```

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
```

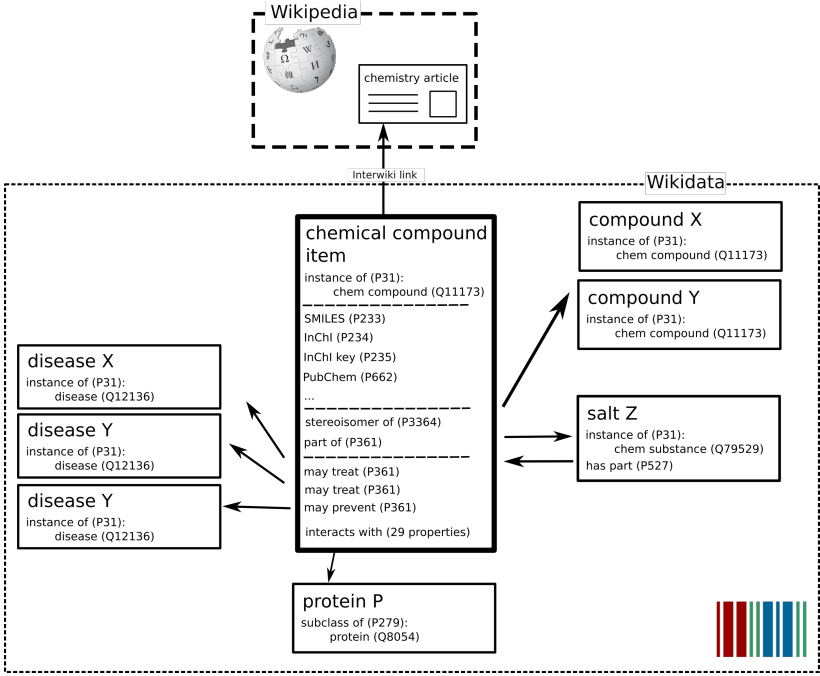
abort (ctl-enter)

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



SuLab / **WikidataIntegrator**

Watch 7 | Star 16 | Fork 6

Code | Issues 4 | Pull requests 1 | Projects 0 | Pulse | Graphs

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

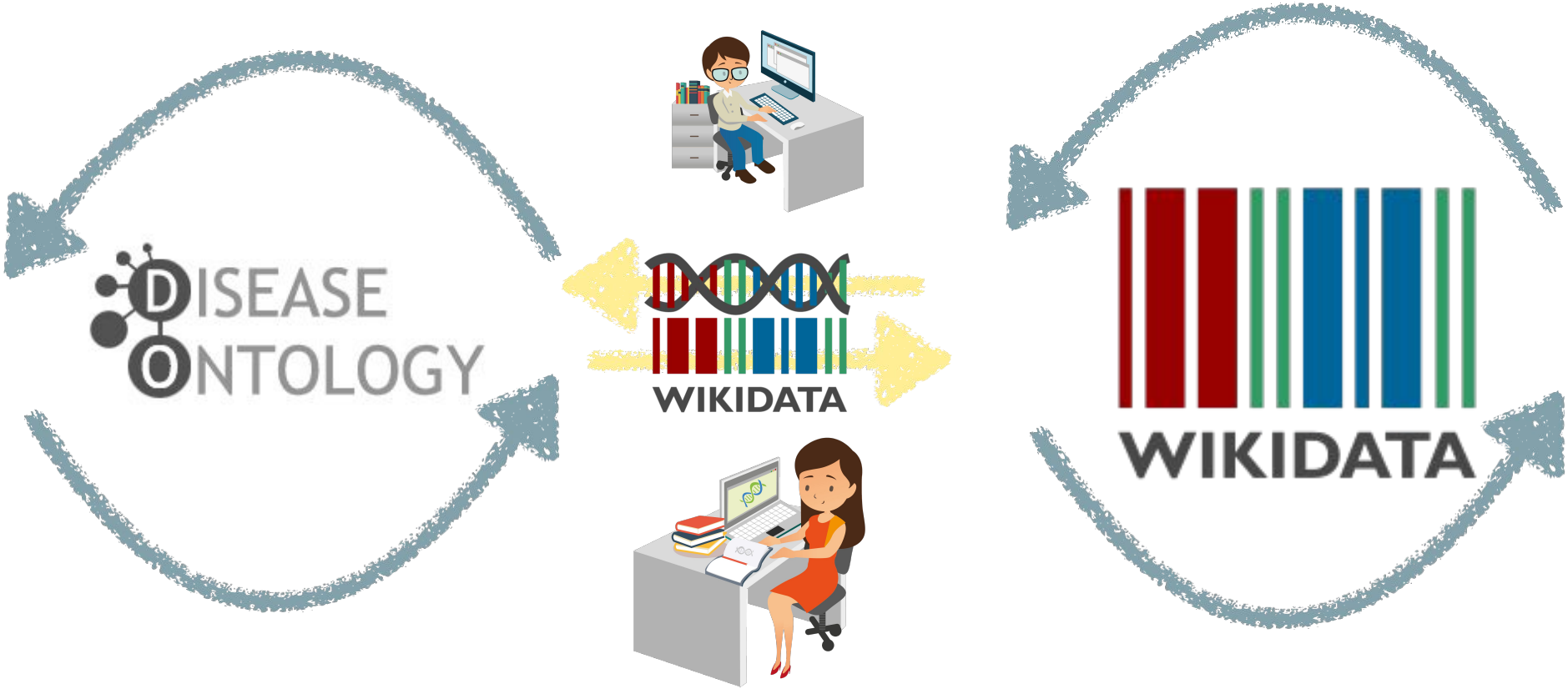
Build Queue

Running Bots

All | **Running** | +

S	Name	Last Success ↑	Last Failure
	ProteinBot_homo_sapiens	1 day 21 hr - #12	N/A
	GOBot_bigmem	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



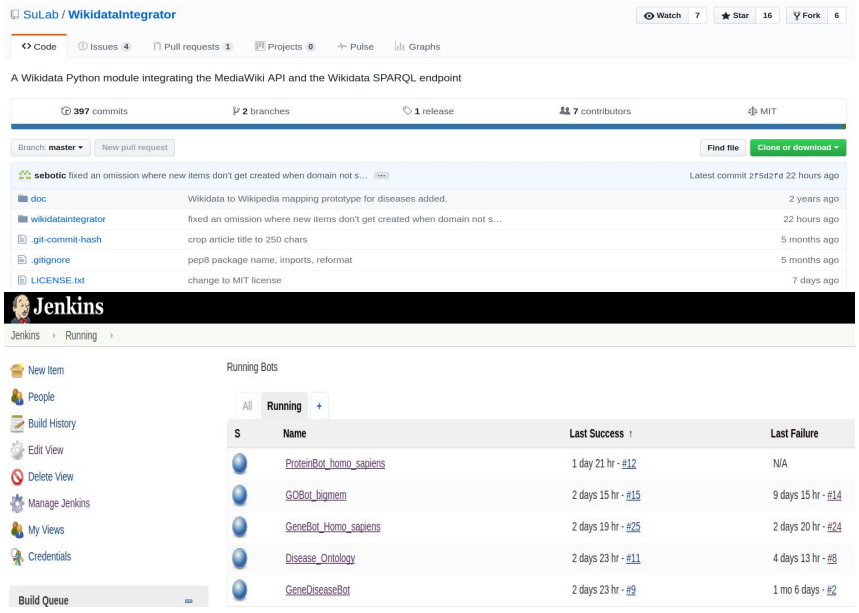
Wikidata Integrator

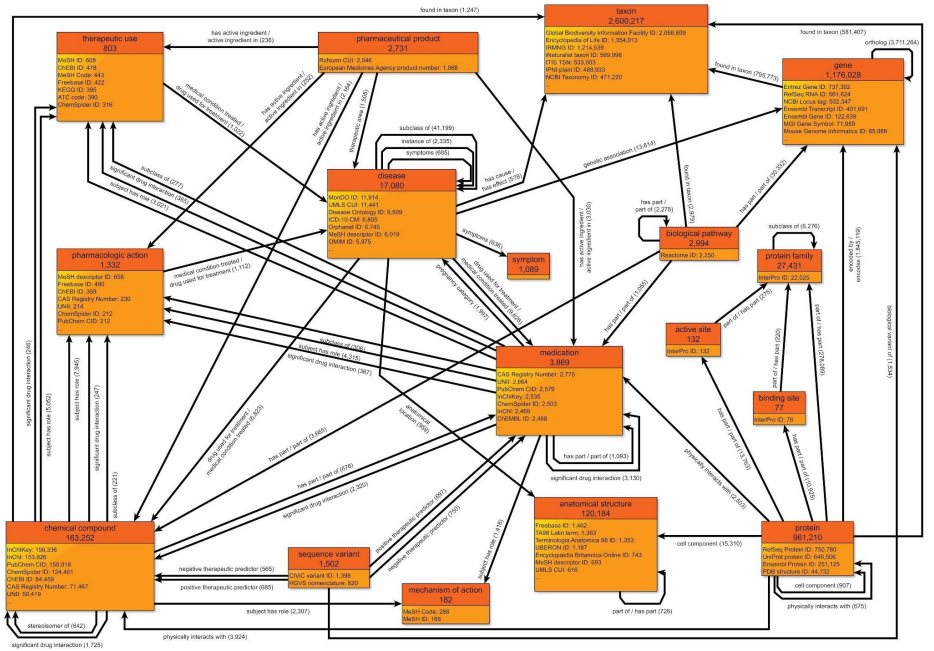
- Python library
- Wikidata Integrator primarily caters Wikidata. Working on other Wikibases or Mediawiki's is possible, but:
  - Wikibase Integrator (https://github.com/LeMyst/WikibaseIntegrator)
  - PyWikibot (https://www.mediawiki.org/wiki/Manual:Pywikibot)
- MIT license
- Github: https://github.com/SuLab/WikidataIntegrator
- Pypi: https://pypi.org/project/wikidataintegrator/
- Community project (i.e. pull requests are welcome)

Example notebooks

- PAWS (https://public.paws.wmcloud.org/User:Andrawaag/WDI_example.ipynb?kernel_name=python3)
- Google (https://colab.research.google.com/drive/1cdth8YUNnlq1zEv22GB3M0t2LIlJhgXT?usp=sharing)
- OBO bot (PAWS)
- DOI helper function (PAWS)

Gene Wiki





Source: https://elifesciences.org/articles/52614