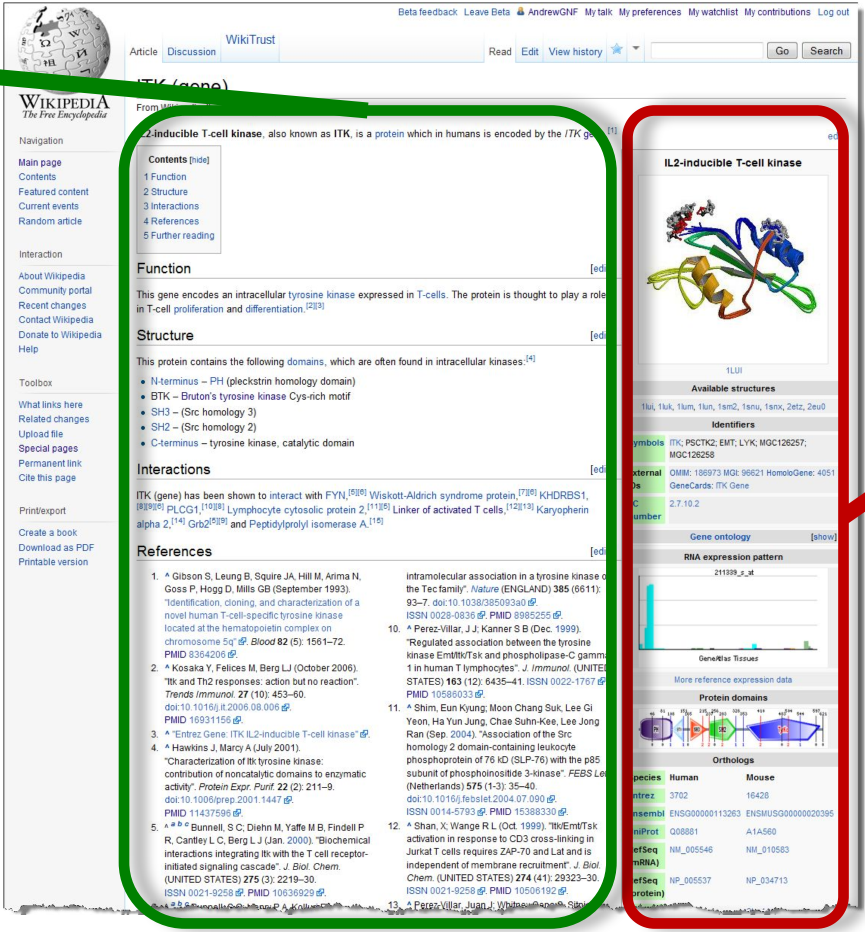
Introduction to Gene Wiki: Using wikidata as a proxy to scholarly articles

Andra Waagmeester[1]

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

The Gene Wiki project, circa 2008

Summarized knowledge via crowdsourcing



ITK (gene)

From Wi

2-inducible T-cell kinase, also known as ITK, is a protein which in humans is encoded by the ITK ge [1]

Contents [hide]

Function [edi

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 [edi

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 [edi

ITK (gene) has been shown to interact with FYN,[5][6] Wiskott-Aldrich syndrome protein,[7][6] KHDRBS1,[8][9][6] PLCG1,[10][8] Lymphocyte cytosolic protein 2,[11][5] Linker of activated T cells,[12][13] Karyopherin alpha 2,[14] Grb2[5][9] and Peptidylprolyl isomerase A.[15]

References [edi

1. ^ 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 8364206.
2. ^ 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.2006.08.006. PMID 16931156.
3. ^ "Entrez Gene: ITK IL2-inducible T-cell kinase".
4. ^ 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.
5. ^ a b c 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 10636929.

intramolecular association in a tyrosine kinase o the Tec family". *Nature* (ENGLAND) **385** (6611): 93–7. doi:10.1038/385093a0. ISSN 0028-0836. PMID 8985255.

10. ^ Perez-Villar, J J; Kanner S B (Dec. 1999). "Regulated association between the tyrosine kinase Emt/Itk/Tsk and phospholipase-C gamm 1 in human T lymphocytes". *J. Immunol.* (UNITED STATES) **163** (12): 6435–41. ISSN 0022-1767 PMID 10586033.
11. ^ Shim, Eun Kyung; Moon Chang Suk, Lee Gi Yeon, Ha Yun Jung, Chae Suhn-Kee, 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 Le* (Netherlands) **575** (1-3): 35–40. doi:10.1016/j.febslet.2004.07.090. ISSN 0014-5793. PMID 15388330.
12. ^ Shan, X; Wange R L (Oct. 1999). "Itk/Emt/Tsk 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 10506192.
13. ^ Perez-Villar, Juan J;

ed

IL2-inducible T-cell kinase

Available structures

1lui, 1luk, 1lum, 1lun, 1sm2, 1snu, 1snx, 2etz, 2eu0

Identifiers

ymbols	ITK; PSCTK2; EMT; LYK; MGC126257; MGC126258
xternal Ds	OMIM: 186973 MGI: 96621 HomoloGene: 4051 GeneCards: ITK Gene
C umber	2.7.10.2

Gene ontology [show]

RNA expression pattern

More reference expression data

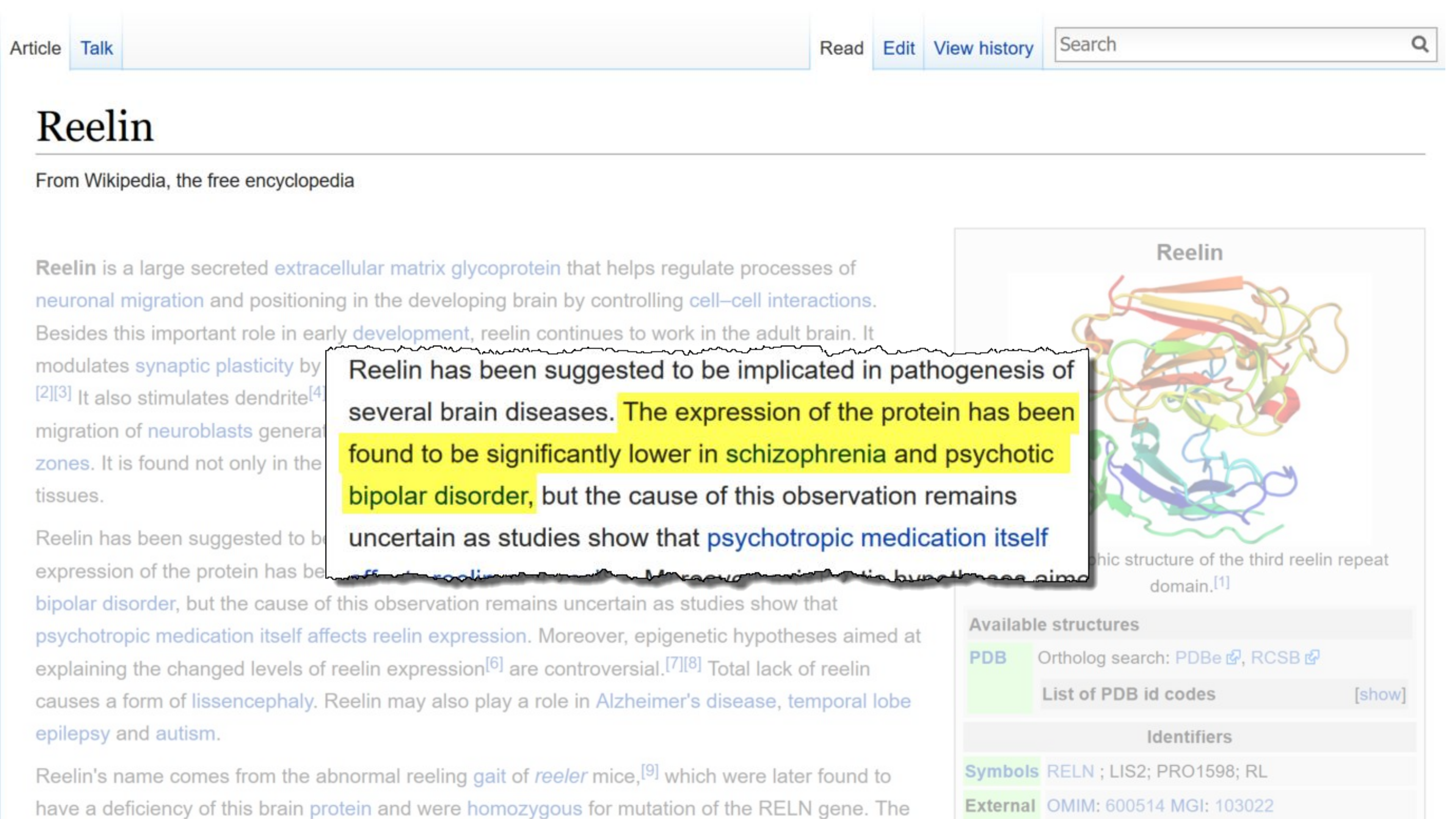
Protein domains

Orthologs

pecies	Human	Mouse
ntrez	3702	16428
nsembl	ENSG00000113263	ENSMUSG00000020395
niProt	Q08881	A1A560
efSeq (mRNA)	NM_005546	NM_010583
efSeq protein)	NP_005537	NP_034713

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 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 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'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

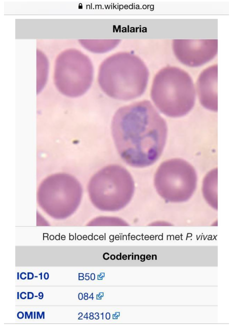
Reelin

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

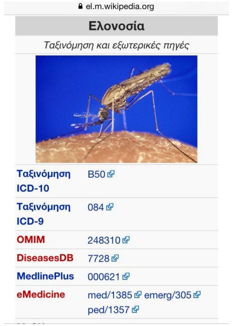
Malaria

Rode bloedcel geïnfecteerd met *P. vivax*

Coderingen

ICD-10	B50
ICD-9	084
OMIM	248310

Greek



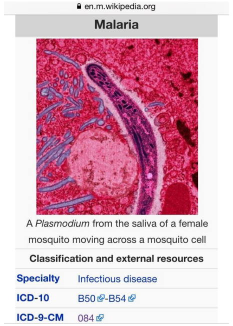
el.m.wikipedia.org

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



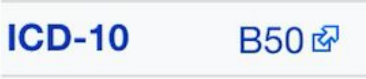
en.m.wikipedia.org

Malaria

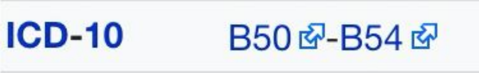
A *Plasmodium* from the saliva of a female mosquito moving across a mosquito cell

Classification and external resources

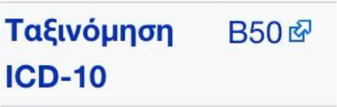
Specialty	Infectious disease
ICD-10	B50-B54
ICD-9-CM	084



ICD-10	B50

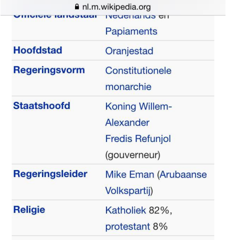


ICD-10	B50-B54



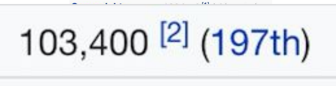
Ταξινόμηση ICD-10	B50

Dutch



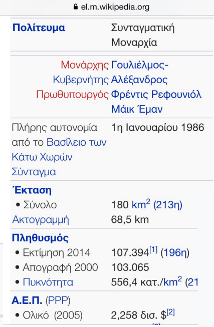
nl.m.wikipedia.org

Officiele landstaal	Nederlands en Papiaments
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)

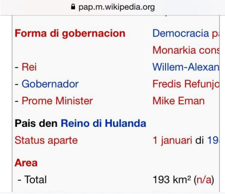
Greek



el.m.wikipedia.org

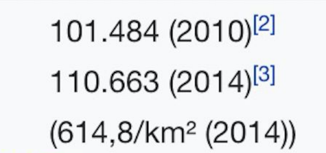
Πολίτευμα	Συνταγματική Μοναρχία
Μονάρχης Κυβερνήτης Πρωθυπουργός	Γουλιέλμος-Αλέξανδρος Φρέντις Ρεφουνιόλ Μάικ Έμαν
Πλήρης αυτονομία από το Βασίλειο των Κάτω Χωρών Σύνταγμα	1η Ιανουαρίου 1986
Έκταση	
• Σύνολο	180 km² (213η)
Ακτογραμμή	68,5 km
Πληθυσμός	
• Εκτίμηση 2014	107.394[1] (196η)
• Απογραφή 2000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

English



pap.m.wikipedia.org

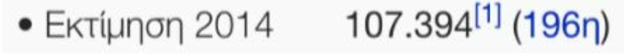
Forma di gobernacion	Democracia p Monarkia cons
- Rei	Willem-Alexan
- Gobernador	Fredis Refunjo
- Prome Minister	Mike Eman
Pais den Reino di Hulanda	
Status aparte	1 januari di 19
Area	
- Total	193 km² (n/a)



101.484 (2010)[2]

110.663 (2014)[3]

(614,8/km² (2014))



• Εκτίμηση 2014	107.394[1] (196η)

Wikidata is to data as Wikipedia is to text

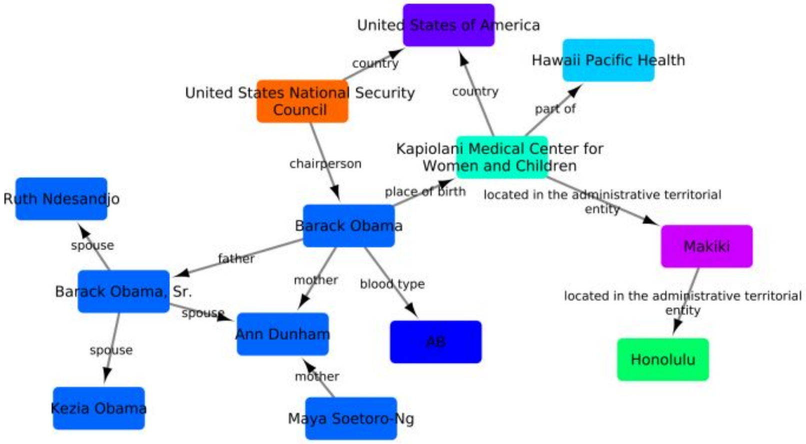
Wikidata is a collaboratively edited knowledge base operated by the Wikimedia Foundation

- 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!

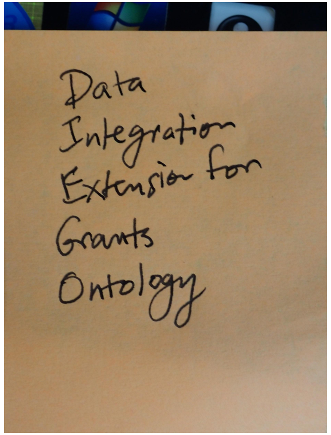
WIKIDATA

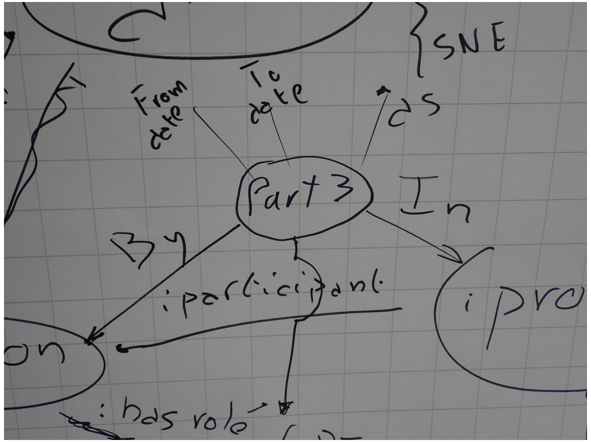
A giant graph of knowledge!

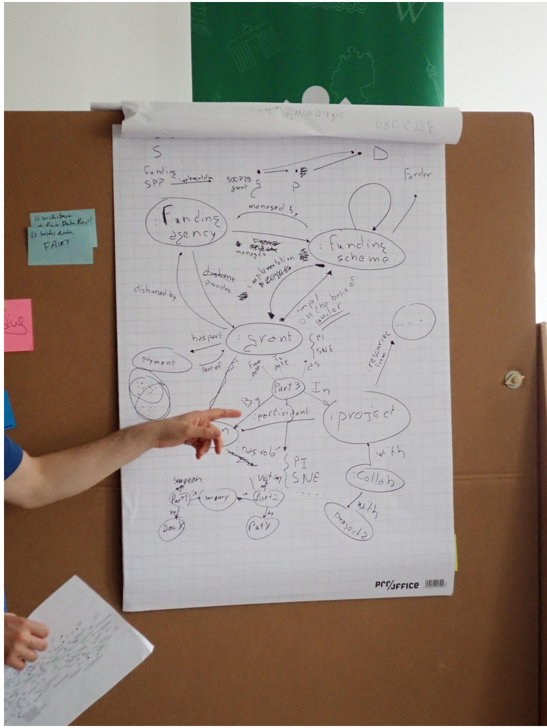


Getting data in..

Community engagement and model discussion



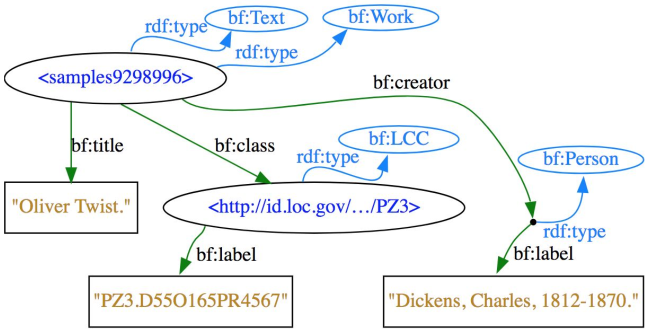


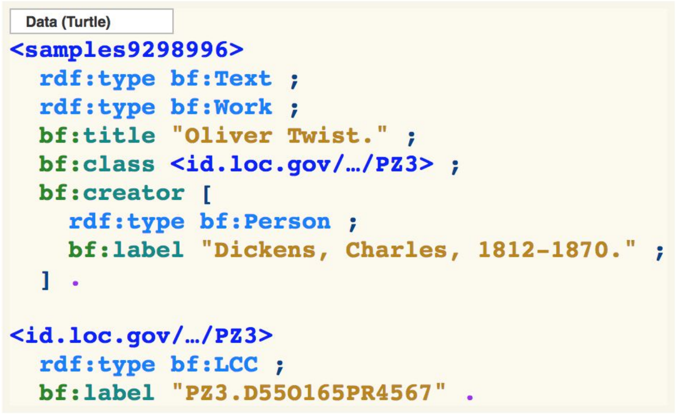


Formally capture and describe model and community consensus

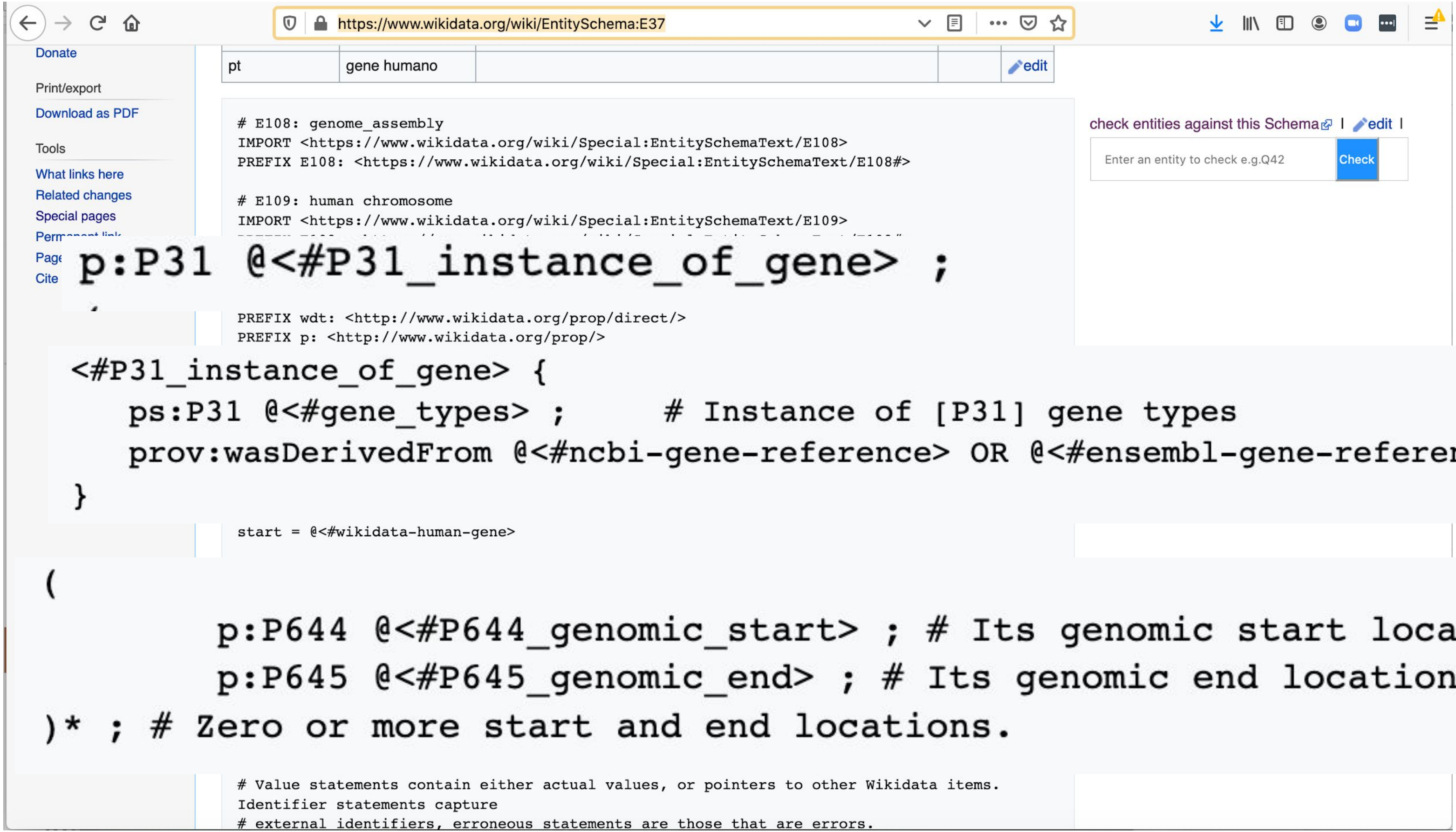
Model development

- 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
Page
Cite
check entities against this Schema | edit |
Enter an entity to check e.g.Q42
Check

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> ;
```

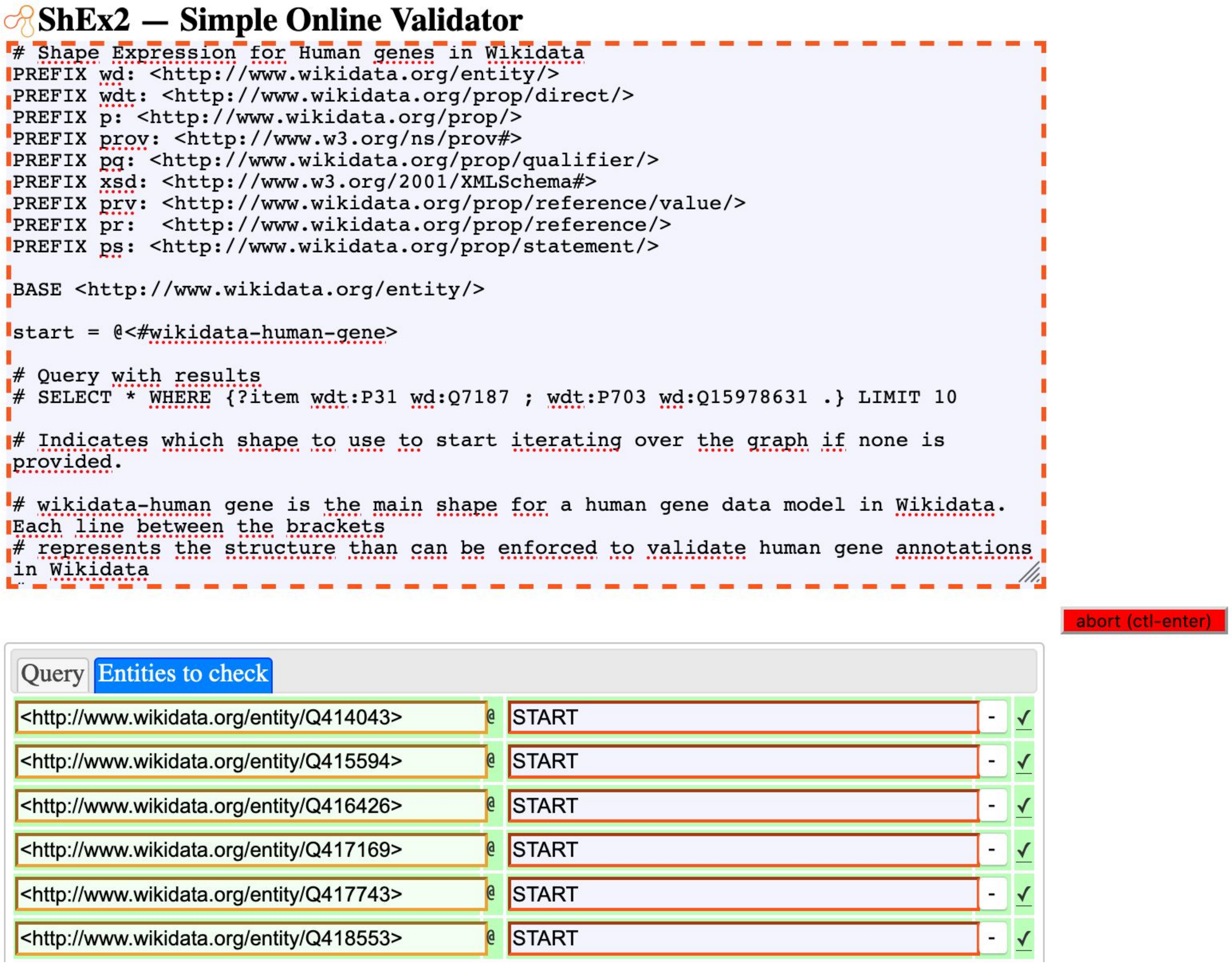
```
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX p: <http://www.wikidata.org/prop/>
```

```
<#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 locatio
          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.
```



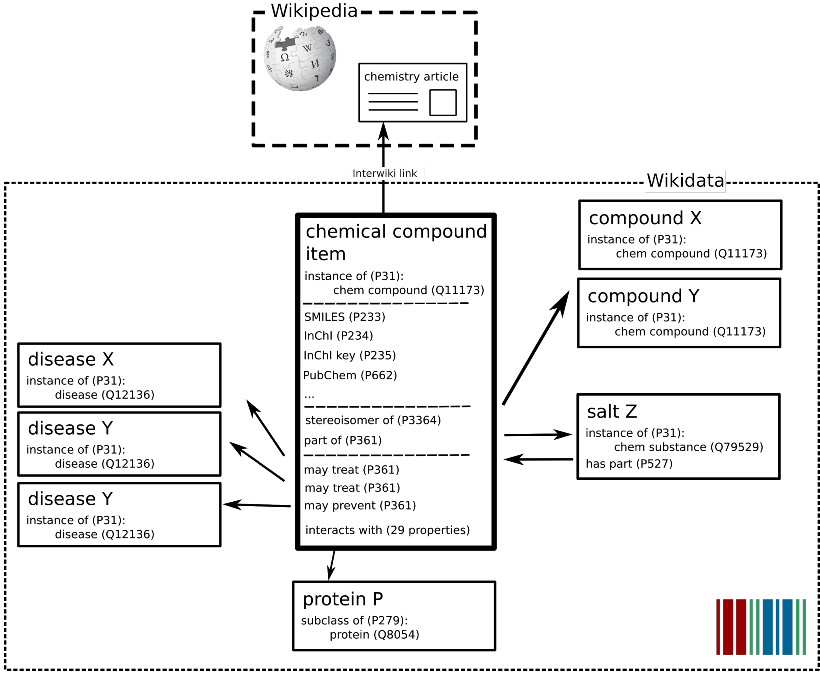
ShEx2 — Simple Online Validator

```
# 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
```

<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



Wikipedia

chemistry article

Interwiki link

Wikidata

chemical compound item

instance of (P31):
chem compound (Q11173)

SMILES (P233)

InChI (P234)

InChI key (P235)

PubChem (P662)

...

stereoisomer of (P3364)

part of (P361)

may treat (P361)

may treat (P361)

may prevent (P361)

interacts with (29 properties)

compound X

instance of (P31):
chem compound (Q11173)

compound Y

instance of (P31):
chem compound (Q11173)

salt Z

instance of (P31):
chem substance (Q79529)

has part (P527)

disease X

instance of (P31):
disease (Q12136)

disease Y

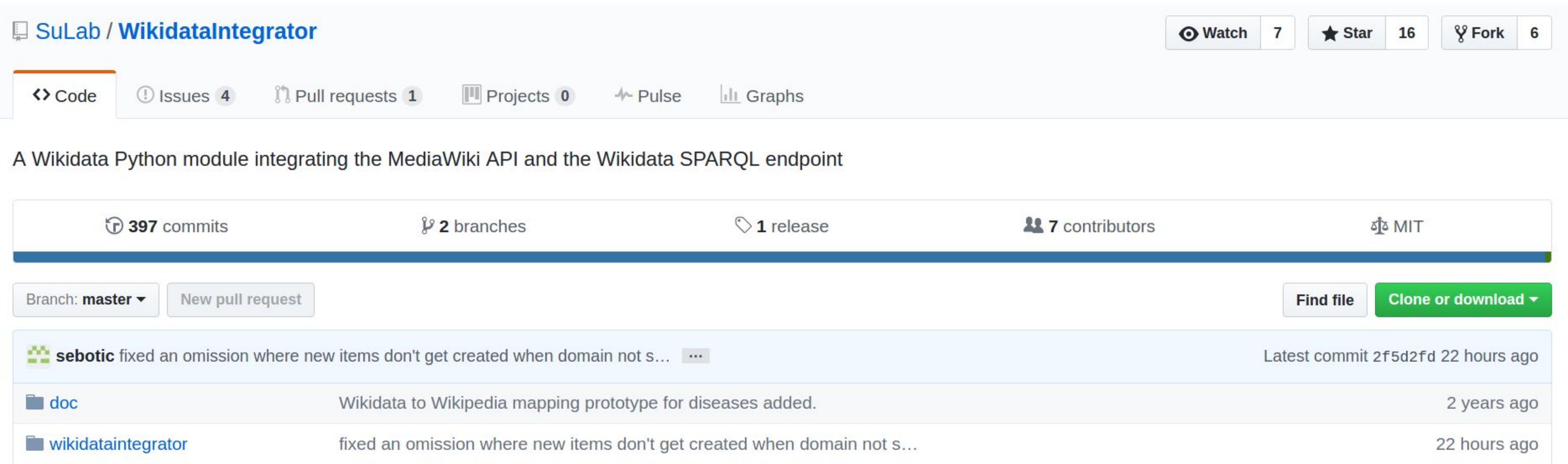
instance of (P31):
disease (Q12136)

disease Y

instance of (P31):
disease (Q12136)

protein P

subclass of (P279):
protein (Q8054)



SuLab / WikidataIntegrator

Watch 7 | Star 16 | Fork 6

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

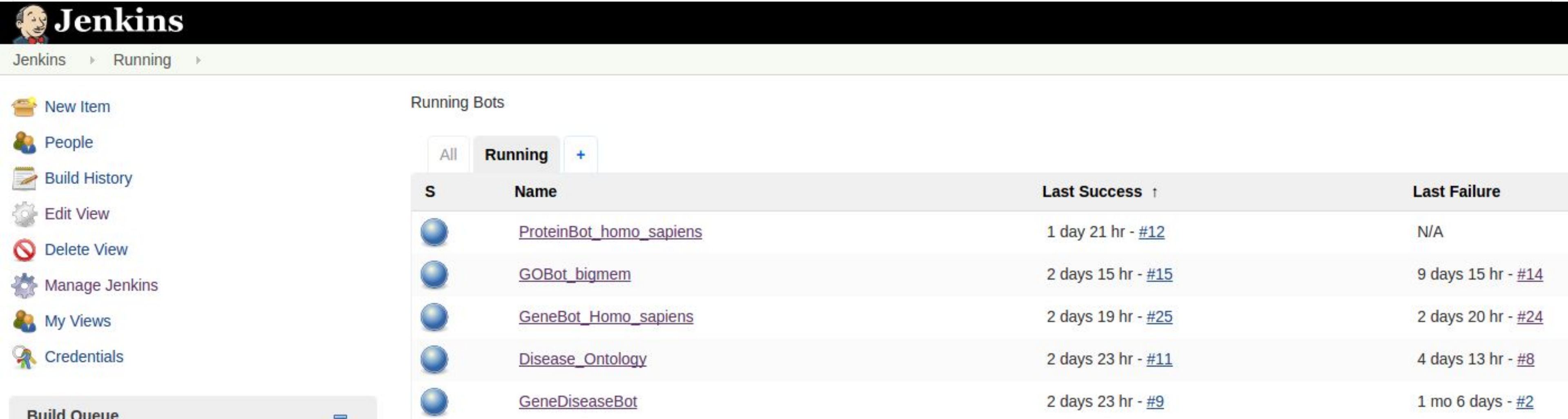
A Wikidata Python module integrating the MediaWiki API and the Wikidata SPARQL endpoint

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

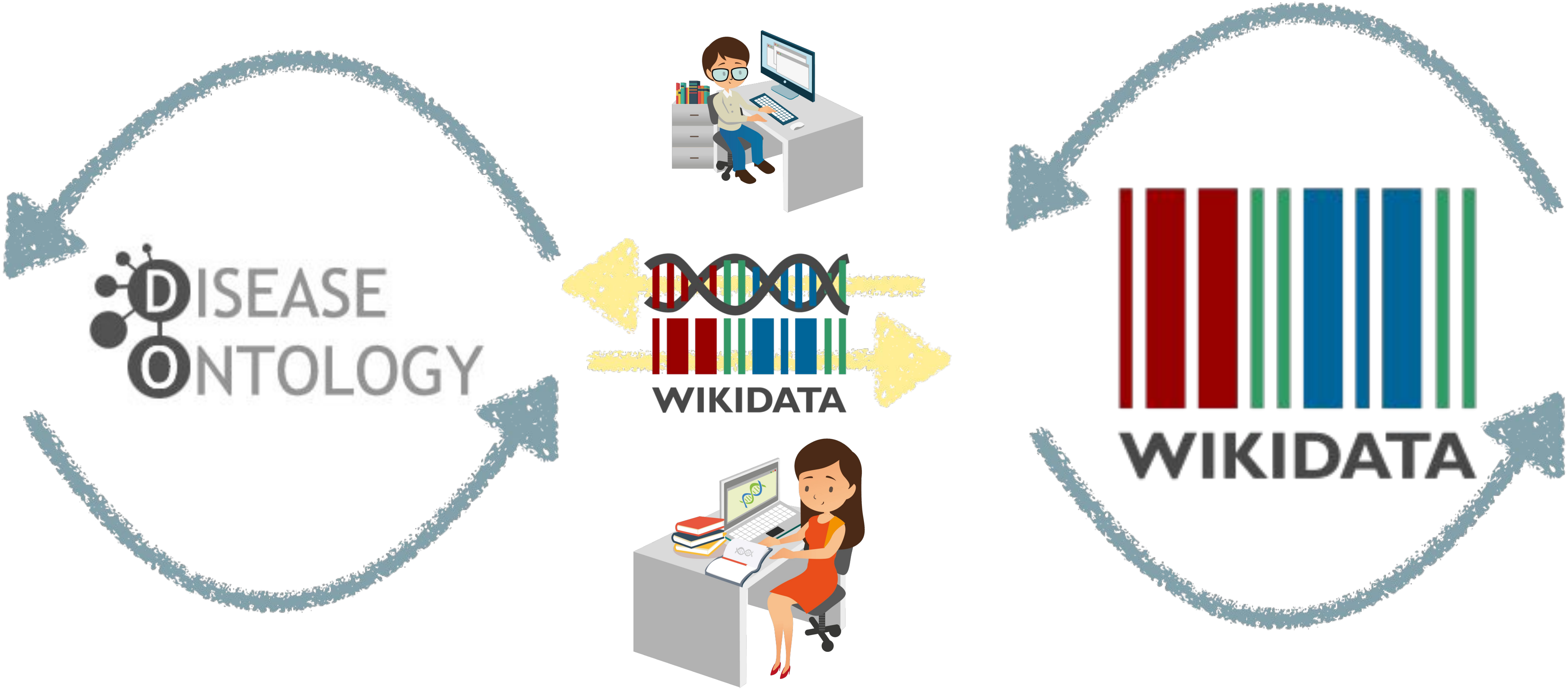
Build Queue

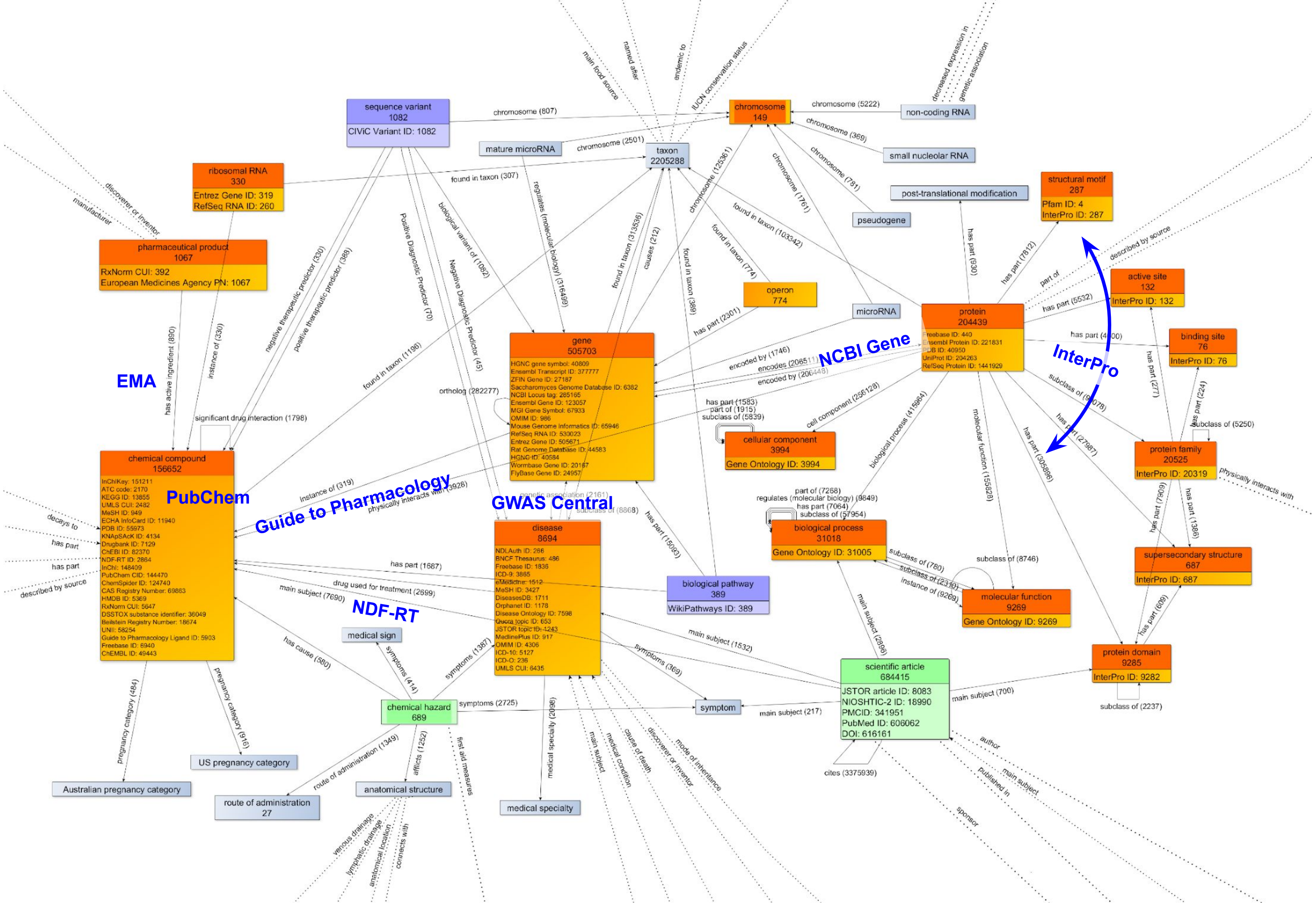
Running Bots

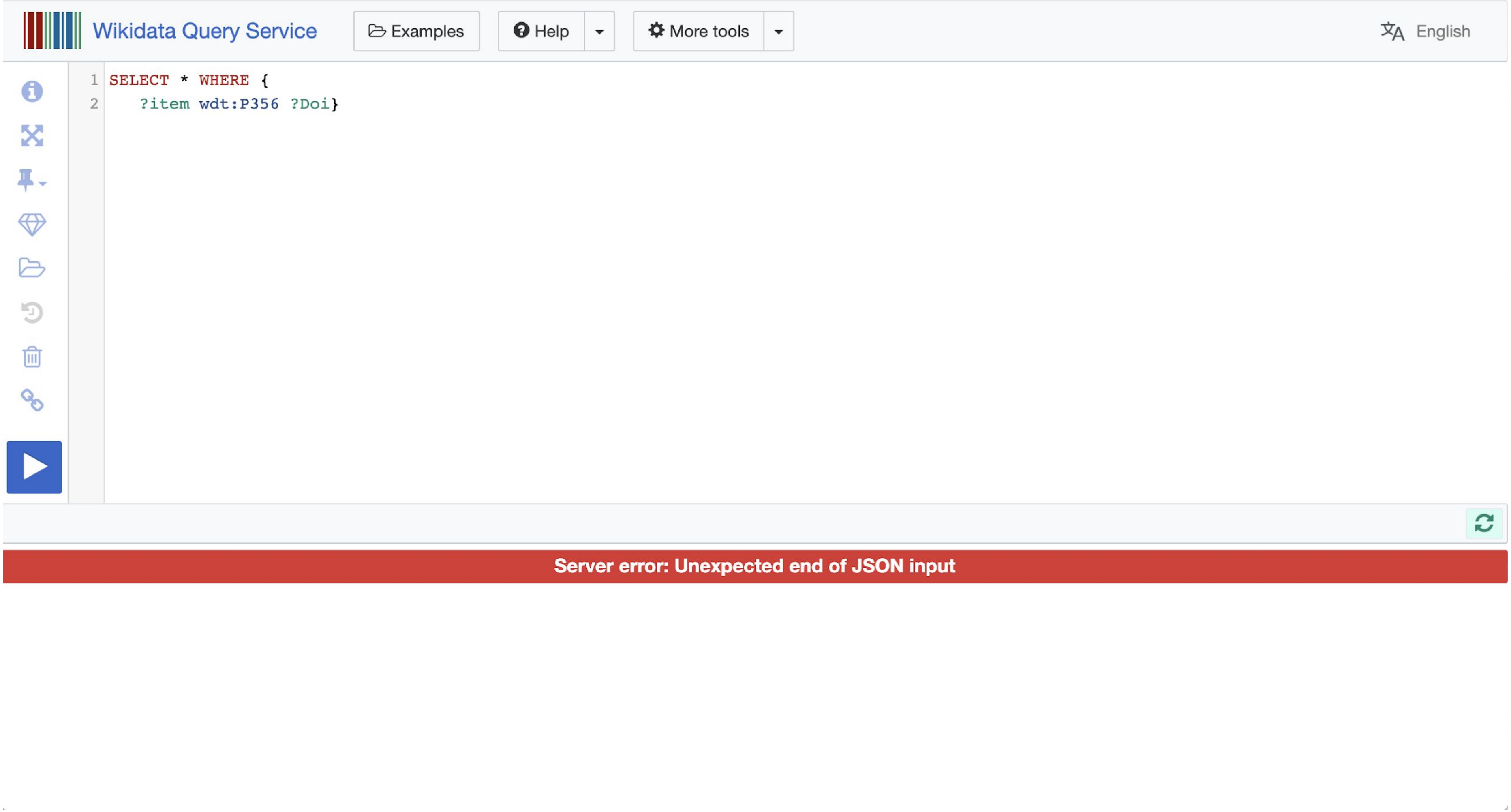
All | Running | +

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




EMA
PubChem
Guide to Pharmacology
GWAS Central
NDF-RT
NCBI Gene
InterPro
sequence variant
1082
CIViC Variant ID: 1082
chromosome
149
non-coding RNA
small nucleolar RNA
mature microRNA
taxon
2205288
ribosomal RNA
330
Entrez Gene ID: 319
RefSeq RNA ID: 260
pharmaceutical product
1067
RxNorm CUI: 392
European Medicines Agency PN: 1067
post-translational modification
pseudogene
structural motif
287
Pfam ID: 4
InterPro ID: 287
operon
774
microRNA
active site
132
InterPro ID: 132
binding site
76
InterPro ID: 76
gene
505703
HGNC gene symbol: 40809
Ensembl Transcript ID: 377777
ZFIN Gene ID: 27187
Saccharomyces Genome Database ID: 6382
NCBI Locus tag: 285165
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
protein
204439
Freebase ID: 440
Ensembl Protein ID: 221831
PDB ID: 40950
UniProt ID: 204263
RefSeq Protein ID: 1441929
cellular component
3994
Gene Ontology ID: 3994
protein family
20525
InterPro ID: 20319
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: 2864
InChI: 148409
PubChem CID: 144470
ChemSpider ID: 124740
CAS Registry Number: 69863
HMDB ID: 5369
RxNorm CUI: 5647
DSSTOX substance identifier: 36049
Beilstein Registry Number: 18674
UNII: 58254
Guide to Pharmacology Ligand ID: 5903
Freebase ID: 6940
ChEMBL ID: 49443
disease
8694
NDLAuth ID: 266
BNCF Thesaurus: 486
Freebase ID: 1836
ICD-9: 3865
eMedicine: 1512
MeSH ID: 3427
DiseasesDB: 1711
Orphanet ID: 1178
Disease Ontology ID: 7598
Quora topic ID: 653
JSTOR topic ID: 1243
MedlinePlus ID: 917
OMIM ID: 4306
ICD-10: 5127
ICD-O: 236
UMLS CUI: 6435
biological process
31018
Gene Ontology ID: 31005
supersecondary structure
687
InterPro ID: 687
biological pathway
389
WikiPathways ID: 389
molecular function
9269
Gene Ontology ID: 9269
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
main food source
named after
endemic to
IUCN conservation status
decreased expression in
genetic association
chromosome (807)
chromosome (5222)
chromosome (369)
chromosome (2501)
chromosome (781)
chromosome (1761)
chromosome (125361)
found in taxon (307)
found in taxon (103342)
found in taxon (774)
found in taxon (313536)
found in taxon (389)
found in taxon (1196)
causes (212)
regulates (molecular biology) (316499)
biological variant of (1082)
Positive Diagnostic Predictor (70)
Negative Diagnostic Predictor (45)
negative therapeutic predictor (330)
positive therapeutic predictor (389)
discoverer or inventor
manufacturer
has active ingredient (890)
instance of (330)
significant drug interaction (1798)
ortholog (282277)
has part (2301)
has part (930)
has part (7812)
part of
described by source
has part (5532)
has part (4400)
has part (277)
has part (224)
subclass of (5250)
encoded by (1746)
encodes (206511)
encoded by (206448)
has part (1583)
part of (1915)
subclass of (5839)
cell component (256128)
biological process (415964)
molecular function (158828)
has part (305896)
has part (27987)
subclass of (9078)
physically interacts with
instance of (319)
physically interacts with (3928)
genetic association (2161)
subclass of (8868)
has part (15093)
part of (7268)
regulates (molecular biology) (9849)
has part (7064)
subclass of (57954)
subclass of (780)
subclass of (2310)
instance of (9269)
subclass of (8746)
has part (7909)
has part (1386)
has part (609)
decays to
has part
has part
described by source
has part (1687)
drug used for treatment (2699)
main subject (7690)
main subject (1532)
symptoms (369)
has cause (580)
symptoms (414)
symptoms (1387)
symptoms (2725)
main subject (217)
main subject (2896)
main subject (700)
subclass of (2237)
pregnancy category (484)
pregnancy category (916)
route of administration (1349)
afflicts (1252)
first aid measures
medical specialty (2098)
main subject
medical condition
cause of death
discoverer or inventor
mode of inheritance
cites (3375939)
author
main subject
published in
sponsor
venous drainage
lymphatic drainage
anatomical location
connects with



```
SELECT * WHERE {
   ?item wdt:P356 ?Doi}
```

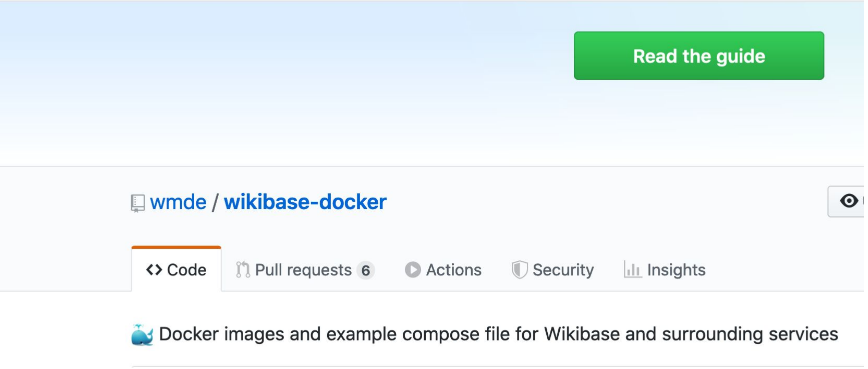
Server error: Unexpected end of JSON input

Wikibase and WBStack

WIKIBASE

github.com/wmde/wikibase-docker

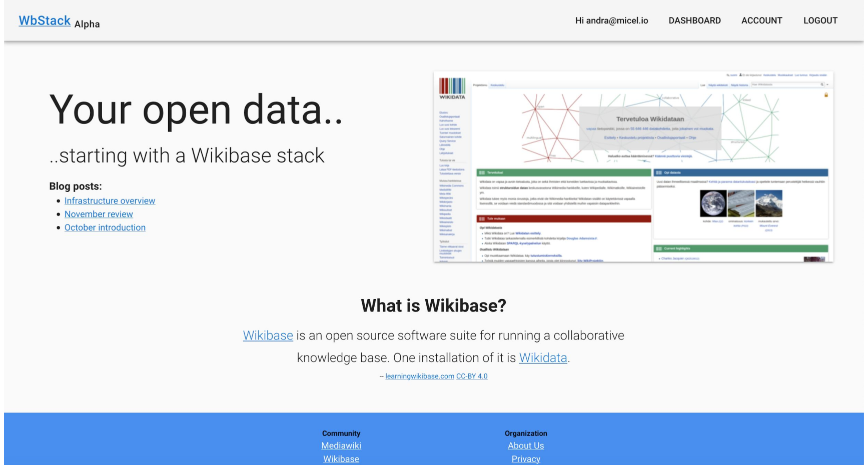
wbstack.com



Read the guide

wmde / wikibase-docker

Docker images and example compose file for Wikibase and surrounding services



Your open data..

..starting with a Wikibase stack

Blog posts:

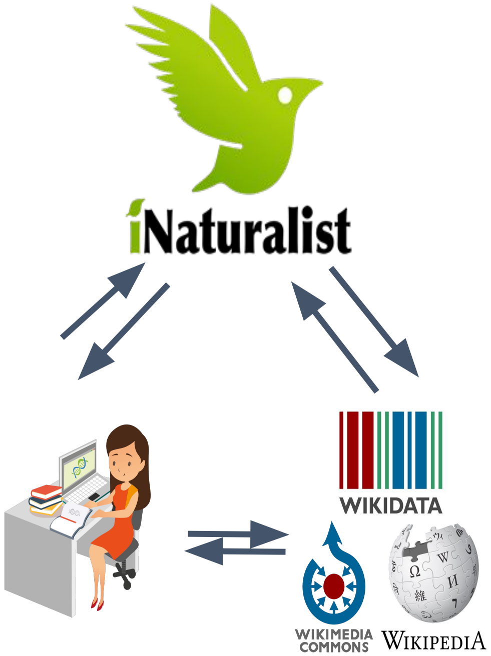
- Infrastructure overview
- November review
- October introduction

What is Wikibase?

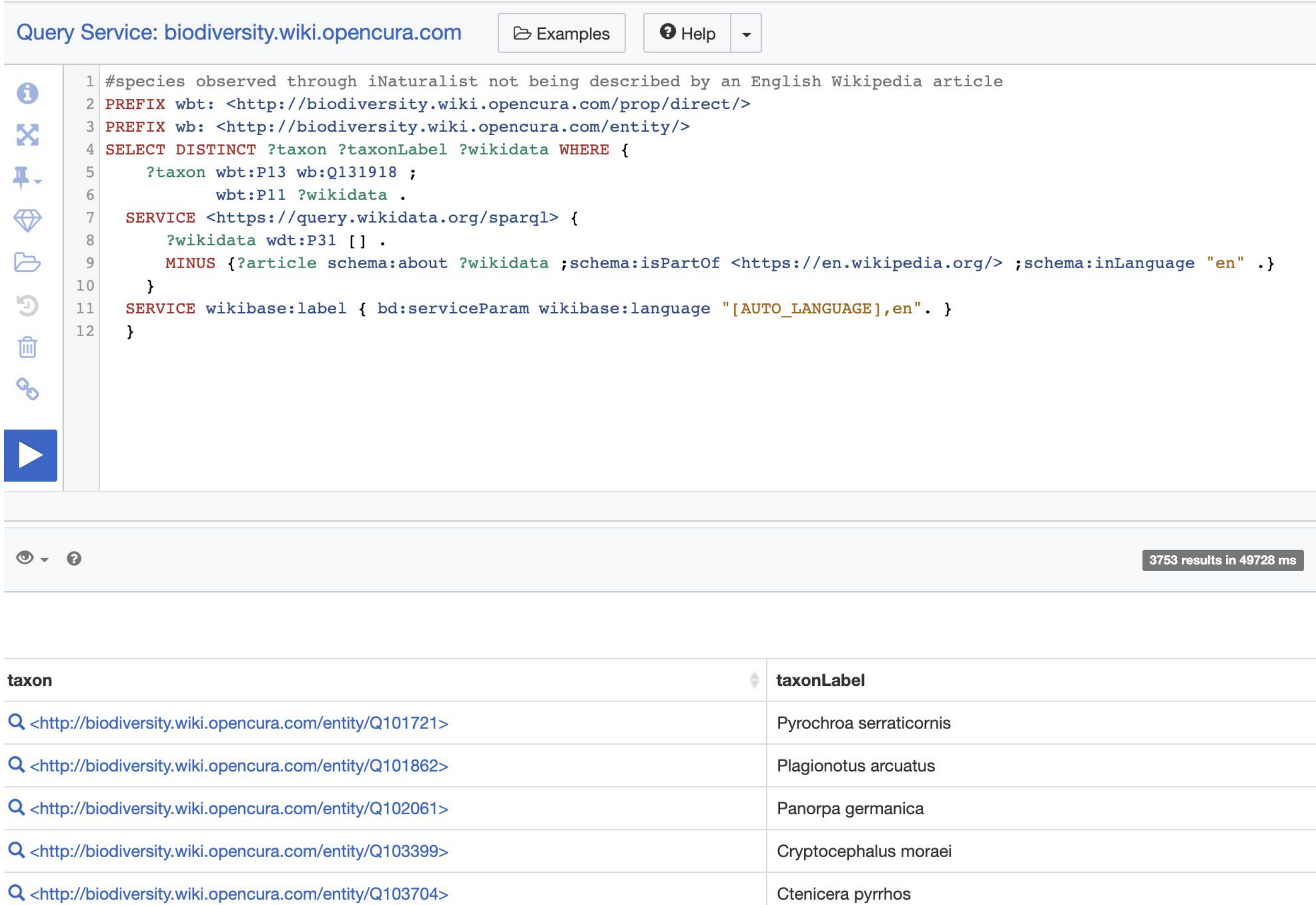
Wikibase is an open source software suite for running a collaborative knowledge base. One installation of it is Wikidata.

– learningwikibase.com CC-BY 4.0

Introducing Wikiproject Biodiversity



https://www.wikidata.org/wiki/Wikidata:WikiProject_Biodiversity



```
#species observed through iNaturalist not being described by an English Wikipedia article
PREFIX wbt: <http://biodiversity.wiki.opencura.com/prop/direct/>
PREFIX wb: <http://biodiversity.wiki.opencura.com/entity/>
SELECT DISTINCT ?taxon ?taxonLabel ?wikidata WHERE {
    ?taxon wbt:P13 wb:Q131918 ;
           wbt:P11 ?wikidata .
  SERVICE <https://query.wikidata.org/sparql> {
      ?wikidata wdt:P31 [] .
      MINUS {?article schema:about ?wikidata ;schema:isPartOf <https://en.wikipedia.org/> ;schema:inLanguage "en" .}
    }
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
  }
```

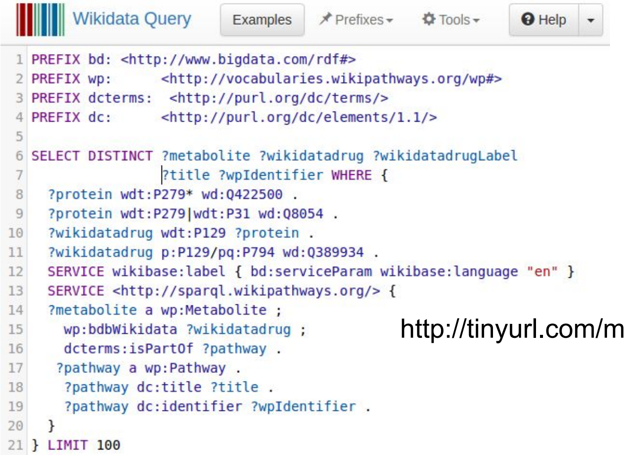
3753 results in 49728 ms

taxon	taxonLabel
<http://biodiversity.wiki.opencura.com/entity/Q101721>	Pyrochroa serraticornis
<http://biodiversity.wiki.opencura.com/entity/Q101862>	Plagionotus arcuatus
<http://biodiversity.wiki.opencura.com/entity/Q102061>	Panorpa germanica
<http://biodiversity.wiki.opencura.com/entity/Q103399>	Cryptocephalus moraei
<http://biodiversity.wiki.opencura.com/entity/Q103704>	Ctenicera pyrrhos

https://tinyurl.com/y5mqqkgl

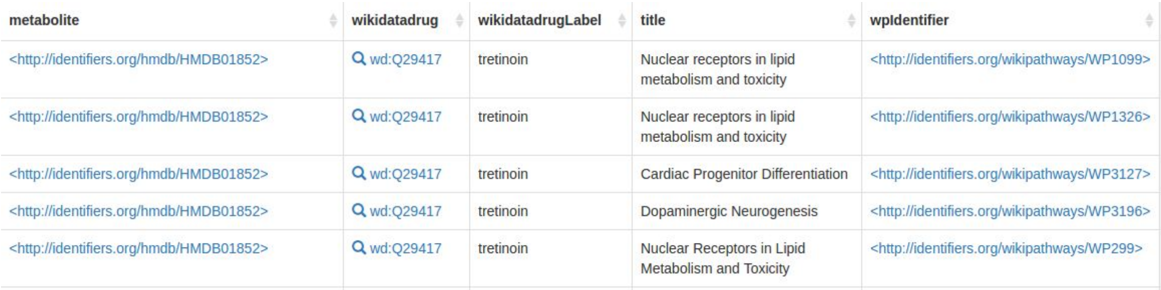
Incomplete... but doesn't need to be complete

Example: Get drugs that act as channel blockers from Wikidata, get the pathways that these drugs are part of from Wikipathways



```
PREFIX bd: <http://www.bigdata.com/rdf#>
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 ?metabolite ?wikidatadrug ?wikidatadrugLabel
                    ?title ?wpIdentifier WHERE {
  ?protein wdt:P279* wd:Q422500 .
  ?protein wdt:P279|wdt:P31 wd:Q8054 .
  ?wikidatadrug wdt:P129 ?protein .
  ?wikidatadrug p:P129/pq:P794 wd:Q389934 .
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
  SERVICE <http://sparql.wikipathways.org/> {
  ?metabolite a wp:Metabolite ;
    wp:bdbWikidata ?wikidatadrug ;
    dcterms:isPartOf ?pathway .
   ?pathway a wp:Pathway .
    ?pathway dc:title ?title .
    ?pathway dc:identifier ?wpIdentifier .
  }
} LIMIT 100
```

http://tinyurl.com/m8g7q7p



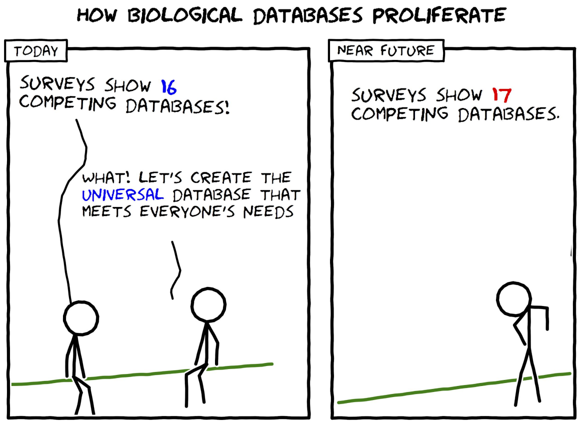
metabolite	wikidatadrug	wikidatadrugLabel	title	wpIdentifier
<http://identifiers.org/hmdb/HMDB01852>	wd:Q29417	tretinoin	Nuclear receptors in lipid metabolism and toxicity	<http://identifiers.org/wikipathways/WP1099>
<http://identifiers.org/hmdb/HMDB01852>	wd:Q29417	tretinoin	Nuclear receptors in lipid metabolism and toxicity	<http://identifiers.org/wikipathways/WP1326>
<http://identifiers.org/hmdb/HMDB01852>	wd:Q29417	tretinoin	Cardiac Progenitor Differentiation	<http://identifiers.org/wikipathways/WP3127>
<http://identifiers.org/hmdb/HMDB01852>	wd:Q29417	tretinoin	Dopaminergic Neurogenesis	<http://identifiers.org/wikipathways/WP3196>
<http://identifiers.org/hmdb/HMDB01852>	wd:Q29417	tretinoin	Nuclear Receptors in Lipid Metabolism and Toxicity	<http://identifiers.org/wikipathways/WP299>

Not trying to create the

Can be linked to greater semantic web

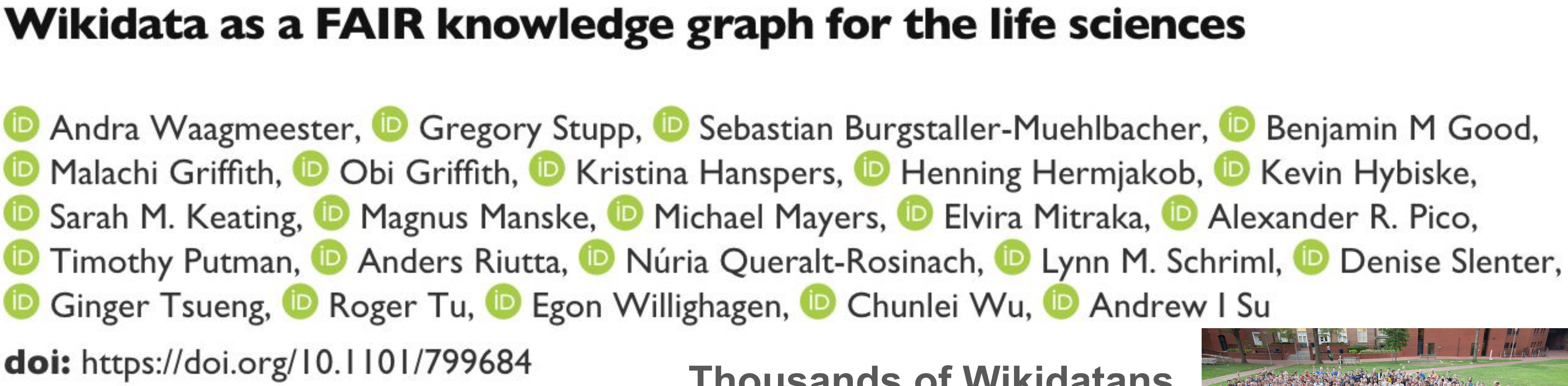
Federated queries: The ability to make one query across multiple sources

Interoperability!


HOW BIOLOGICAL DATABASES PROLIFERATE
TODAY
SURVEYS SHOW 16 COMPETING DATABASES!
WHAT! LET'S CREATE THE UNIVERSAL DATABASE THAT MEETS EVERYONE'S NEEDS
NEAR FUTURE
SURVEYS SHOW 17 COMPETING DATABASES.

https://doi.org/10.1371/journal.pcbi.1005128.g001

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)

References

- A protocol for adding knowledge to Wikidata, a case report (Q90557988)
- Wikidata as a knowledge graph for the life sciences (Q87830400)
- Using Shape Expressions (ShEx) to Share RDF Data Models and to Guide Curation with Rigorous Validation (Q64225211)
- Wikidata as an intuitive resource towards semantic data modeling in data FAIRification (Q59925812)
- Wikidata as a semantic framework for the Gene Wiki initiative (Q23712646)

Hands-on

- Search for a Wikipedia article that cites a DOI; (e.g. https://en.wikipedia.org/wiki/Cetiosauriscus)
- Check if that DOI already exists in Wikidata (e.g. https://w.wiki/j9q)
- in Wikidata on a scholarly article using a DOI; https://boiling-ridge-97667.herokuapp.com/
- Annotate the wikidata item with metadata e.g main subject (e.g. https://www.wikidata.org/wiki/Q100968025)
- Disambiguate its authors (i.e. resolve its author string to author): https://sourcemd.toolforge.org/new_resolve_authors.php
- Explore that article using Scholia: https://scholia.toolforge.org/work/Q100968025