tinyurl.com/DC2019GeneWiki

FAIRification of biomedical knowledge

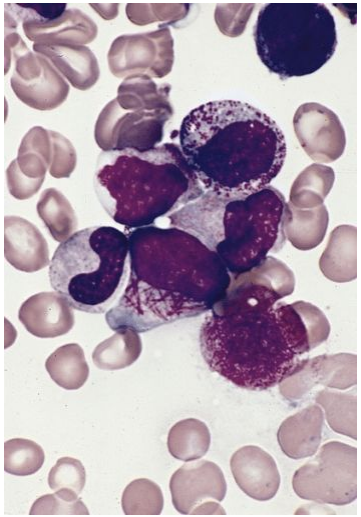
Andra Waagmeester[1]

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

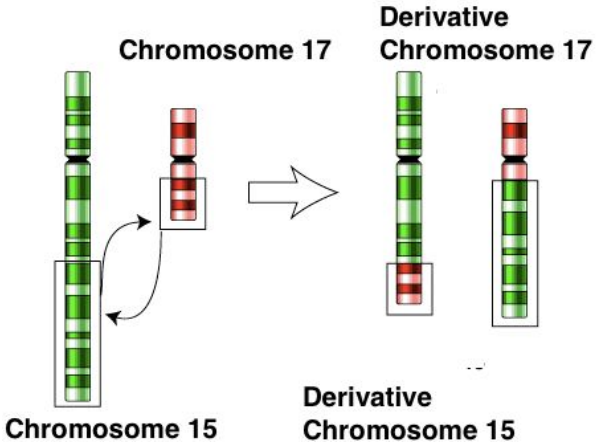
1) Micelio, Ekeren, Antwerp, Belgium

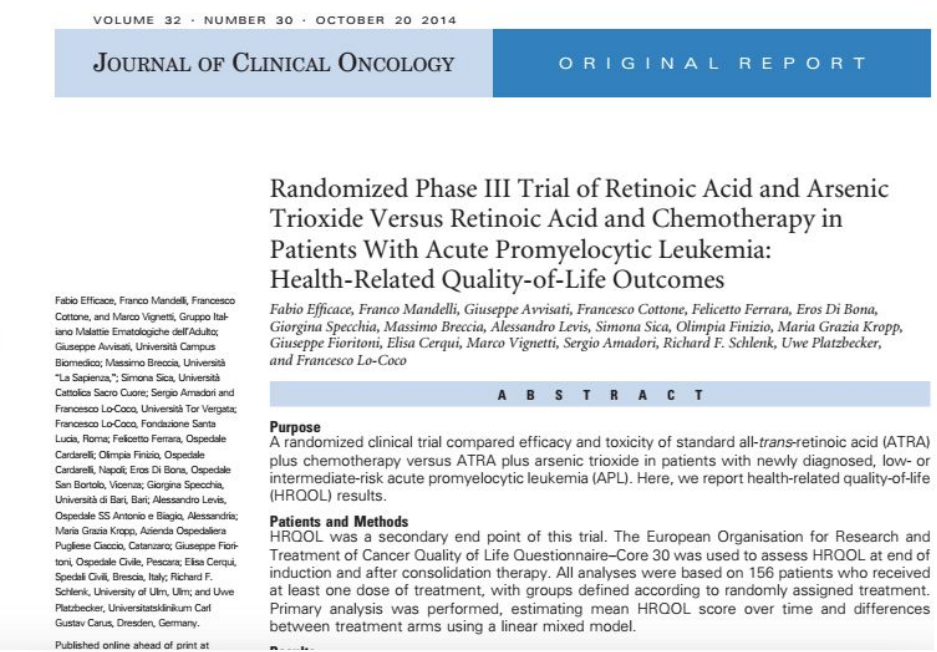
WHY? Acute promyelocytic leukemia

Cancer of the white blood cells



Chromosomal translocation on the RARα gene





VOLUME 32 · NUMBER 30 · OCTOBER 20 2014

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Randomized Phase III Trial of Retinoic Acid and Arsenic Trioxide Versus Retinoic Acid and Chemotherapy in Patients With Acute Promyelocytic Leukemia: Health-Related Quality-of-Life Outcomes

Fabio Efficace, Franco Mandelli, Giuseppe Avvisati, Francesco Cottone, Felicetto Ferrara, Eros Di Bona, Giorgina Specchia, Massimo Breccia, Alessandro Levis, Simona Sica, Olimpia Finizio, Maria Grazia Kropp, Giuseppe Fioritoni, Elisa Cerqui, Marco Vignetti, Sergio Amadori, Richard F. Schlenk, Uwe Platzbecker, and Francesco Lo-Coco

Fabio Efficace, Franco Mandelli, Francesco Cottone, and Marco Vignetti, Gruppo Italiano Malattie Ematologiche dell'Adulto; Giuseppe Avvisati, Università Campus Biomedico; Massimo Breccia, Università "La Sapienza,"; Simona Sica, Università Cattolica Sacro Cuore; Sergio Amadori and Francesco Lo-Coco, Università Tor Vergata; Francesco Lo-Coco, Fondazione Santa Lucia, Roma; Felicetto Ferrara, Ospedale Cardarelli; Olimpia Finizio, Ospedale Cardarelli, Napoli; Eros Di Bona, Ospedale San Bortolo, Vicenza; Giorgina Specchia, Università di Bari, Bari; Alessandro Levis, Ospedale SS Antonio e Biagio, Alessandria; Maria Grazia Kropp, Azienda Ospedaliera Pugliese Ciaccio, Catanzaro; Giuseppe Fioritoni, Ospedale Civile, Pescara; Elisa Cerqui, Spedali Civili, Brescia, Italy; Richard F. Schlenk, University of Ulm, Ulm; and Uwe Platzbecker, Universitatsklinikum Carl Gustav Carus, Dresden, Germany.

Published online ahead of print at

A B S T R A C T

Purpose

A randomized clinical trial compared efficacy and toxicity of standard all-*trans*-retinoic acid (ATRA) plus chemotherapy versus ATRA plus arsenic trioxide in patients with newly diagnosed, low- or intermediate-risk acute promyelocytic leukemia (APL). Here, we report health-related quality-of-life (HRQOL) results.

Patients and Methods

HRQOL was a secondary end point of this trial. The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire–Core 30 was used to assess HRQOL at end of induction and after consolidation therapy. All analyses were based on 156 patients who received at least one dose of treatment, with groups defined according to randomly assigned treatment. Primary analysis was performed, estimating mean HRQOL score over time and differences between treatment arms using a linear mixed model.

Findings support the use of retinoic acid plus arsenic trioxide as prefered first-line treatment

Effects of arsenic trioxide known for decades in China

Original papers were published in the Chinese language and in journals that are obscure even to most Chinese readers

SCIENCE CHINA
Life Sciences

• REVIEW •

June 2013 Vol.56 No.6: 495–502
doi: 10.1007/s11427-013-4487-z

A drug from poison: how the therapeutic effect of arsenic trioxide on acute promyelocytic leukemia was discovered

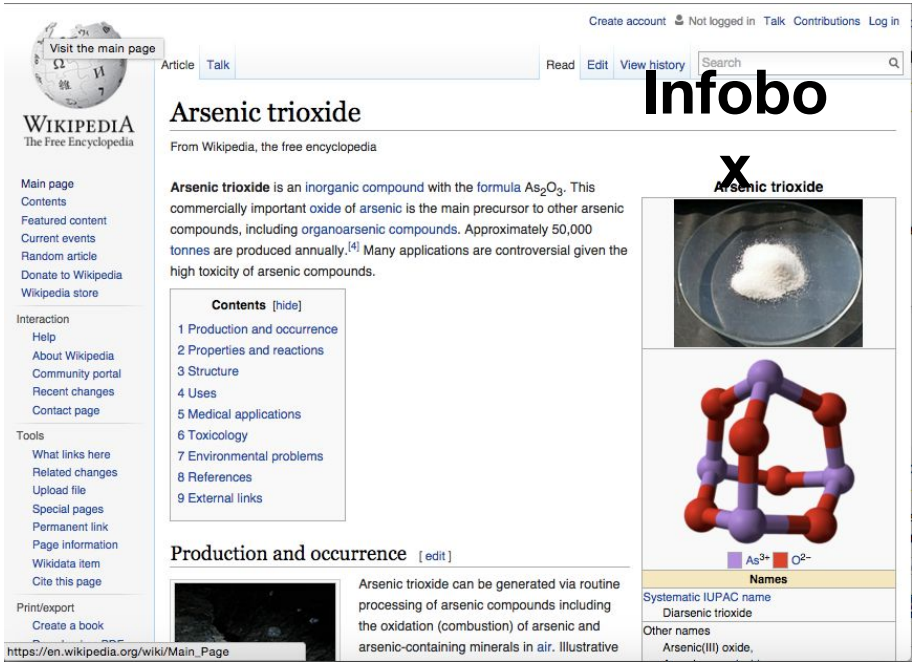
RAO Yi[1*], LI RunHong[2] & ZHANG DaQing[2]

[1]*Peking-Tsinghua Center for Life Sciences at Peking University School of Life Sciences, Beijing 100871, China;*
[2]*Peking University Health Sciences Center, Beijing 100871, China*

Received March 27, 2013; accepted April 5, 2013; published online May 3, 2013

Source: http://jco.ascopubs.org/content/early/2014/09/17/JCO.2014.55.3453

Multilingual Wikipedia & Infoboxes

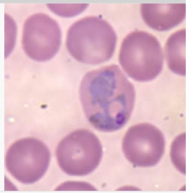


Language independent info boxes

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

	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 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]

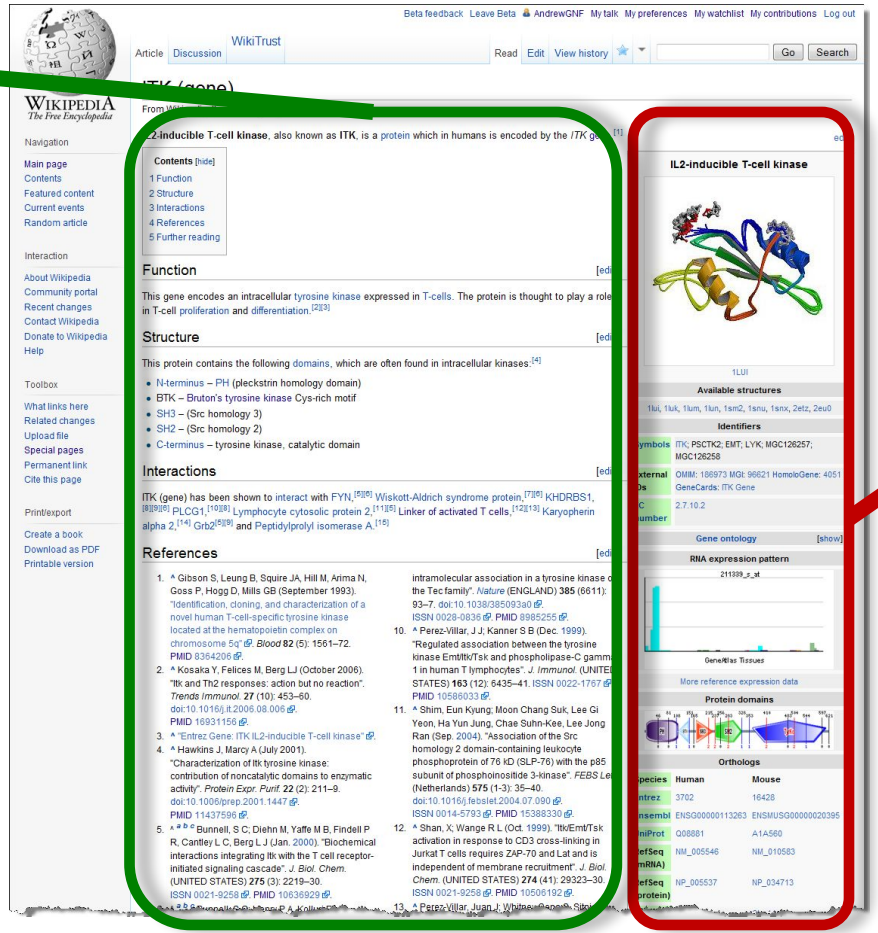
110.663 (2014)[3]

(614,8/km² (2014))

• Εκτίμηση 2014 107.394[1] (196η)

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

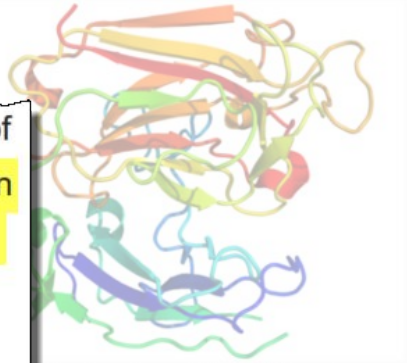
Reelin is a large secreted extracellular matrix glycoprotein that helps regulate processes of neuronal migration and positioning in the developing brain by controlling cell–cell interactions. Besides this important role in early development, reelin continues to work in the adult brain. It modulates synaptic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

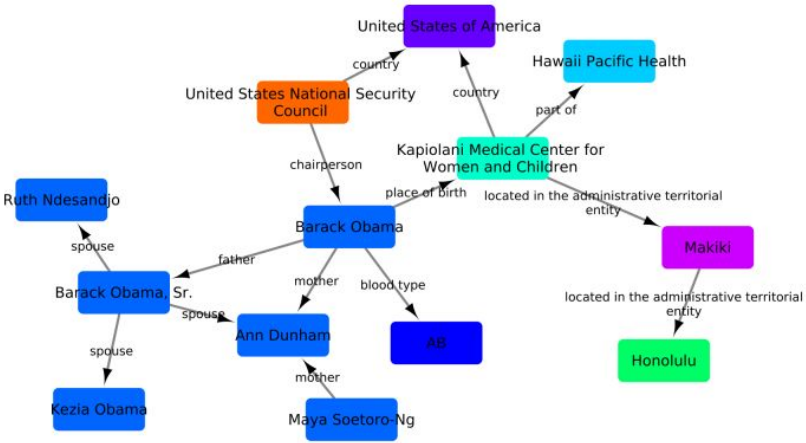
Wikidata is to data as Wikipedia is to text

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!



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			

```
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

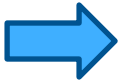
gene	geneLabel	gene	geneLabel	gene	geneLabel	gene	geneLabel
Q14912759	SLC22A5	Q18027370	IGSF3	Q18035037	RAD50	Q18027836	IL6R
Q14914243	PSAP	Q18033424	IL18R1	Q18033919	XPR1	Q18030409	PDE4D
Q14907990	SLC30A8	Q18045382	HPSE2	Q18042132	GSDMB	Q18030185	NOTCH4
Q18035589	C6orf10	Q18027822	IL2RB	Q18036729	RAP1GAP2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

Computing on provenance

"Retrieve genes with GWAS association with asthma and gene product is localized to membrane (non-IEA)"



15 genes

gene	geneLabel		gene	geneLabel		gene	geneLabel
Q14912759	SLC22A5		Q18045382	HPSE2		Q17917202	ERBB4
Q14914243	PSAP		Q18027822	IL2RB		Q18027836	IL6R
Q14907990	SLC30A8		Q14903974	SMAD3		Q18030409	PDE4D
Q18027370	IGSF3		Q18035037	RAD50		Q18030185	NOTCH4
Q18033424	IL18R1		Q18036729	RAP1GAP2		Q18026947	HLA-DQA1

```
6   ?gene wdt:P31 wd:Q7187 ;      # gene is subclass of "gene"
7         wdt:P688 ?protein ;     # gene encodes a protein
8         rdfs:label ?geneLabel .
9   FILTER (lang(?geneLabel) = "en")
10  ?protein p:P681 ?s .                          # protein's cell component statement
11    ?s ps:P681 ?cp .                            # get statement value
12    FILTER NOT EXISTS {?s pq:P459 wd:Q23190881 .}  # determination method is not IEA
13    ?cp wdt:P279*|wdt:P361* wd:Q14349455 .       # statement value is 'part of' or 'subclass of' membrane
14
```

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 ?gen
2  (group_concat(DISTINCT ?geneLabel; separator
3    ?gene wdt:P2293 ?diseaseGA .          # gene
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
```

http://bit.ly/bosc2017_wikidata

Opportunistic integration

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

4 diseases / 6 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

... and show associated 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."

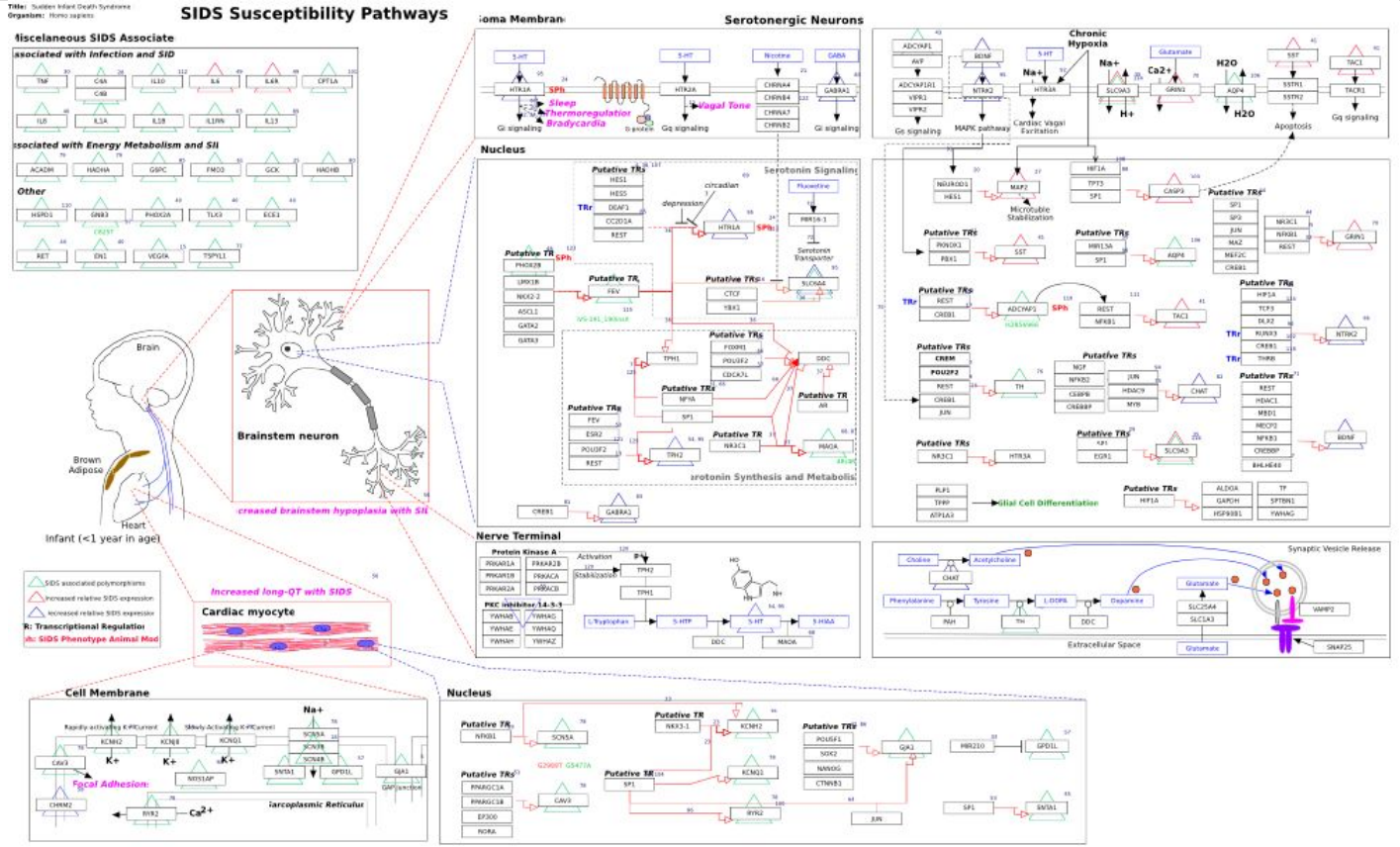
16 genes / 59 pathways

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
....	.....

```
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

From Wikidata to an external SPARQL endpoint (Wikipathways)



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

Wikidata

Wikipathways

Wikidata

Try me....

From a remote SPARQL endpoint to Wikidata

UniProt

SPARQL Downloads Documentation/Help Contact

Your query

Add common prefixes

```
SELECT DISTINCT ?wd_item ?physically_interacts_with ?interactswithLabel ?type ?iri ?uniprot ?text  WHERE  {
  {SELECT * WHERE { ?iri a up:Protein ;
                    up:organism taxon:9606 ;
                    up:annotation ?annotation .
                ?annotation a up:Natural_Variant_Annotation ;
                            rdfs:comment ?text .
                FILTER (CONTAINS(?text, 'loss of function'))
  }}
  SERVICE <https://query.wikidata.org/bigdata/namespace/wdq/sparql> {
    VALUES ?use {wd:Q427492}
  ?wd_item wdt:P352 ?uniprot ;
           wdt:P129 ?physically_interacts_with ;
      wdt:P2888 ?iri ;
      wdt:P703 wd:Q15978631 .
  ?wd_item p:P129 ?phys_interacts_with_node .
  ?phys_interacts_with_node ps:P129 ?physically_interacts_with ;
                            pq:P366 ?use .
  ?physically_interacts_with wdt:P31 ?type ;
                             rdfs:label ?interactswithLabel .
  FILTER (lang(?interactswithLabel) = "en")
  }}
```

UniProt

Wikidata

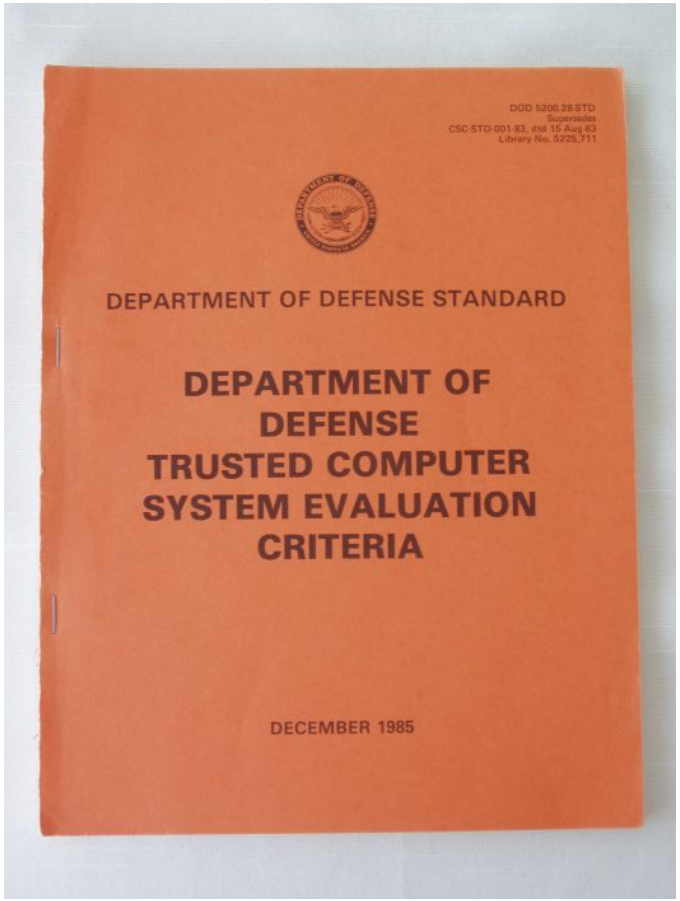
Submit Query Cancel

EMBL-EBI PIR SIB

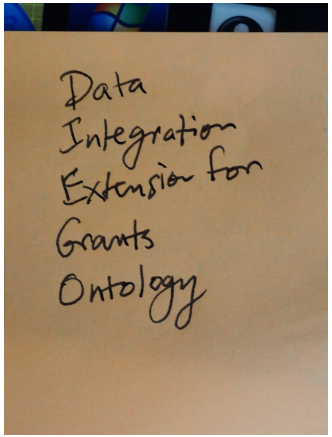
“***People think RDF is a pain because it is complicated. The truth is even worse. RDF is painfully simplistic, but it allows you to work with real-world data and problems that are horribly complicated. While you can avoid RDF, it is harder to avoid complicated data and complicated computer problems.”*** Dan Brickley, Schema.org and Google Libby Miller, BBC

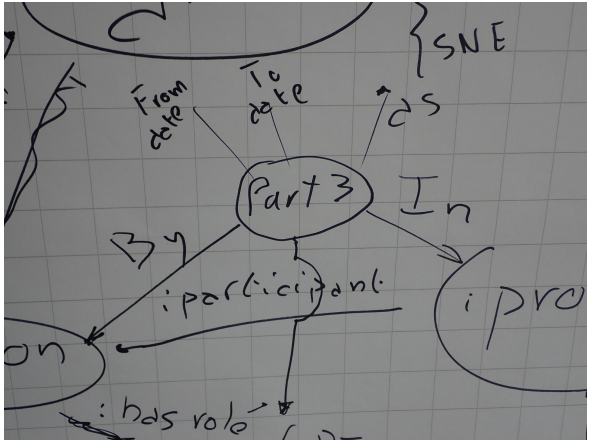
Licenses and Wikidata

Define access rights



Community engagement and model discussion

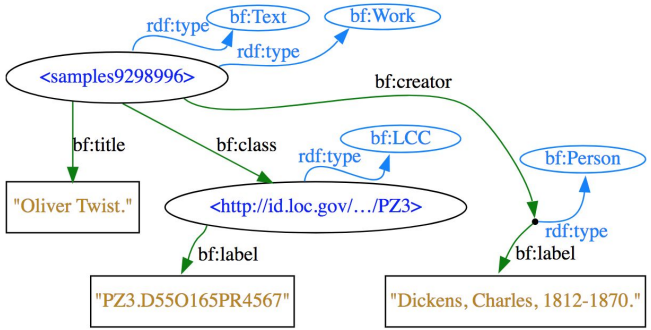




Formally capture and describe model and community consensus

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

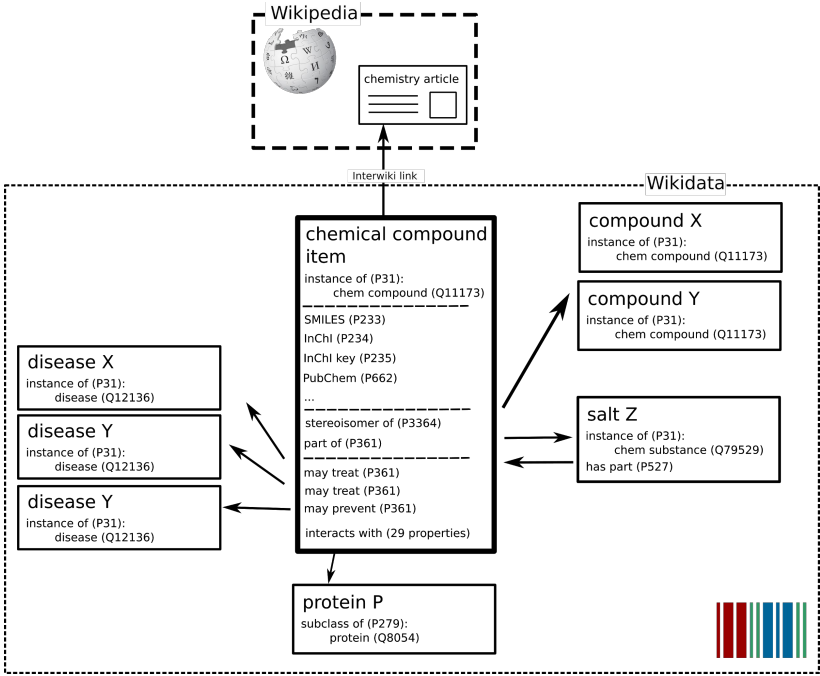


Data (Turtle)

```
<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" .
```

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

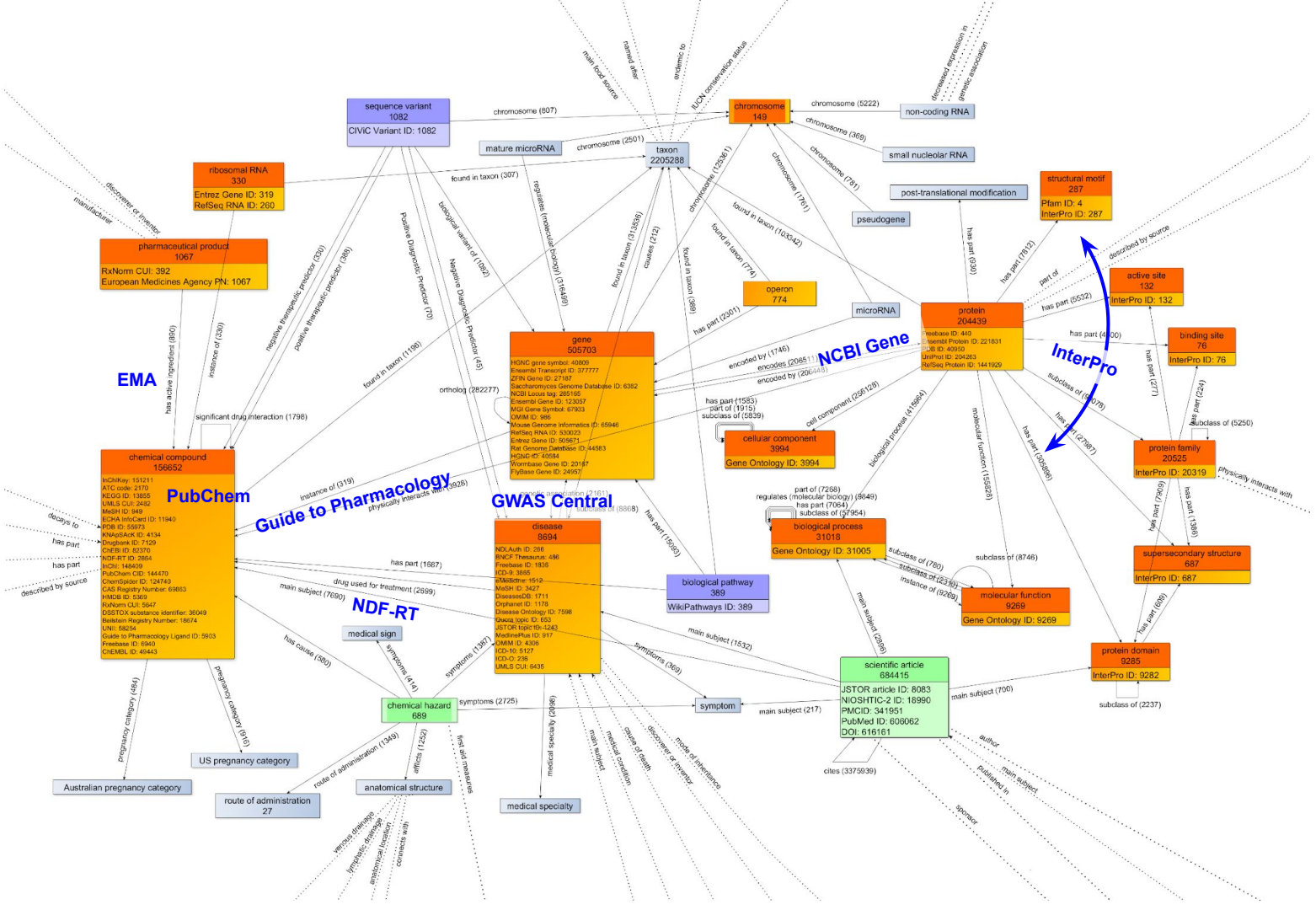
- New Item
- People
- Build History
- Edit View
- Delete View
- Manage Jenkins
- My Views
- Credentials

Build Queue

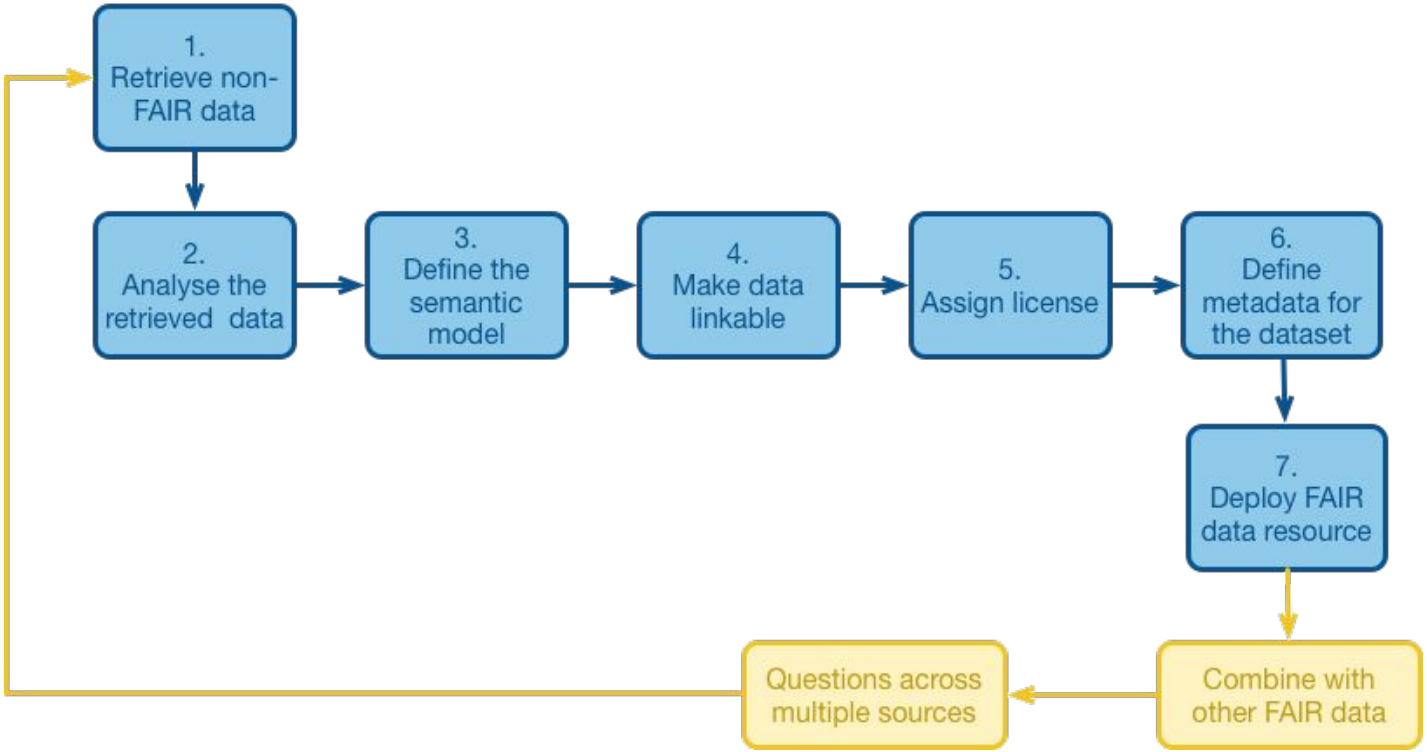
Running Bots

All | **Running** | +

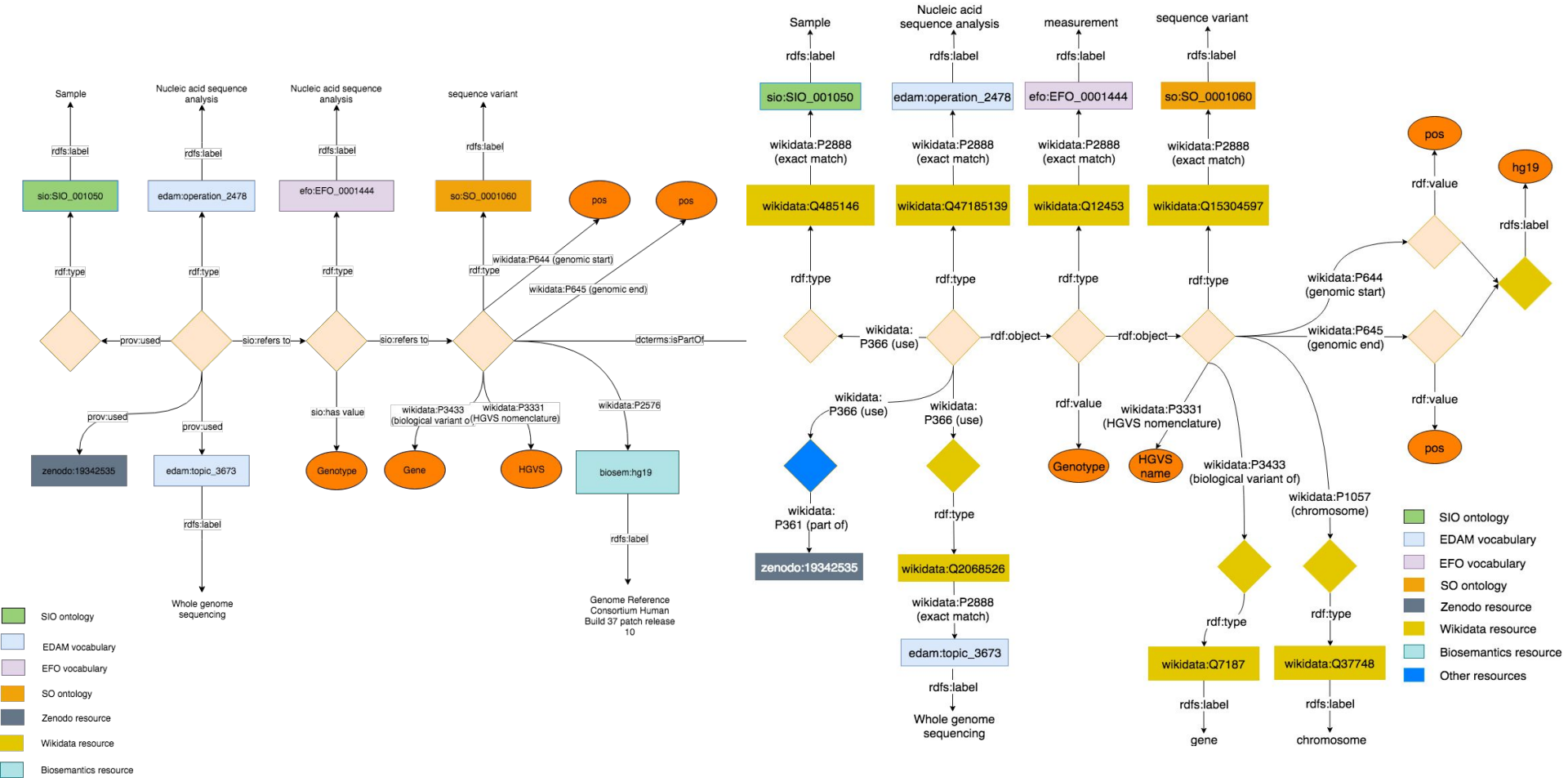
S	Name	Last Success ↑	Last Failure
	ProteinBot_homo_sapiens	1 day 21 hr - #12	N/A
	GOBot_bigmem	2 days 15 hr - #15	9 days 15 hr - #14
	GeneBot_Homo_sapiens	2 days 19 hr - #25	2 days 20 hr - #24
	Disease_Ontology	2 days 23 hr - #11	4 days 13 hr - #8
	GeneDiseaseBot	2 days 23 hr - #9	1 mo 6 days - #2


sequence variant
1082
CIViC Variant ID: 1082
chromosome (807)
chromosome
149
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)
pseudogene
post-translational modification
structural motif
287
Pfam ID: 4
InterPro ID: 287
pharmaceutical product
1067
RxNorm CUI: 392
European Medicines Agency PN: 1067
operon
774
microRNA
protein
204439
Freebase ID: 440
Ensembl Protein ID: 221831
IUPHAR/BPS ID: 40950
UniProt ID: 294263
RefSeq Protein ID: 1441929
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: 830023
Entrez Gene ID: 505671
Rat Genome Database ID: 44583
HGNC ID: 40594
Wormbase Gene ID: 20167
FlyBase Gene ID: 24957
EMA
NCBI Gene
InterPro
PubChem
Guide to Pharmacology
GWAS Central
NDF-RT
chemical compound
156652
InChIKey: 151211
ATC code: 2170
KEGG ID: 13855
UMLS CUI: 2482
MeSH ID: 949
ECHA InfoCard ID: 11940
PDB ID: 55673
KNApSAcK ID: 4134
Drugbank ID: 7129
ChEBI ID: 82370
NDF-RT ID: 2864
InChI: 148409
PubChem CID: 144470
ChemSpider ID: 124740
CAS Registry Number: 69883
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: 286
BNCF Thesaurus: 496
Freebase ID: 1936
ICD-9: 3855
eMedicine: 1512
MeSH ID: 3427
DiseasesDB: 1711
Orphanet ID: 1178
Disease Ontology ID: 7598
Gwoza topic ID: 653
JSTOR topic ID: 1243
MedlinePlus ID: 917
OMIM ID: 4306
ICD-10: 5127
ICD-O: 238
UMLS CUI: 6435
cellular component
3994
Gene Ontology ID: 3994
biological process
31018
Gene Ontology ID: 31005
molecular function
9269
Gene Ontology ID: 9269
biological pathway
389
WikiPathways ID: 389
protein family
20525
InterPro ID: 20319
supersecondary structure
687
InterPro ID: 687
protein domain
9285
InterPro ID: 9282
scientific article
684415
JSTOR article ID: 8083
NIOSHTIC-2 ID: 18990
PMCID: 341951
PubMed ID: 606062
DOI: 616161
medical sign
chemical hazard
689
symptom
US pregnancy category
Australian pregnancy category
route of administration
27
anatomical structure
medical specialty

FAIRification process



Source: https://www.go-fair.org/fair-principles/fairification-process/


Sample
rdfs:label
sio:SIO_001050
rdf:type
Nucleic acid sequence analysis
rdfs:label
edam:operation_2478
rdf:type
Nucleic acid sequence analysis
rdfs:label
efo:EFO_0001444
rdf:type
sequence variant
rdfs:label
so:SO_0001060
rdf:type
pos
pos
wikidata:P644 (genomic start)
wikidata:P645 (genomic end)
prov:used
sio:refers to
sio:refers to
dcterms:isPartOf
prov:used
prov:used
sio:has value
wikidata:P3433 (biological variant o
wikidata:P3331 (HGVS nomenclature)
wikidata:P2576
zenodo:19342535
edam:topic_3673
Genotype
Gene
HGVS
biosem:hg19
rdfs:label
rdfs:label
Whole genome sequencing
Genome Reference Consortium Human Build 37 patch release 10
Sample
rdfs:label
sio:SIO_001050
wikidata:P2888 (exact match)
wikidata:Q485146
rdf:type
Nucleic acid sequence analysis
rdfs:label
edam:operation_2478
wikidata:P2888 (exact match)
wikidata:Q47185139
rdf:type
measurement
rdfs:label
efo:EFO_0001444
wikidata:P2888 (exact match)
wikidata:Q12453
rdf:type
sequence variant
rdfs:label
so:SO_0001060
wikidata:P2888 (exact match)
wikidata:Q15304597
rdf:type
wikidata: P366 (use)
rdf:object
rdf:object
wikidata: P366 (use)
wikidata: P366 (use)
wikidata: P361 (part of)
zenodo:19342535
rdf:type
wikidata:Q2068526
wikidata:P2888 (exact match)
edam:topic_3673
rdfs:label
Whole genome sequencing
rdf:value
Genotype
wikidata:P3331 (HGVS nomenclature)
HGVS name
wikidata:P3433 (biological variant of)
wikidata:P1057 (chromosome)
rdf:type
wikidata:Q7187
rdfs:label
gene
rdf:type
wikidata:Q37748
rdfs:label
chromosome
wikidata:P644 (genomic start)
wikidata:P645 (genomic end)
pos
rdf:value
hg19
rdfs:label
rdf:value
pos
SIO ontology
EDAM vocabulary
EFO vocabulary
SO ontology
Zenodo resource
Wikidata resource
Biosemantics resource
Other resources

Let's make FAIR data. Introducing iNaturalist

iNaturalist

Who you are

You'll need to make an **iNaturalist account** and please only post your own personal observations

Where you saw it

Record both the coordinates of the encounter as well as their accuracy. You can obscure the location from the public

What you saw

Choose a group of organisms like **butterflies** or better yet a specific organism like the **Monarch butterfly**. If you provide evidence you can leave this blank and the **community can help**

When you saw it

Record the date of your encounter, not the date you post it to iNaturalist

Evidence of what you saw

By including evidence like a **photo or sound**, the community can help add, improve, or confirm the identification of the organism you encountered. Help the community by taking clear well framed photos, by including multiple photos from different angles

https://www.inaturalist.org/home

Finally a an example of cross-pollination between two communities