

Synergizing Biomedical Ontologies with Genomics Databases

WSBO-2021: Workshop on Synergizing Biomedical Ontologies

Chris Mungall

Berkeley Lab

cjmungall@lbl.gov

 @chrismungall



monarch
INITIATIVE



*The biomedical field is especially rich with similar ontologies, including BAO and OBI, MONDO and DOID, DTO and PR, CLO and BRENDa, UBERON and FMA, and others. **Extensive mapping efforts** are often required to maintain associations between these ontologies, making integration expensive and cumbersome*

-- from WSBO workshop description

Synergization problem: not unique to ontologies



Homo sapiens (human): 7157

Entry	7157	CDS	T01001
Gene name	TP53, BCC7, BMFS5, LFS1, P53, TRP53		
Definition	(RefSeq) tumor protein p53		
KO	K04451	tumor protein p53	

NCBI-GeneID: 7157

NCBI-ProteinID: NP_000537

OMIM: 191170

HGNC: 11998

Ensembl: ENSG00000141510

Vega: OTTHUMG00000162125

Pharos: P04637 (Tchem)

UniProt: P04637 K7PPA8 Q53GA5

UniProtKB - P04637 (P53_HUMAN)

Organism-specific databases

HGNC ⁱ	HGNC:11998, TP53
MIM ⁱ	191170, gene+phenotype
neXtProt ⁱ	NX_P04637
VEuPathDB ⁱ	HostDB:ENSG00000141510.16



GCID: GC17M007661

External IDs for TP53 Gene

HGNC: 11998 NCBI Entrez Gene: 7157 Ensembl: ENSG00000141510

OMIM®: 191170 UniProtKB/Swiss-Prot: P04637

Previous GeneCards Identifiers for TP53 Gene

GC17P008026, GC17M008311, GC17M007514, GC17M007772, GC17M007512,
GC17M007565, GC17M007465

- “Bioinformaticians would rather share their toothbrushes than their gene identifiers”
 - To paraphrase Michael Ashburner

Ontologies to the rescue... or not....

tp53

Enter a class, e.g. Melanoma

Show advanced options

Search

Matches in 23 ontologies

TP53 - Interlinking Ontology for Biological Concepts (IOBC)

<http://purl.jp/bio/4/id/200906049453084405>

[details](#) - [visualize](#) - 1 more from this ontology

TP53 - Cell Line Ontology [by Mahadevan] (MCCL)

<http://www.semanticweb.org/pallabi/d/ontologies/2014/2/untitled-ontology-11#TP53>

[details](#) - [visualize](#) - 19 more from this ontology

TP53 - Gene Expression Ontology (GEXO)

<http://identifiers.org/ncbigene/7157>

The protein-coding gene TP53 (tumor protein p53) located on the ch

[details](#) - [visualize](#) - 2 more from this ontology

TP53 - Ontology of Genes and Genomes (OGG)

http://purl.obolibrary.org/obo/OGG_3000007157

Other designations: antigen NY-CO-13|cellular tumor antigen p53|m
related protein 53|tumor protein 53

TP53 - Alzheimer's disease ontology (ADO)

<http://scai.fraunhofer.de/AlzheimerOntology#TP53>

[details](#) - [visualize](#)

TP53 - Ontology for MicroRNA Target (OMIT)

http://purl.obolibrary.org/obo/OMIT_0035832

tumor protein p53

[details](#) - [visualize](#)

TP53 - Breast Tissue Cell Lines Ontology (MCBCC)

http://purl.obolibrary.org/obo/MCBCC_0000065#TP53

TP53 gene - Logical Observation Identifier Names and Codes (LOINC)

<http://purl.bioontology.org/ontology/LNC/LP19769-6>

TP53 protein, human - Robert Hoehndorf Version of MeSH (RH-MESH)

<http://phenoebrowser.net/ontologies/mesh/mesh.owl#C495901>

P53 - Holistic Ontology of Rare Diseases (HORD)

https://cdn.rawgit.com/laiasubirats/rarediseasesontology/master/ICD10_1.0.owl#P53

tumor protein p53 - Orphanet Rare Disease Ontology (ORDO)

http://www.orpha.net/ORDO/Orphanet_120204

Ontologies to the rescue... or not....

tp53

Enter a class, e.g. Melanoma

Show advanced options

Search

Matches in 23 ontologies

TP53 - Interlinking Ontology for Biological Concepts (IOBC)

<http://purl.jp/bio/4/id/200906049453084405>

[details](#) - [visualize](#) - 1 more from this ontology

TP53 - Cell Line Ontology [by Mahadevan] (MC)

<http://www.semanticweb.org/pallabi/d/ontologies/2014/2/untitled-ontology-1>

[details](#) - [visualize](#) - 19 more from this ontology

TP53 - Gene Expression Ontology (GEXO)

<http://identifiers.org/ncbigene/7157>

The protein-coding gene TP53 (tumor protein p53) |

[details](#) - [visualize](#) - 2 more from this ontology

TP53 - Ontology of Genes and Genomes (OGG)

http://purl.obolibrary.org/obo/OGG_3000007157

Other designations: antigen NY-CO-13|cellular tumor related protein 53|tumor protein 53

TP53 - Alzheimer's disease ontology (ADO)

<http://scail.fraunhofer.de/AlzheimerOntology#TP53>

[details](#) - [visualize](#)

TP53 - Ontology for MicroRNA Target (OMIT)

http://purl.obolibrary.org/obo/OMIT_0035832

tumor protein p53

[details](#) - [visualize](#)

p53 gene (Code CDR0000042948)

Terms & Properties

Synonym Details

Relationships

Mappings

View All

Terms & Properties

Physician Data Query Code: CDR0000042948

NCI Thesaurus Code: C17359 ([see NCI Thesaurus info](#))

NCI Metathesaurus Link: C0079419 ([see NCI Metathesaurus info](#))

Ontologies to the rescue... or not....

p53 2768677

<http://identifiers.org/ncbigene/2768677>

The protein-coding gene p53 located on the chromosome 3R mapped at 94D

Ontology: Cell Cycle Ontology CCO

tumor protein p53 Orphanet:120204

http://www.orpha.net/ORDO/Orphanet_120204

Ontology: Orphanet Rare Disease Ontology ORDO

Cellular Tumor Antigen p53 NCIT:C17387

http://purl.obolibrary.org/obo/NCIT_C17387

Cellular tumor antigen p53 (393 aa, ~44 kDa) is encoded by the human TP53 both the cell cycle and apoptosis.

Ontology: NCI Thesaurus OBO Edition NCIT

p53 (mol) IMR:0100659

http://purl.obolibrary.org/obo/IMR_0100659

Ontology: TOXic Process Ontology (TXPO) TXPO

Genes, p53 OMIT:0016706

http://purl.obolibrary.org/obo/OMIT_0016706

Ontology: Ontology for MIRNA Target OMIT

TP53 OMIT:0035832

http://purl.obolibrary.org/obo/OMIT_0035832

Ontology: Ontology for MIRNA Target OMIT

TP53 OGG:3000007157

http://purl.obolibrary.org/obo/OGG_3000007157

Ontology: OGG: Ontology of Genes and Genomes OGG

Also in: TXPO

TP53 7157

<http://identifiers.org/ncbigene/7157>

The protein-coding gene TP53 (tumor protein p53) located on the chro

Ontology: Cell Cycle Ontology CCO

Also in: GeXO ReTO ReXO

Tp53 24842

<http://identifiers.org/ncbigene/24842>

The protein-coding gene Tp53 (tumor protein p53) located on the chromosome 10 mapped at 10q2

Ontology: Regulation of Gene Expression Ontology ReXO

Also in: GeXO ReTO ReXO

TP53 11998

<http://identifiers.org/hgnc/11998>

Ontology: Mondo Disease Ontology MONDO

TP53 Gene NCIT:C17359

http://purl.obolibrary.org/obo/NCIT_C17359

This gene plays a critical role in cell cycle regulation and has tumor suppressor activity.

Ontology: NCI Thesaurus OBO Edition NCIT

cellular tumor antigen p53 (human)

PR:P04637

http://purl.obolibrary.org/obo/PR_P04637

A cellular tumor antigen p53 that is encoded in the genome of human.

Ontology: PRotein Ontology (PRO) PR

TP53 (human) gene_symbol:report

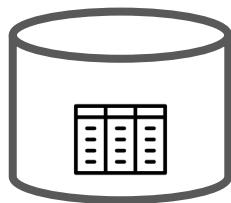
http://www.genenames.org/cgi-bin/gene_symbol_report?hgnc_id=11998

A protein coding gene TP53 in human.

Ontology: PRotein Ontology (PRO) PR

Not just genes!

Database/Knowledgebase-like



CHEMBL
PubChem
DrugBank
RxNorm
MODs, NCBI,
ENSEMBL

RNA central

ENSEMBL
MODs
NCBI
...

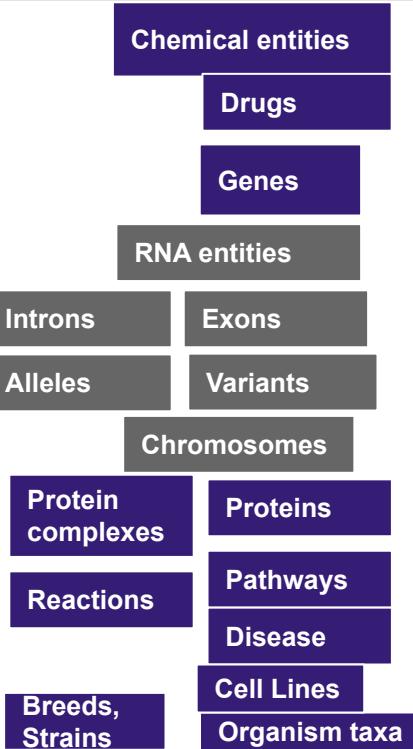
IntAct
BioGRID

Rhea, Reactome,
BioCyc, KEGG

OMIM, Orphanet, ...

Coriell

NCBI, EOL, GTDB



CHEBI
DRON,
OGG,
PR

PRO

GO, MetaCyc

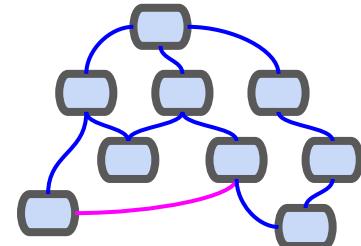
DOID, Mondo, NCIT, ...

CLO, BTO

NCBITaxon, RS

SNOMED
MESH
...

Ontology-like



How do we break out of the loop?

1. I need an ontology-like representation of entities in an omics database or similar
 - a. Use cases:
 - i. Text mining
 - ii. Ontology axiomatization
 - iii. Knowledge Graph or semantic database representation
 - iv. Uniform generic representation
2. I survey existing ontologies
3. I'm an ontologist and therefore opinionated, so I reject everything that exists
4. Make a new ontology
5. Someone else comes along....

Simple Rules for Database Ontologization

Simple Rules for Database Ontologization

1. Don't

Simple Rules for Database Ontologization

1. Don't *if you really must, then...*
2. **Don't mint your own IDs for the same concepts** *no new mappings!*

Simple Rules for Database Ontologization

1. Don't *if you really must, then...*
2. **Don't mint your own IDs** for the same concepts *no new mappings!*
3. **Conflate** rather than proliferate concepts *see Schulz et al **

* Schulz, S., Spackman, K., James, .. (2011).
Scalable representations of diseases in biomedical ontologies. *Journal of Biomedical Semantics*, 2 Suppl 2, S6.

Simple Rules for Database Ontologization

1. Don't *if you really must, then...*
2. **Don't mint your own IDs** for the same concepts *no new mappings!*
3. **Conflate** rather than proliferate concepts *see Schulz et al **
4. **Use a direct automated translation**
 - a. Curation should happen in the upstream database
 - b. Ideally work collaborative with the resource
 - c. *they are the authorities!*

* Schulz, S., Spackman, K., James, .. (2011).
Scalable representations of diseases in biomedical ontologies. *Journal of Biomedical Semantics*, 2 Suppl 2, S6.

Simple Rules for Database Ontologization

1. Don't *if you really must, then...*
2. **Don't mint your own IDs** for the same concepts *no new mappings!*
3. **Conflate** rather than proliferate concepts *see Schulz et al **
4. **Use a direct automated translation**
 - a. Curation should happen in the upstream database
 - b. Ideally work collaborative with the resource
5. **Keep it simple**
 - a. Don't **over-ontologize** or introduce unintended ontological commitments
 - b. Don't introduce your own classification

* Schulz, S., Spackman, K., James, .. (2011).
Scalable representations of diseases in biomedical ontologies. *Journal of Biomedical Semantics*, 2 Suppl 2, S6.

Simple Rules for Database Ontologization

1. Don't *if you really must, then...*
2. Don't mint your own IDs for the same concepts *no new mappings!*
3. Conflate rather than proliferate concepts *see Schulz et al **
4. Use a direct automated translation
 - a. Curation should happen in the upstream database
 - b. Ideally work collaborative with the resource
5. Keep it simple
 - a. Don't over-ontologize or introduce unintended ontological commitments
 - b. Don't introduce your own classification
6. Recognize the limitations of OWL and ontologies *see Rector et al †*

* Schulz, S., Spackman, K., James, .. (2011).
Scalable representations of diseases in biomedical ontologies. *Journal of Biomedical Semantics*, 2 Suppl 2, S6.

† Rector, A., Schulz, S., Rodrigues, J. M., Chute, C. G., & Solbrig, H. (2019). **On beyond Gruber: “Ontologies” in today’s biomedical information systems and the limits of OWL.** *Journal of Biomedical Informatics*: X, 2, 100002.

Case study: NCBI Taxon

Class: *Homo sapiens*

Term IRI: http://purl.obolibrary.org/obo/NCBITaxon_9606

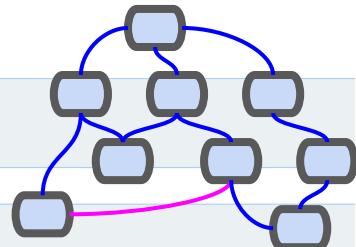
Annotations

- database_cross_reference: GC_ID:1
- has_exact_synonym: man; human
- has_iri_namespace: ncbi_taxonomy
- has_rank: species

Class Hierarchy

Thing

- + root
 - + cellular organisms
 - + Eukaryota
 - + Opisthokonta
 - + Metazoa
 - + Fumetazoa
 - + Bilateria
 - + Deuterostomia
 - + Chordata
 - + Craniata <chordates>
 - + Vertebrata <vertebrates>
 - + Gnathostomata <vertebrates>
 - + Teleostomi
 - + Euteleostomi
 - + Sarcopterygii
 - + Dipnotetrapodomorpha
 - + Tetrapoda
 - + Amniota
 - + Mammalia
 - + Theria <mammals>
 - + Eutheria
 - + Boreoeutheria
 - + Euarchontoglires
 - + Primates
 - + Haplorrhini
 - + Simiiformes
 - + Catarrhini
 - + Hominoidea
 - + Homidae
 - + Hominidae
 - + Homininae
 - + Homo
 - Homo heidelbergensis
 - environmental samples <primates>
 - **Homo sapiens**
 - Homo sapiens neanderthalensis
 - Homo sapiens subsp. 'Denisova'



Credit:

Technical: James Overton, Jim Balhoff, Heiko Dietze, Peter Midford, Frederic Bastian

Content: NCBI taxonomy team

NCBI Taxonomy Browser

Search for [] as [] complete name lock

Display 3 levels using filter: none

Nucleotide Protein Structure Gene Genome Popset SNP Conserved Domains GEO Datasets PubMed Central

Gene SRA Experiments LinkOut BLAST GEO Profiles Protein Clusters Identical Protein Groups SPARCLE

Bio Project Bio Sample Bio Systems Assembly dbVar Registry Host Viral Host Probe

PubChem BioAssay

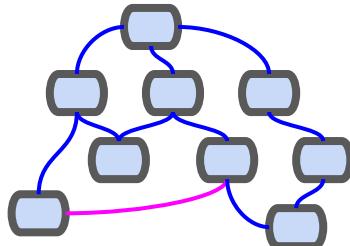
Lineage (full): cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Euarchontoglires; Primates; Haplorrhini; Simiiformes; Catarrhini; Hominoidea; Hominidae; Homininae; Homo

- **Homo sapiens** (human) Click on organism name to get more information.
 - **Homo sapiens neanderthalensis** (Neandertal)
 - **Homo sapiens subsp. 'Denisova'** (Denisova hominin)

- ✓ Reuses ids
- ✓ Direct translation *content requests -> NCBI*
- ✓* No additional classification
 - Additional groupings as *layers*
- ✓ Simple, no over-ontologization

Counter-example: taxa in SNOMED

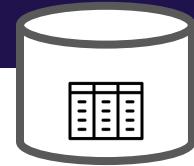
Order Primates
Callitrichidae
Family Cebidae
Family Cercopithecidiae
Family Cheirogaleidae
Family Daubentonidae
Family Galagonidae
Family Hominidae
Genus Homo
Homo sapiens
Genus Pan
Genus Pongo
Gorilla
Family Indriidae
Family Lemuridae
Family Loridae
Family Megaladapidae
Family Tarsiidae
Hylobatidae
Monkey
Johnston's mangabey
Nisnas monkey
Saimiri
Sapajou
Sifaka
Pitheciidae
Order Rodentia
Order Scandentia
Order Sirenia
Pilosa
Proboscidea
Soricomorpha
Tubulidentata
Infraclass Metatheria
Class Reptilia
Fish
Chauliodus macouni
Chilorhinus suensonii
Chlopsis bicolor
Chondrichthyes
Cod
Dysomma anguillare
Eel
Moray eel



Details	Visualization	Notes (0)	Class Mappings (21)	🔗
Preferred Name				Monkey
Synonyms				Monkey (organism)
ID				http://purl.bioontology.org/ontology/SNOMEDCT/54121002
Active				1
altLabel				Monkey (organism)
CASE SIGNIFICANCE ID				900000000000448009 90000000000020002
CTV3ID				X79ov
cui				C0026447
DEFINITION STATUS ID				90000000000074008
Effective time				20020131
notation				54121002
prefLabel				Monkey

- ✗ Reuses ids
- ✗ Direct translation
- ✗ No additional classification *monkey, fish*
- ✓ Simple, no over-ontologization

Case study: Chromosome “ontology”



Screenshot of a web browser showing the OBO library interface for the chromosome ontology. The left panel displays a tree view of chromosome entities, including human chromosomes 1 through 22, X, and Y, along with mouse, rat, and zebrafish chromosomes. The right panel shows the UCSC Genome Browser for the Human Dec. 2013 (GRCh38/hg38) Assembly, specifically focusing on chromosome 1. The browser interface includes a navigation bar, a search bar, and a genomic track viewer.

Not yet submitted
to OBO

- ?? Reuses ids
- ✓ Direct translation
- ✓ No additional classification
- ✓ Simple, no over-ontologization

Credit:

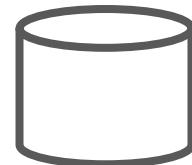
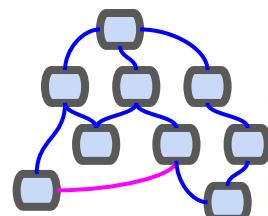
Technical: Kent Shefchek, Tom Conlon, Nico Matentzoglu, Monarch R24

Content: UCSC

<https://github.com/monarch-initiative/monochrom>

Case study: pathway database ontologization

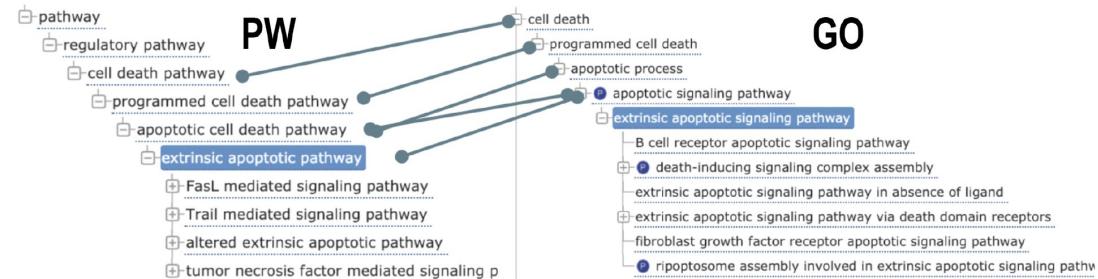
Resource	Elements categorized	Examples of databases that use resource
GO Biological Process (GO-BP)	Pathways/Processes [signaling + biochem]	GO; Reactome; multiple others
RGD Pathway Ontology (PW)	Pathways [signaling and biochemical]	RGD; WikiPathways
NCIT Biological Process (NCIT-BP)	Pathways/Processes [signaling]	NCIT
MetaCyc Pathways (MetaCyc-PWY)	Pathways [biochemical]	BioCyc; EcoCyc; PlantCyc; IMG/M; GO-xrefs
KEGG Pathways (KEGG-P)	Pathways	KEGG; IMG/M
UniPathway (UniPathway)	Pathways [biochemical]	UniProt
SEED SubSystems (SEED-SS)	Pathways [biochemical]	DOE KBase, RAST, PATRIC
IMG Pathways (IMG-PWY)	Pathways [biochemical]	IMG
INOH Pathway Ontology (INOH)	Pathways	INOH
Metabolic Pathways DB (MPW)	Pathways [biochemical]	IMG
GO Molecular Function (GO-MF)	Molecular Activities	GO; Reactome
Enzyme Commission (EC)	Molecular Activities [enzymes]	multiple
Rhea database (RHEA)	Molecular Activities	UniProtKB
MetaCyc Reactions (MetaCyc-RXN)	Molecular Activities	RHEA (xrefs)
SEED Function Roles (SEED-FR)	Molecular Activities	DOE KBase, RAST, PATRIC
KEGG Reactions (KEGG-R)	Molecular Activities	KEGG; IMG/M
IMG Roles (IMG-FR)	Molecular Activities	IMG
ChEBI main ontology (ChEBI)	Chemical Entities	GO-BP [logical axioms]
KEGG Compounds (KEGG-C)	Chemical Entities	ChEBI [xrefs]
HMDB	Chemical Entities	various
ChEBI roles (ChEBI-R)	Roles	ChEBI; GO-BP [logical axioms]
GO Cellular Component (GO-CC)	Cell Components	GO; Reactome
CCO (BioCyc-CCO)	Cell Components	BioCyc



Integrating pathway databases and ontologies

Pathway and Reaction Ontologies

- In process of mapping
 - Available in SSSOM registry soon
- See Jim's talk in ~1hr



Credit:

Tom Hayman (RGD / PW)

Jim Balhoff (RENCI / GO)

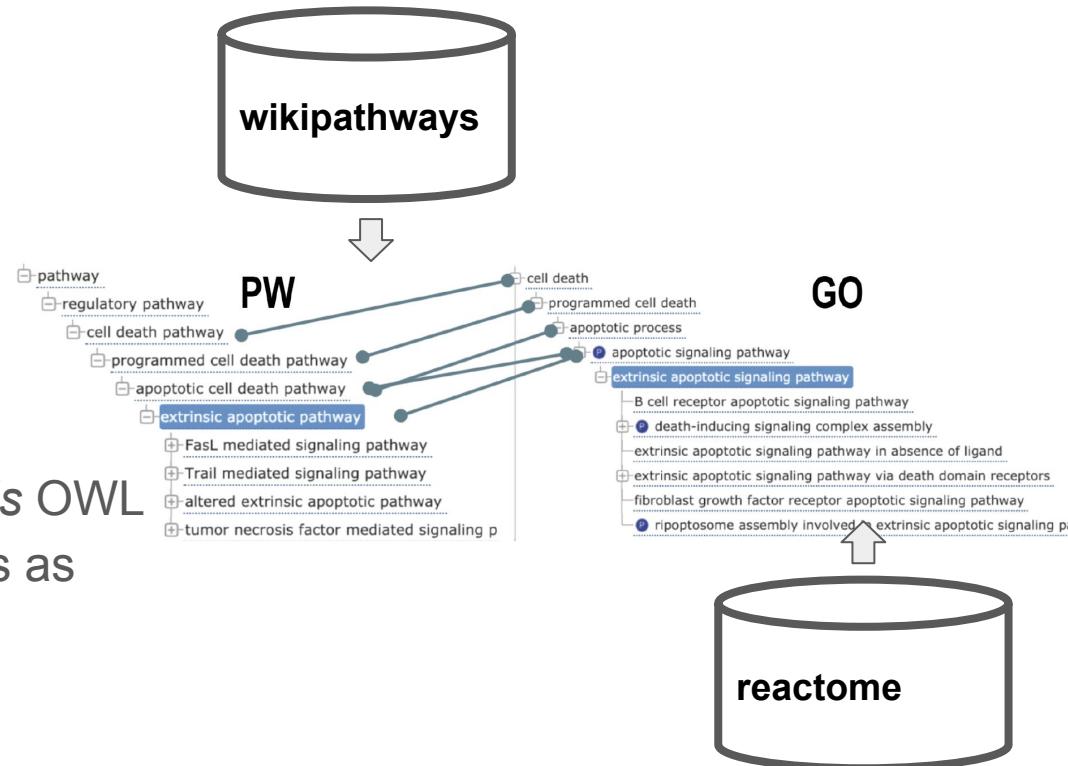
Synergizing pathway databases and ontologies

Pathway Databases

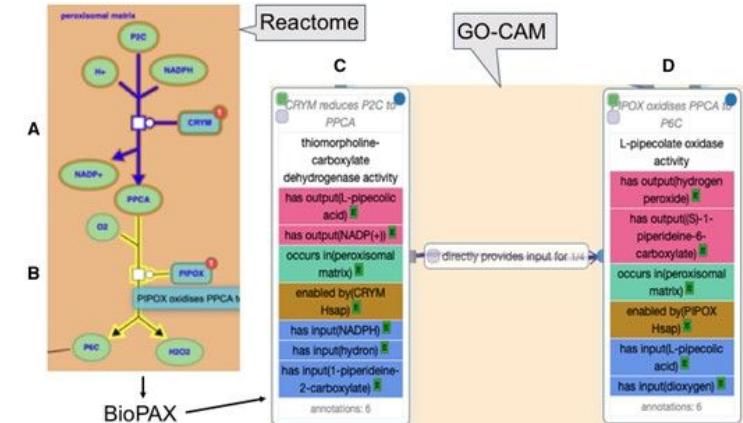
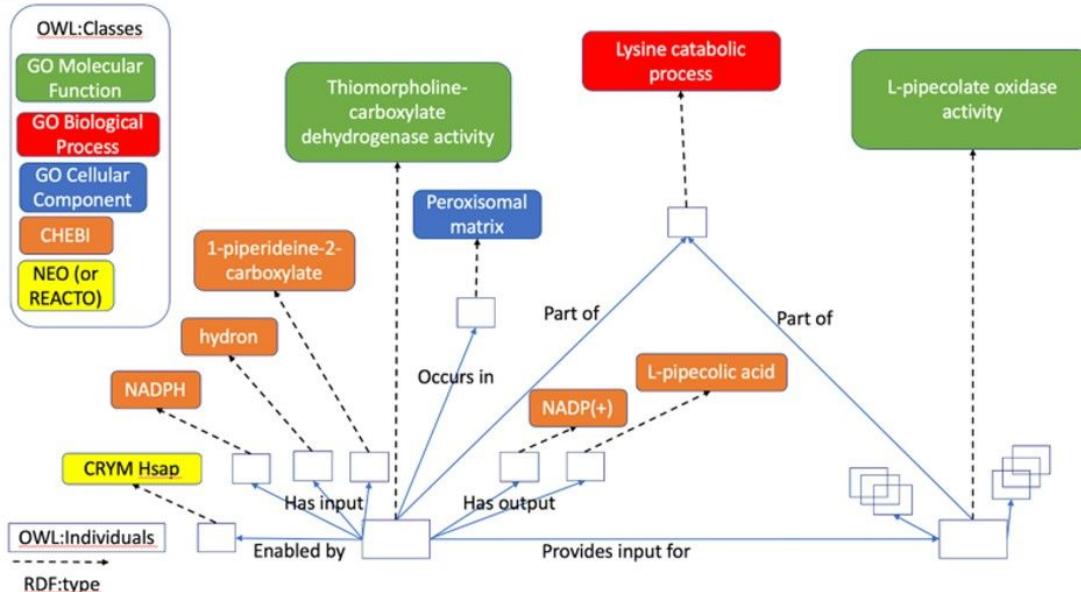
- Multiple databases
- BioPAX OWL insufficient for integration

Approach

- GO CAM *Causal Activity Models* OWL
- Treat pathway database entities as **instances** of ontology classes
- Prototype with Reactome



Synergizing Reactome and GO



Good, B. M., Van Auken, K., Hill, D., P. (2021). Reactome and the Gene Ontology: Digital convergence of data resources. *Bioinformatics* . <https://doi.org/10.1093/bioinformatics/btab325>

Credit:

Peter d'Eustachio, Ben Good, Kimberly Van Auken, David Hill, Huiayu Mi, Jim Balhoff, Seth Carbon, Dustin Ebert, Laurent-Philippe, Judy Blake, Reactome Curators, GO curators, Paul Thomas, GO+Reactome grant

What ontology to use for molecular entities?

- Requirement
 - Annotate to genes, proteins, ncRNAs, complexes from **any species**
 - Use **native IDs** from authoritative databases (e.g ZFIN, SGD, UniProt)
 - Use gene IDs and protein IDs **interchangeably**
 - **OWL** representation
- Approach
 - Create **NEO ontology**: Noctua Entity Ontology
 - Use identifiers.org PURLs
 - Automated translation to OWL from database representation (GPI)
 - Complemented by REACTO
 - Direct translation of Reactome Entities
 - Reuses ids
 - Direct translation
 - No additional classification
 - Simple, no over-ontologization
 - Conflates (g/p) rather than proliferates

Conclusions

- Synergization/harmonization problem not unique to ontologies
- Ontologies can make the situation worse, not better
 - Proliferation of more IDs, more mappings required
- We need rules, guidelines, and design patterns for how to make an ontology rendering of entities in a genomics database
 - We haven't necessarily agreed what these are yet
 - Direct translation, instantiation
 - *It may not be necessary to make an ontology rendering at all!*
 - Need more formal evaluation of existing approaches

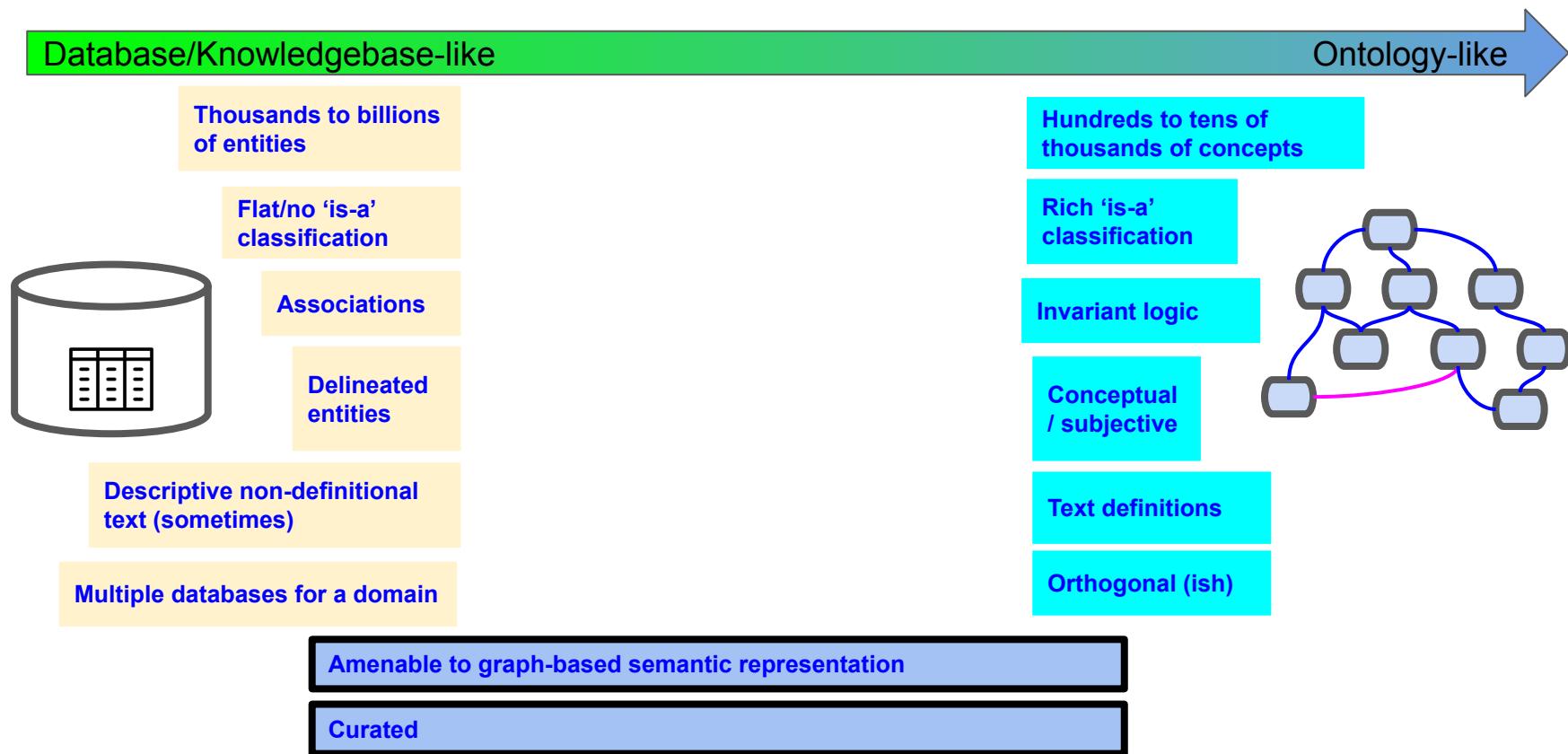
Join the discussion:

- #databases-as-ontologies on OBO slack

end

Old slides follow

There are many things in common...



This is the end....

Reasons for duplicating entities in ontologies

Why?

Required for ontology axioms

Unified semantic representation
of knowledge

Named entities for text mining

“Because we can”

Why not?

Hard to scale

Duplicative effort

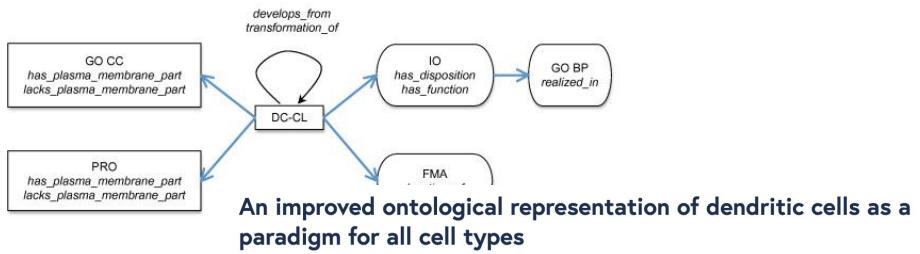
Introduces confusion

Round peg → square hole

Requirement for marker-based cell type definitions

Original Need

- Define cells by surface markers
- Ontology: PRO
- KR: needs to be present as OWL axioms (edges)



Anna Maria Masci, Cecilia N Arighi, Alexander D Diehl, Anne E Lieberman, Chris Mungall,
Richard H Scheuermann, Barry Smith & Lindsay G Cowell

BMC Bioinformatics 10, Article number: 70 (2009) | Cite this article

9252 Accesses | 23 Citations | 3 Altmetric | Metrics



Need within protein ontology to link to genes

OLS / PRotein Ontology (PRO) PR / gene_symbol:report Copy

SHH (human)

 http://www.genenames.org/cgi-bin/gene_symbol_report?hgnc_id=10848 Copy

A protein coding gene SHH in human. [PRO:DNx]

OLS / PRotein Ontology (PRO) PR / MGI:98297 Copy

Shh (mouse)

 <http://www.informatics.jax.org/marker/MGI:98297> Copy

A protein coding gene Shh in mouse. [PRO:DNx]

OLS / PRotein Ontology (PRO) PR / 395615 Copy

SHH (chicken)

 <http://www.ncbi.nlm.nih.gov/gene/395615> Copy

A protein coding gene SHH in chicken. [PRO:DNx]

OLS / PRotein Ontology (PRO) PR / ZDB-GENE-980526-41 Copy

shhb (zebrafish)

 <http://zfin.org/action/marker/view/ZDB-GENE-980526-41> Copy

A protein coding gene shhb in zebrafish. [PRO:DNx]

Other examples

Chemical Entities

- Multiple ontologies need to reference chemical entities in their axioms
- Need grouping classes (e.g. amino acid) as well as specific structures
 - These are not in databases like CHEMBL
 - ChEBI provides this

Dovetailing biology and chemistry: integrating the Gene Ontology with the ChEBI chemical ontology

David P Hill , Nico Adams, Mike Bada, Colin Batchelor, Tanya Z Berardini, Heiko Dietze, Harold J Drabkin, Marcus Ennis, Rebecca E Foulger, Midori A Harris, Janna Hastings, Namrata S Kale, Paula de Matos, Christopher J Mungall, Gareth Owen, Paola Roncaglia, Christoph Steinbeck, Steve Turner & Jane Lomax

Other examples

Chemical Entities

- Multiple ontologies need to reference chemical entities in their axioms
- Need grouping classes (e.g. amino acid) as well as specific structures
 - These are not in databases like CHEMBL
 - ChEBI provides this

Dovetailing biology and chemistry: integrating the Gene Ontology with the ChEBI chemical ontology

David P Hill , Nico Adams, Mike Bada, Colin Batchelor, Tanya Z Berardini, Heiko Dietze, Harold J Drabkin, Marcus Ennis, Rebecca E Foulger, Midori A Harris, Janna Hastings, Namrata S Kale, Paula de Matos, Christopher J Mungall, Gareth Owen, Paola Roncaglia, Christoph Steinbeck, Steve Turner & Jane Lomax

BMC Genomics 14, Article number: 513 (2013) | [Cite this article](#)

Organism Taxa

- Need for taxon constraints
 - Evolutionary knowledge
 - Automated QC
 - Extracting subsets
- We created a direct transform of the NCBI taxonomy database

Formalization of taxon-based constraints to detect inconsistencies in annotation and ontology development

Jennifer I Deegan (née Clark) , Emily C Dimmer & Christopher J Mungall 

BMC Bioinformatics 11, Article number: 530 (2010) | [Cite this article](#)

4840 Accesses | 29 Citations | 0 Altmetric | [Metrics](#)

NCBITaxon: Isomorphic (one-to-one) transform

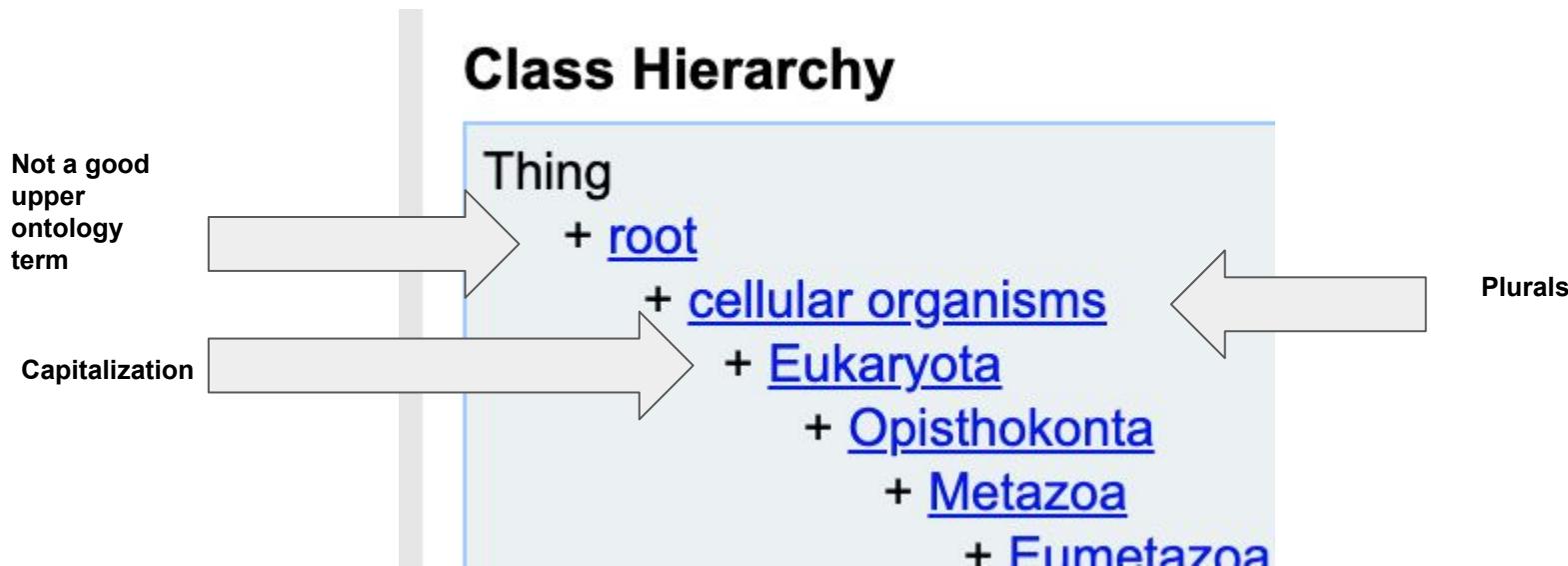
Class: Homo sapiens	
Term IRI: http://purl.obolibrary.org/obo/NCBITaxon_9606	
Annotations	
<ul style="list-style-type: none">• database_cross_reference: GC_ID:1• has_exact_synonym: man; human• has_obo_namespace: ncbi_taxonomy• has_rank: species	
Class Hierarchy	
Thing	
+ root	+ cellular organisms
	+ Eukaryota
	+ Opisthokonta
	+ Metazoa
	+ Eumetazoa
	+ Bilateria
	+ Deuterostomia
	+ Chordata
	+ Craniata <chordates>
	+ Vertebrata <vertebrates>
	+ Gnathostoma <vertebrates>
	+ Teleostomi
	+ Euteleostomi
	+ Sarcopterygii
	- Dipnotetrapodomorpha
	+ Tetrapoda
	+ Amniota
	+ Mammalia
	+ Theria <mammals>
	+ Eutheria
	+ Boreoeutheria
	+ Euarchontoglires
	+ Primates
	+ Haplorrhini
	+ Simiiformes
	+ Catarrhini
	+ Hominoidea
	+ Hominidae
	+ Homininae
	+ Homo
	- Homo heidelbergensis
	+ environmental samples <primates>
	- Homo sapiens
	- Homo sapiens neanderthalensis
	- Homo sapiens subsp. 'Denisova'

Automated Translation

- OBO technical group

Changes in content come from NCBI

Some parts are not very “OBO-esque”



Some parts are not very “OBO-esque”

- [Primates](#)
 - + [Haplorrhini](#)
 - + [Simiiformes](#)
 - + [Catarrhini](#)
 - + [Hominoidea](#)
 - + [Hominidae](#)
 - + [Homininae](#)
 - + [Homo](#)
 - [Homo heidelbergensis](#)
 - + [Homo sapiens](#)
 - [environmental samples <primates>](#)
 - [Homo sapiens environmental sample](#)

Class: ecological metagenomes

Term IRI: http://purl.obolibrary.org/obo/NCBITaxon_410657

Annotations

- **database_cross_reference:** GC_ID:11
- **has_obo_namespace:** ncbi_taxonomy
- **has_related_synonym:** geographical metagenomes

Class Hierarchy

Thing

- + [root](#)
 - + [unclassified entries](#)
 - + [unclassified sequences](#)
 - + [metagenomes](#)
 - [synthetic metagenome](#)
 - [metagenome](#)
 - + [organismal metagenomes](#)
 - [ecological metagenomes](#)
 - [bioreactor metagenome](#)
 - [permafrost metagenome](#)
 - [food fermentation metagenome](#)
 - [aquatic metagenome](#)
 - [activated carbon metagenome](#)
 - [dust metagenome](#)
 - [tomb wall metagenome](#)
 - [power plant metagenome](#)
 - [indoor metagenome](#)
 - [coal metagenome](#)
 - [more...](#)

Other drivers: semantic representation of knowledge

RDF/OWL provides a possible framework for unifying databases

- Graphs
- URIs
- Semantics

OWL doesn't necessitate "ontology"

- But many entities are naturally represented as OWL classes

Semantic e-Science in Biomedicine

Methodology | [Open Access](#) | Published: 09 May 2007

Advancing translational research with the Semantic Web

[Alan Ruttenberg](#), [Tim Clark](#), [William Bug](#), [Matthias Samwald](#), [Olivier Bodenreider](#), [Helen Chen](#), [Donald Doherty](#), [Kerstin Forsberg](#), [Yong Gao](#), [Vipul Kashyap](#), [June Kinoshita](#), [Joanne Luciano](#), [M Scott Marshall](#), [Chimezie Ogbuji](#), [Jonathan Rees](#), [Susie Stephens](#), [Gwendolyn T Wong](#), [Elizabeth Wu](#), [Davide Zaccagnini](#), [Tonya Hongsermeier](#), [Eric Neumann](#), [Ivan Herman](#) & [Kei-Hoi Cheung](#)✉

BMC Bioinformatics 8, Article number: S2 (2007) | [Cite this article](#)

120k Accesses | 155 Citations | 22 Altmetric | [Metrics](#)

The **global scope of identifiers** that follow from the use of URIs offer a path out of the complexities caused by the proliferation of local identifiers for entities of biomedical interest. Too much effort has been spent developing services mapping between, for instance, the gene identifiers used by the many data sources recording information about them.

The Semantic Web schema languages, RDFS and OWL, offer the potential to simplify the management and comprehension of a complicated and rapidly evolving set of relationships that we need to record among the data describing the products of the life and medical sciences. Along with the benefits of the technologies that underlie our current data stores, there are a number of significant disadvantages that the Web schema languages remediate.

Recap so far

There are different use cases for ‘re-representing’ entities from existing databases in ontologies

- Ontology axioms
- Integrated semantic representation of knowledge

Different approaches to doing this

- Automated translation
- Manual ontologization

The current situation is confusing with a lot of duplication

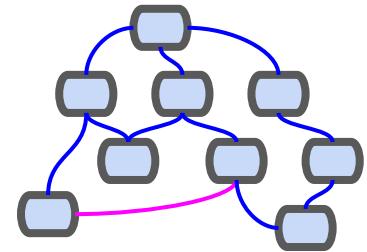
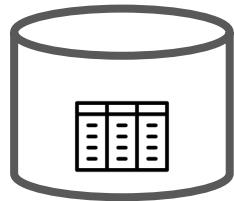
Next:

- Framework for understanding distinctions better

Spectrum of features rather than binary distinction

Database/Knowledgebase-like

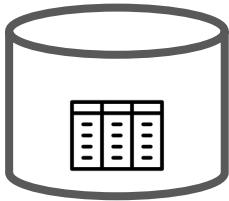
Ontology-like



Feature: scale

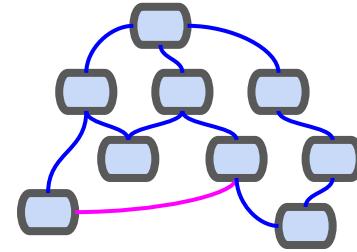
Database/Knowledgebase-like

Thousands to billions
of entities



Ontology-like

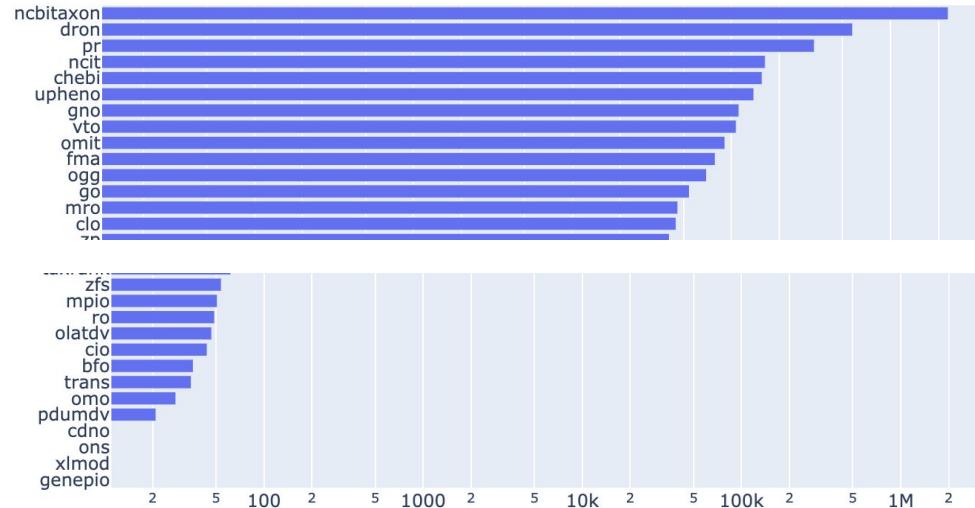
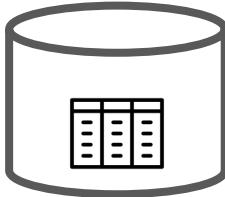
Hundreds to tens of
thousands of concepts



Feature: scale

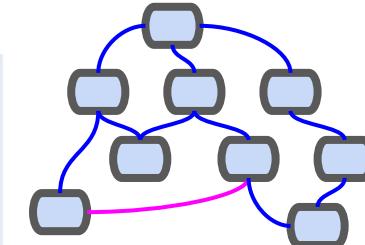
Database/Knowledgebase-like

Thousands to billions
of entities



Ontology-like

Hundreds to tens of
thousands of concepts



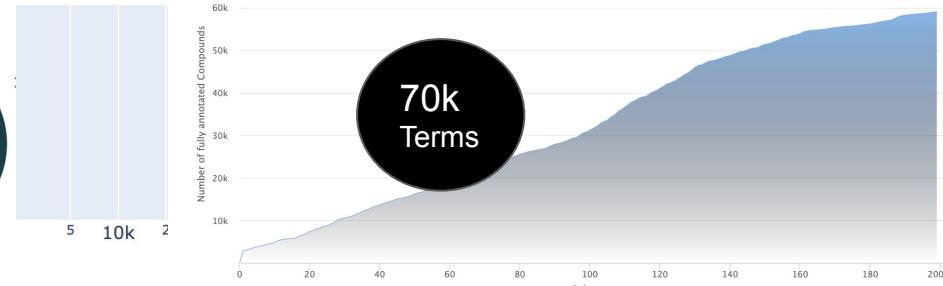
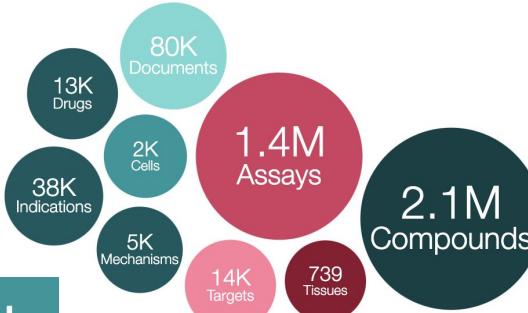
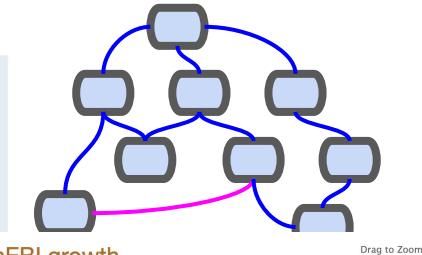
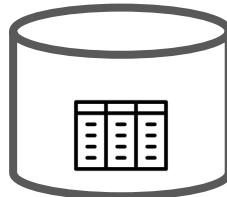
Feature: scale

Database/Knowledgebase-like

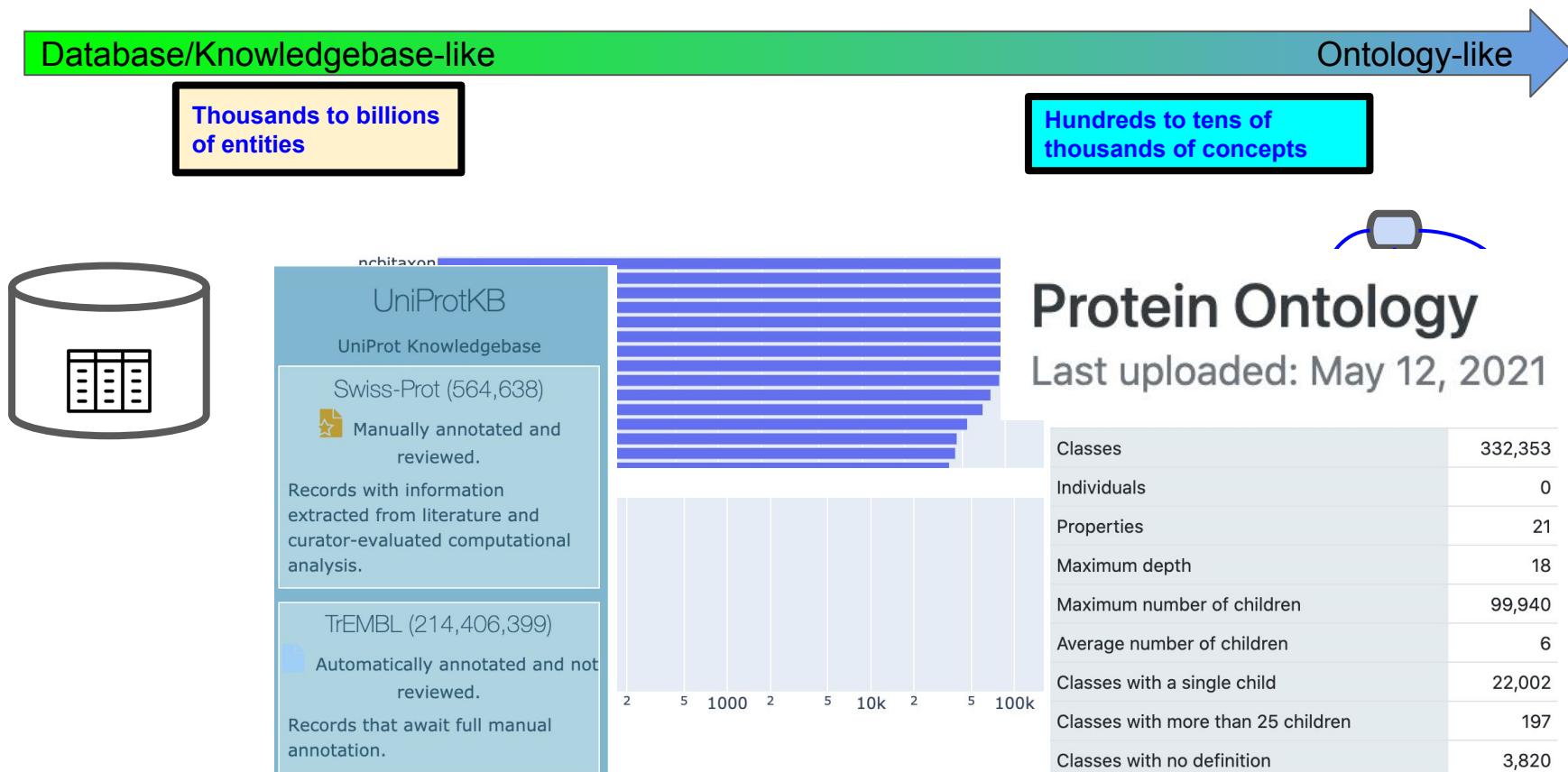
Ontology-like

Thousands to billions
of entities

Hundreds to tens of
thousands of concepts



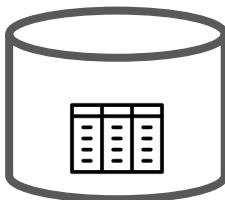
Feature: scale



Feature: scale

Database/Knowledgebase-like

Thousands to billions
of entities



Ontology-like

Hundreds to tens of
thousands of concepts

> Nucleic Acids Res. 2021 Jan 8;49(D1):D751-D763. doi: 10.1093/nar/gkaa939.

The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities

I-Min A Chen ¹, Ken Chu ¹, Krishnaveni Palaniappan ¹, Anna Ratner ¹, Jinghua Huang ¹,
Marcel Huntemann ¹, Patrick Hajek ¹, Stephan Ritter ¹, Neha Varghese ¹, Rekha Seshadri ¹,
Simon Roux ¹, Tanja Woyke ¹, Emiley A Eloe-Fadrosh ¹, Natalia N Ivanova ¹, Nikos C Kyripides ¹

Affiliations – collapse

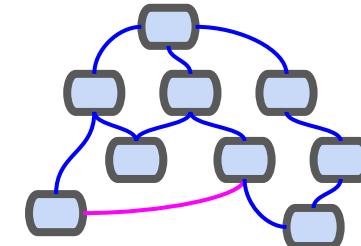
Affiliation

¹ Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, 1
Cyclotron Road, Berkeley, CA 94720, USA.

PMID: 33119741 PMCID: PMC7778900 DOI: [10.1093/nar/gkaa939](https://doi.org/10.1093/nar/gkaa939)



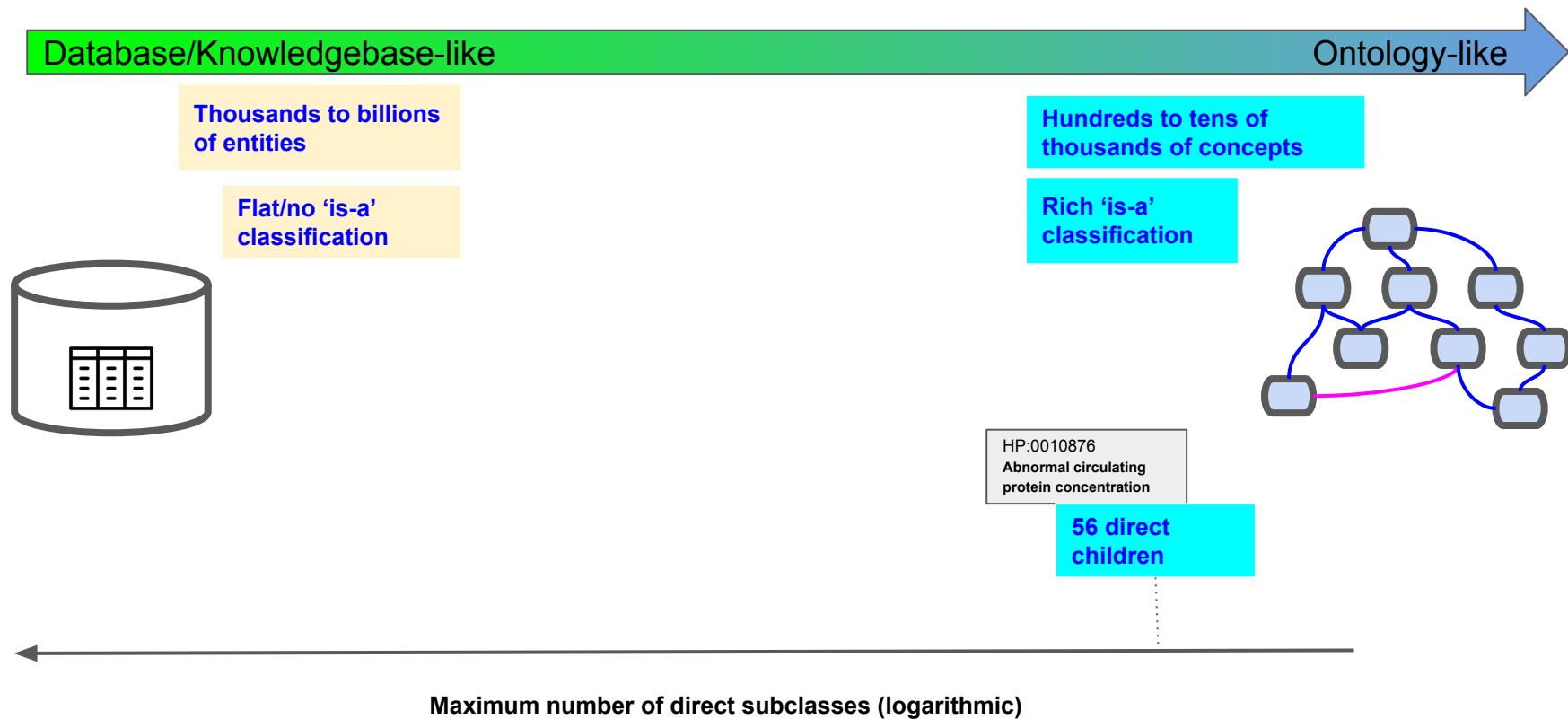
Currently 65 billion genes



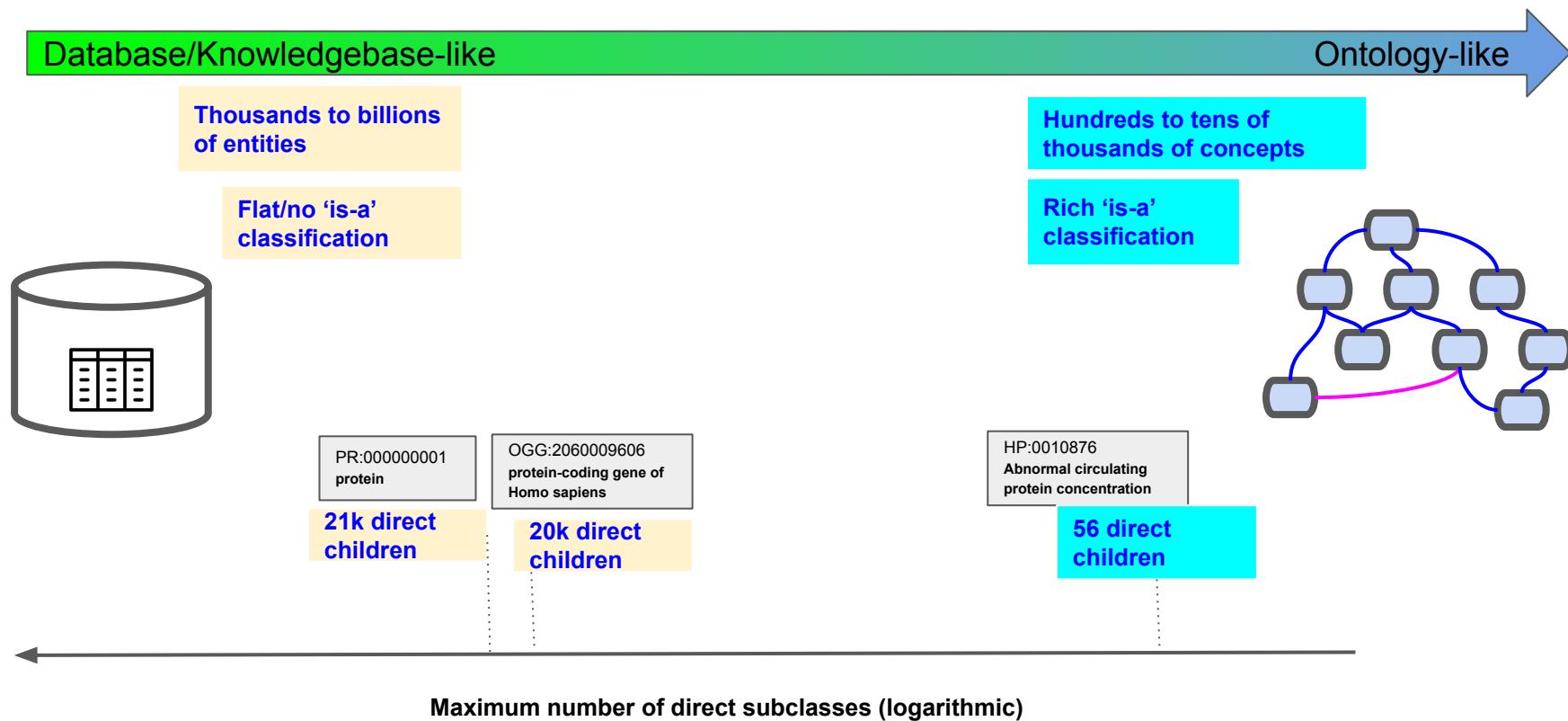
Feature: hierarchy



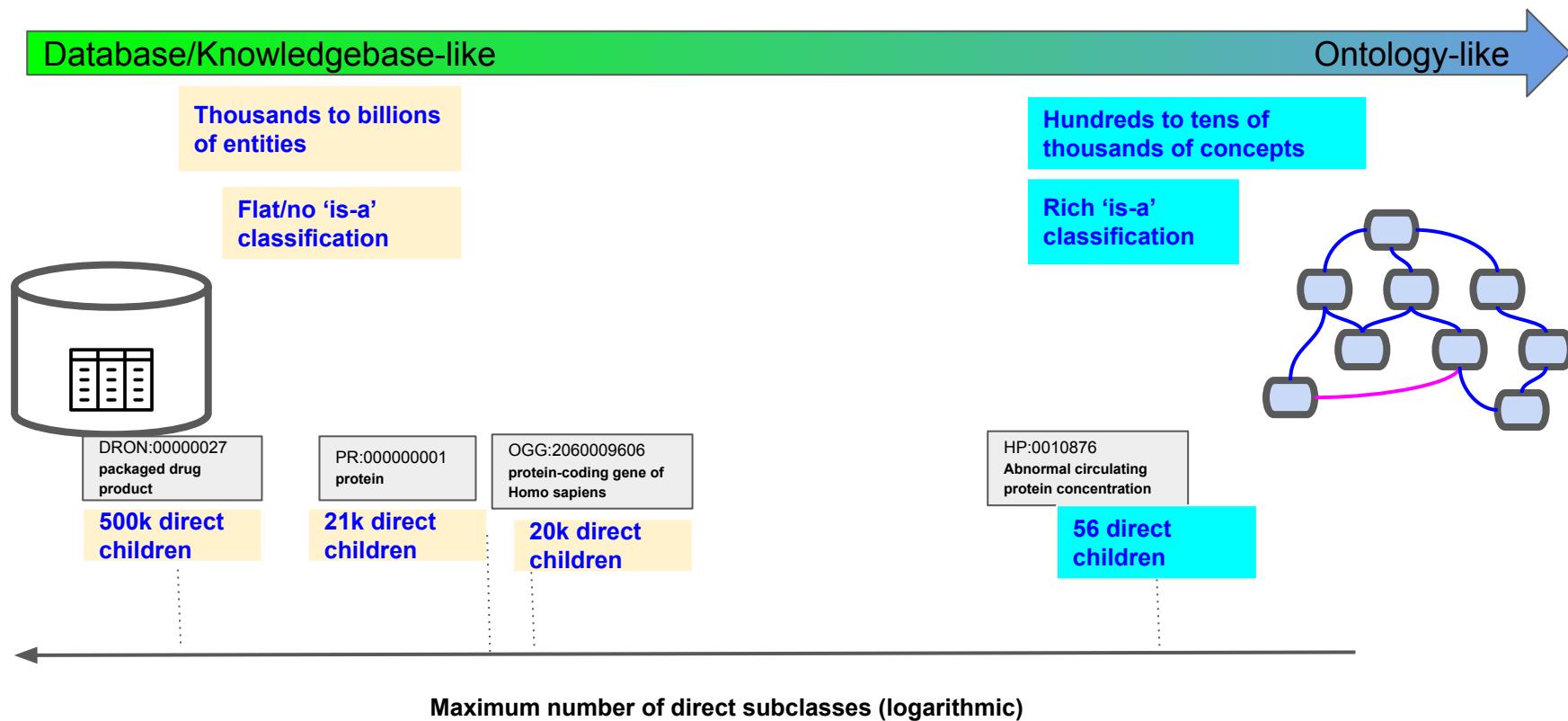
Feature: hierarchy



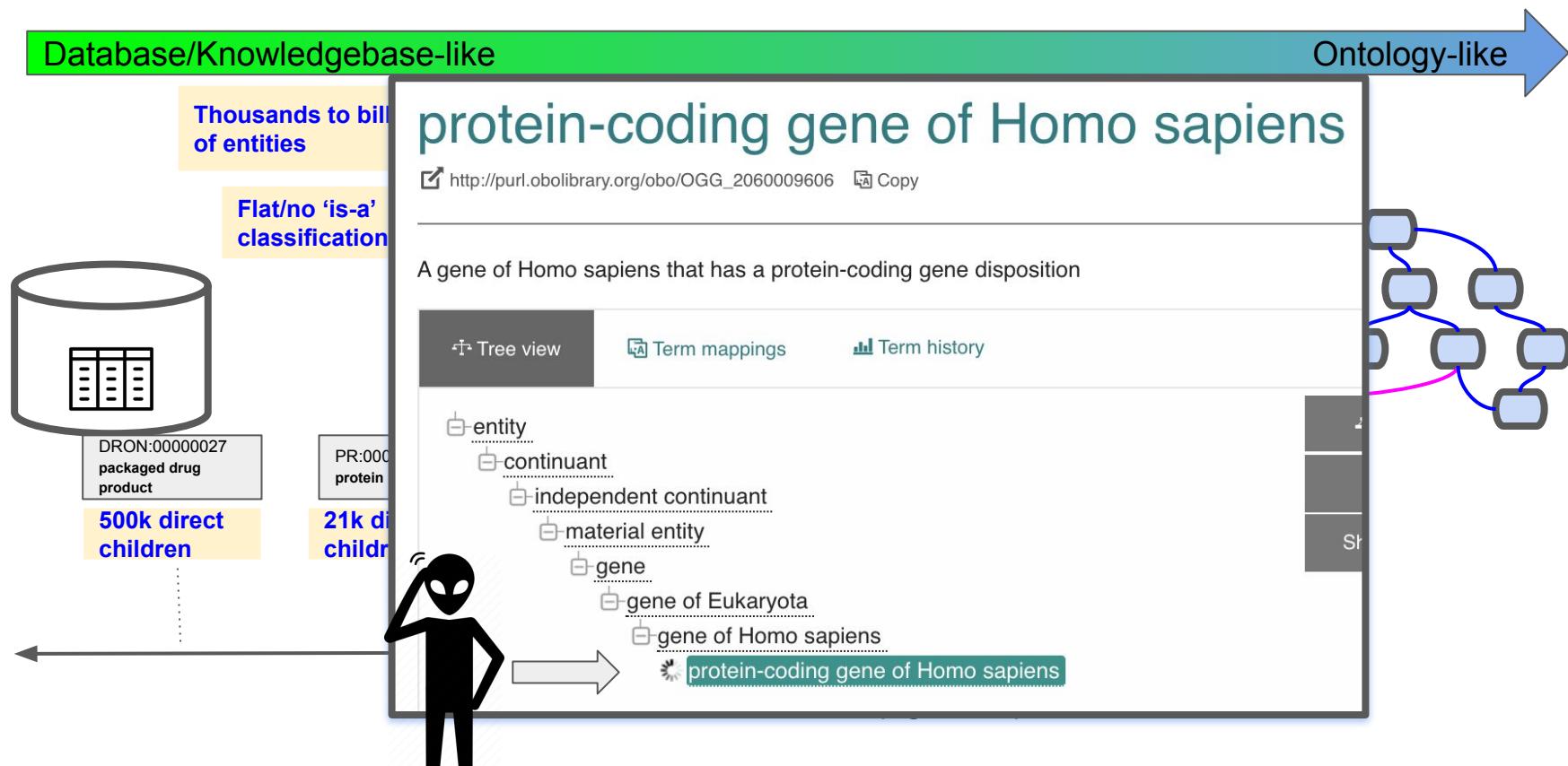
Feature: hierarchy



Feature: hierarchy



Spinning wheel of death



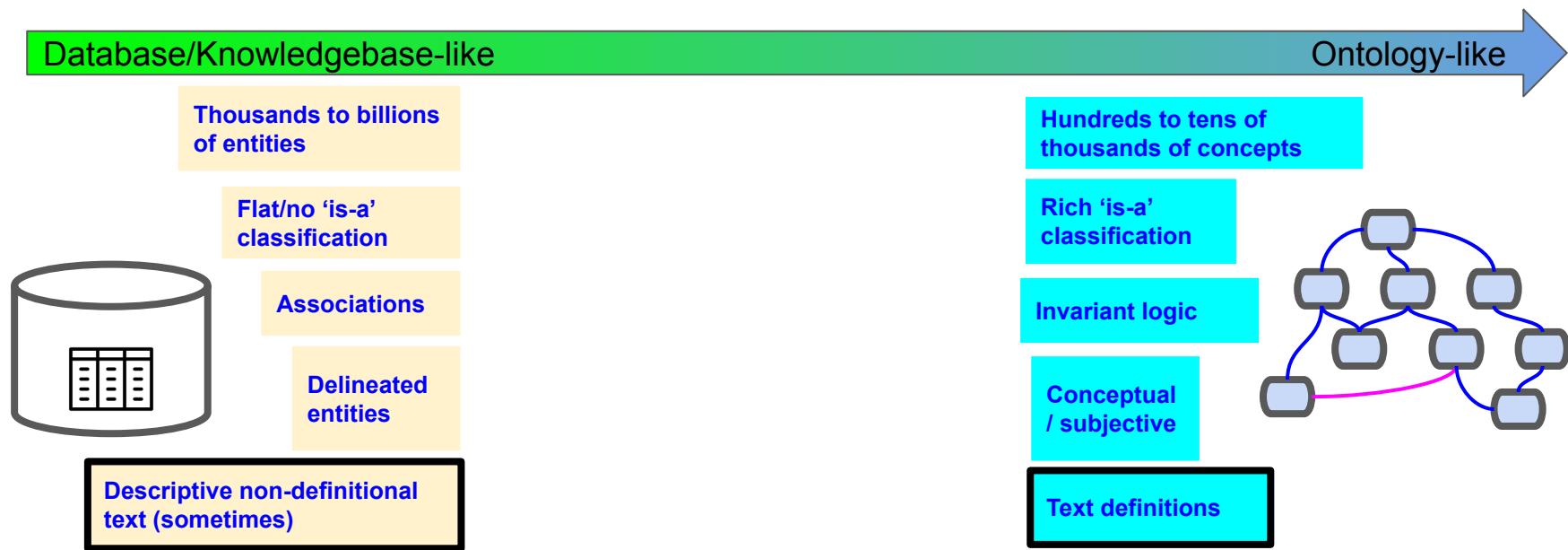
Feature: knowledge representation



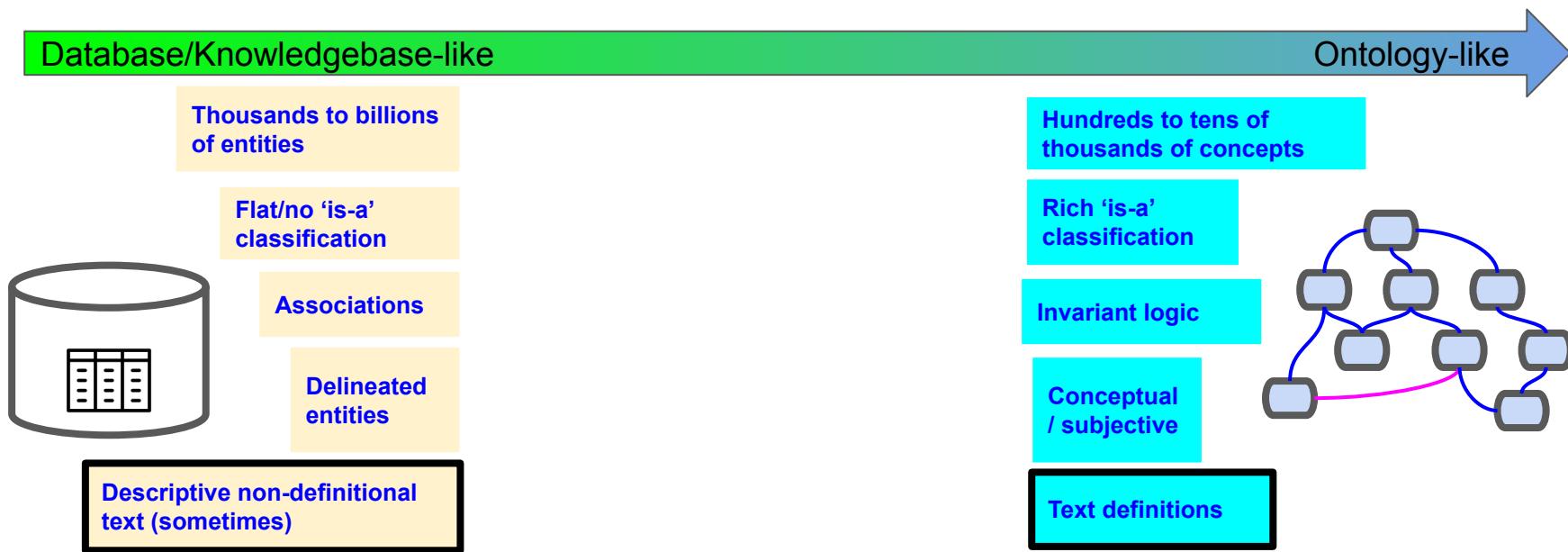
Feature: discrete entities



Feature: textual definitions and descriptive names



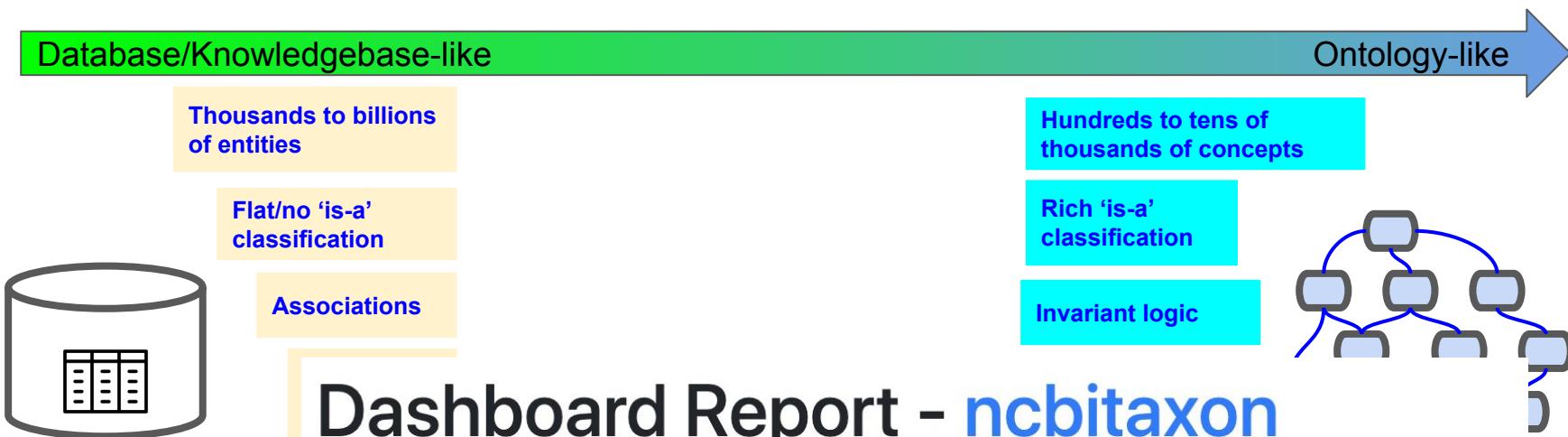
Feature: textual definitions and descriptive names



This gene encodes a protein that is instrumental in patterning the early embryo. It has been implicated as the key inductive signal in patterning of the ventral neural tube, the anterior-posterior limb axis, and the ventral somites. Of three human proteins showing sequence and functional similarity to the sonic hedgehog protein of Drosophila, this protein is the most similar. The protein is made as a precursor that is autocatalytically cleaved; the N-terminal portion is soluble and contains the signalling ... [MORE]

"A sonic hedgehog protein that is encoded in the genome of human"

Feature: textual definitions and descriptive names



Dashboard Report - ncbitaxon

Descriptive non-d
text (sometimes) Version: <http://purl.obolibrary.org/obo/ncbitaxon/2021-02-15/ncbitaxon.owl>

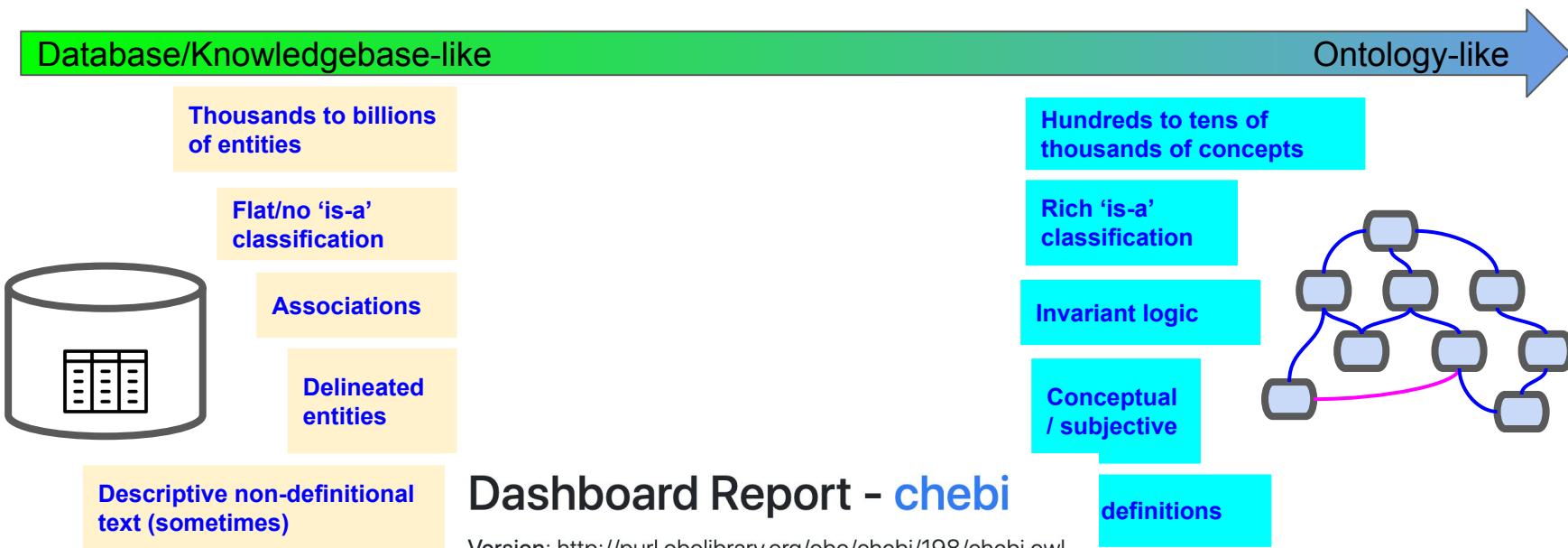
Date run:

Textual Definitions



2308118 missing definitions. See ROBOT Report for details.

Feature: textual definitions and descriptive names

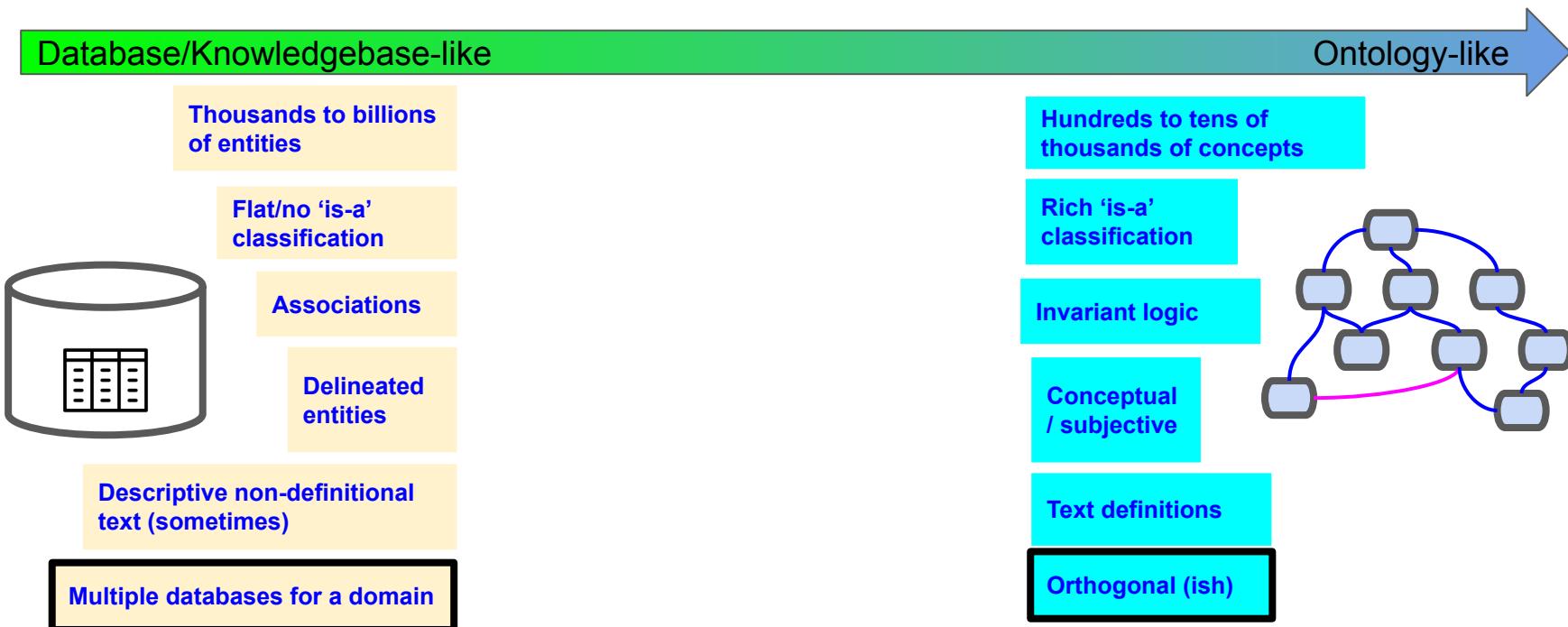


Textual Definitions

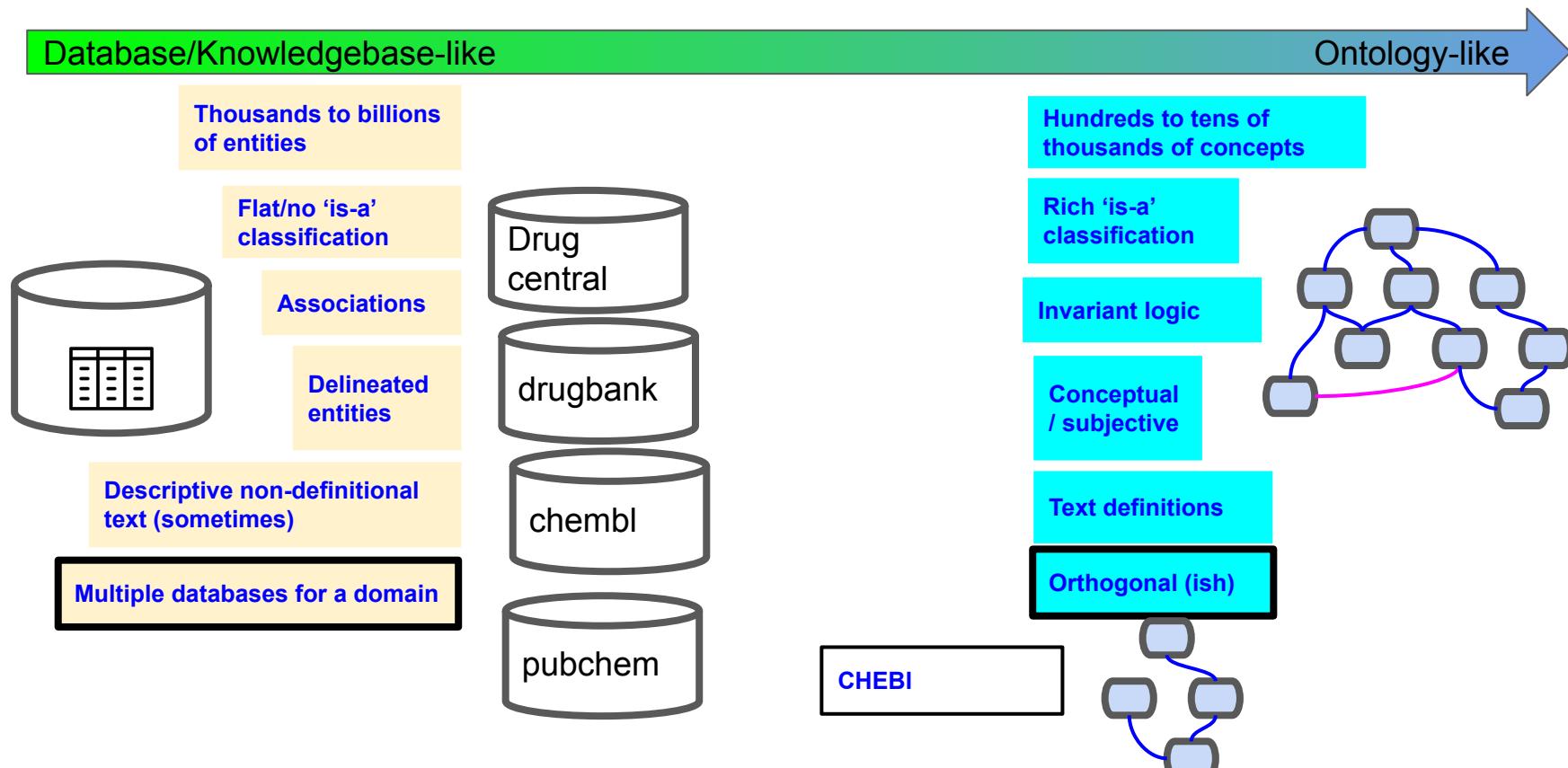


1050 duplicate definitions. 87284 missing definitions.
See ROBOT Report for details.

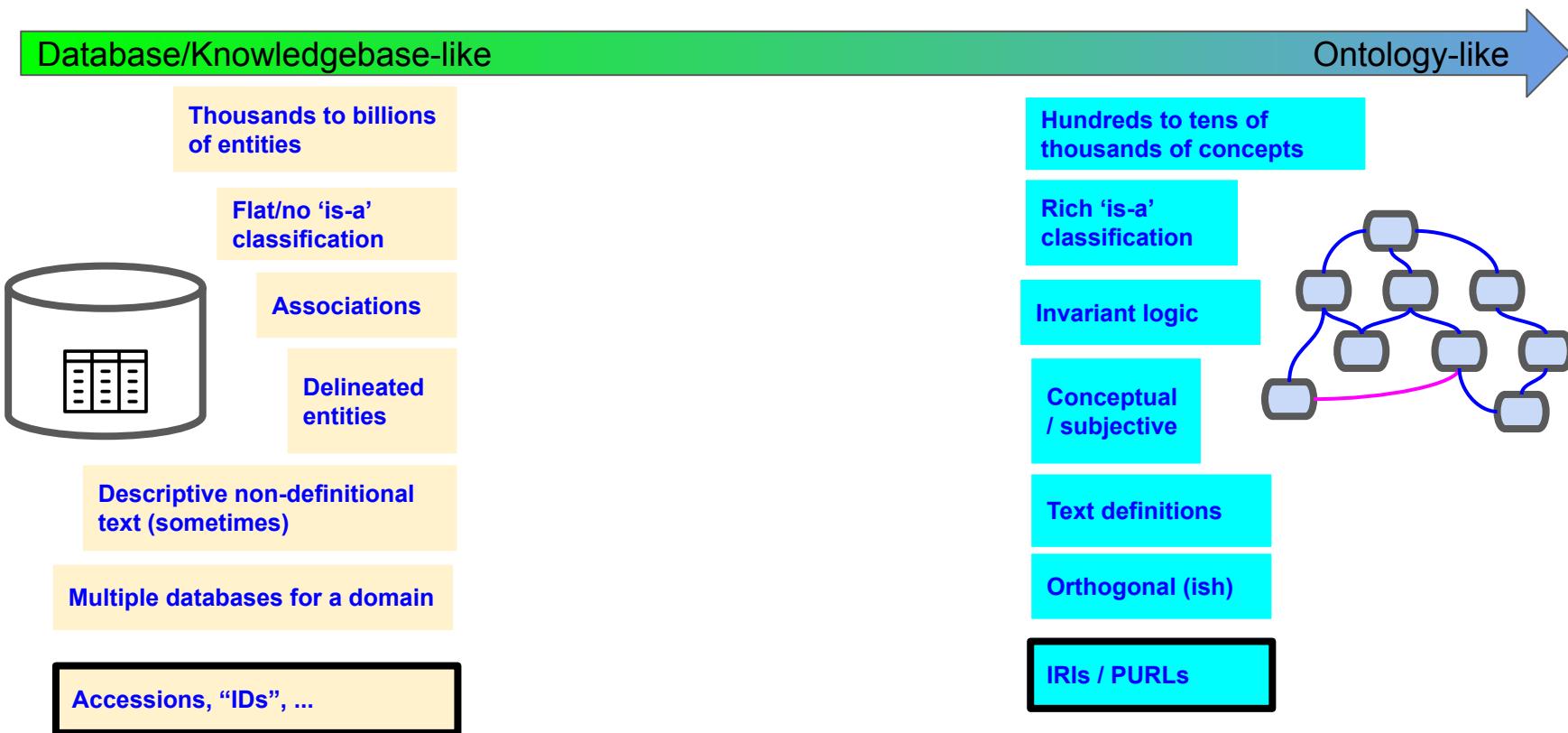
Feature: orthogonality



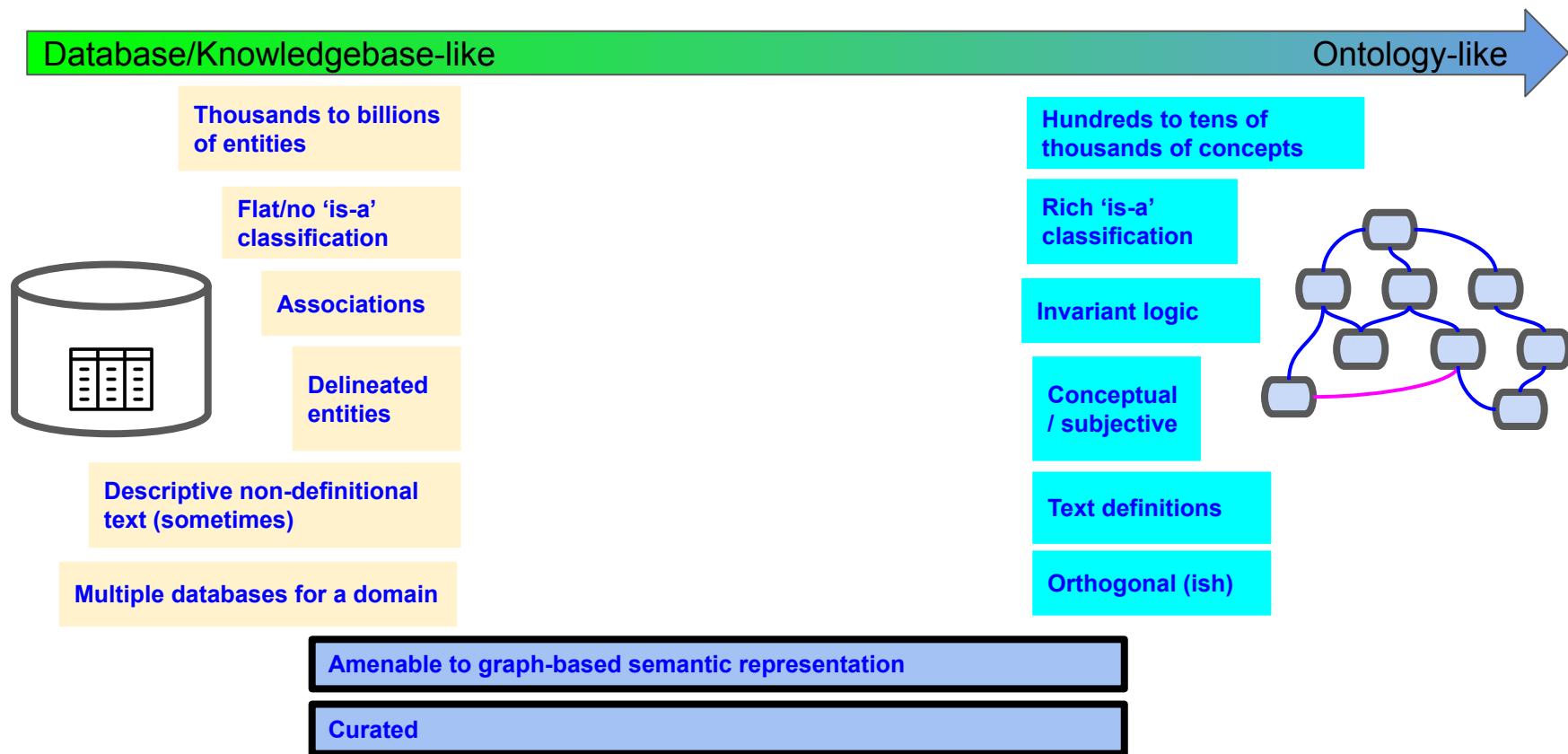
Feature: orthogonality



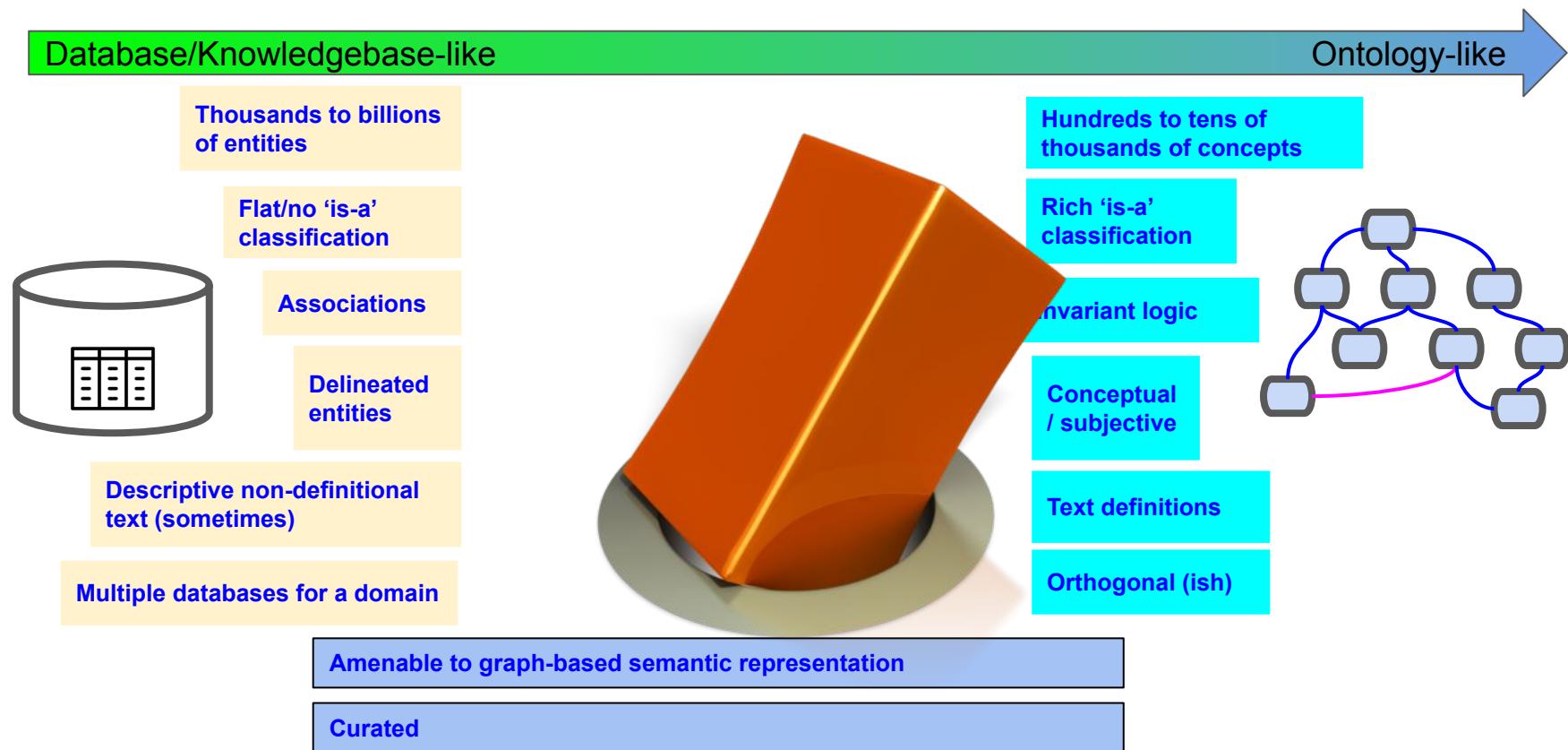
Feature: identifiers



There are many things in common...



...But we still have a mismatch



The Knowledge Representation quandary

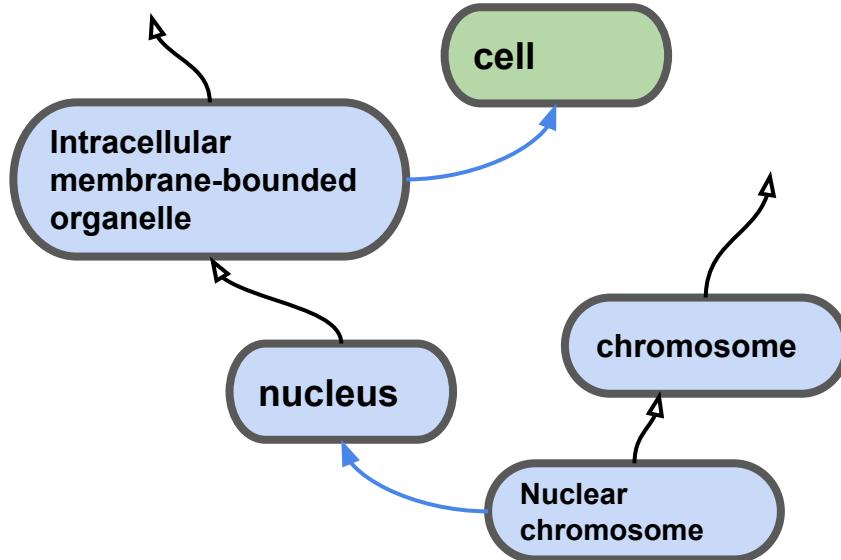
Ontologies arose out of the field of Knowledge Representation (KR)

OWL (OWL-DL) is the main representation language

And yet

- Class-based OWL ontologies have limits on Knowledge Representation

Ontologies work well for ‘invariant’ knowledge



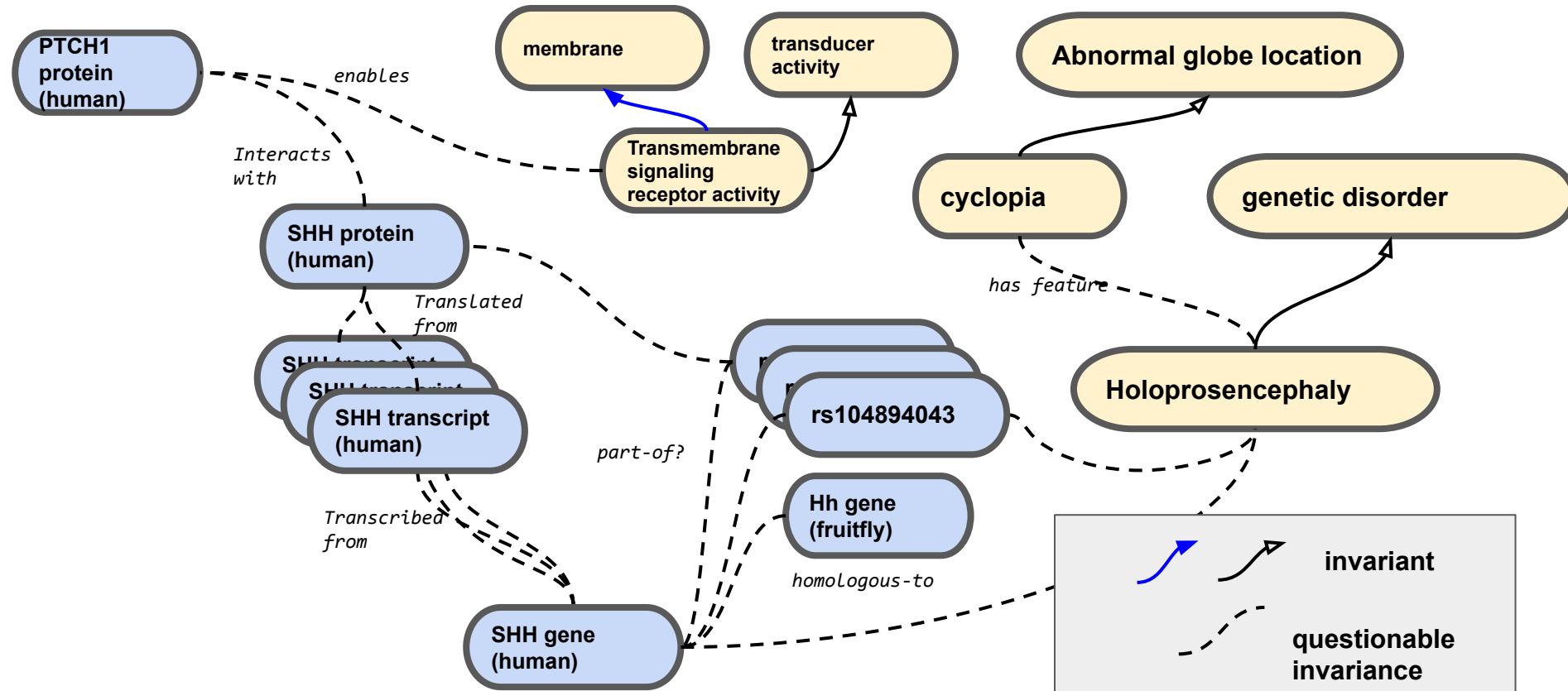
OWL axioms (*invariant*):

nucleus subClassOf IMBO

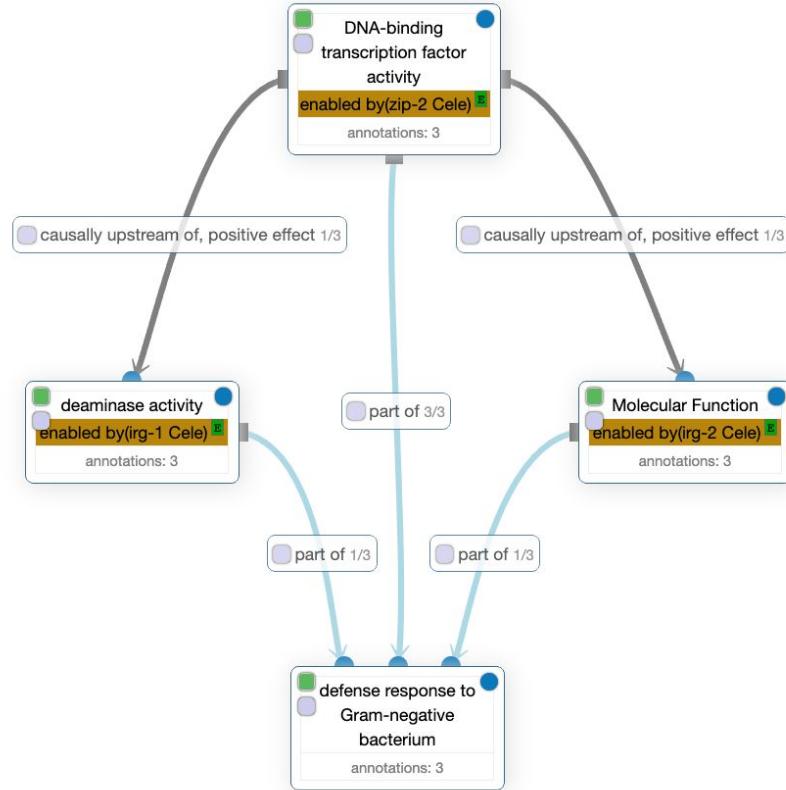
IMBO subClassOf part-of some cell

nuclear chromosome EquivalentTo chromosome and part-of some nucleus

non-invariant knowledge presents challenges



GO-CAMs leverage OWL but are not ontologies



Purpose

- Knowledgebase of gene function
- Richer than standard annotations

Use of OWL

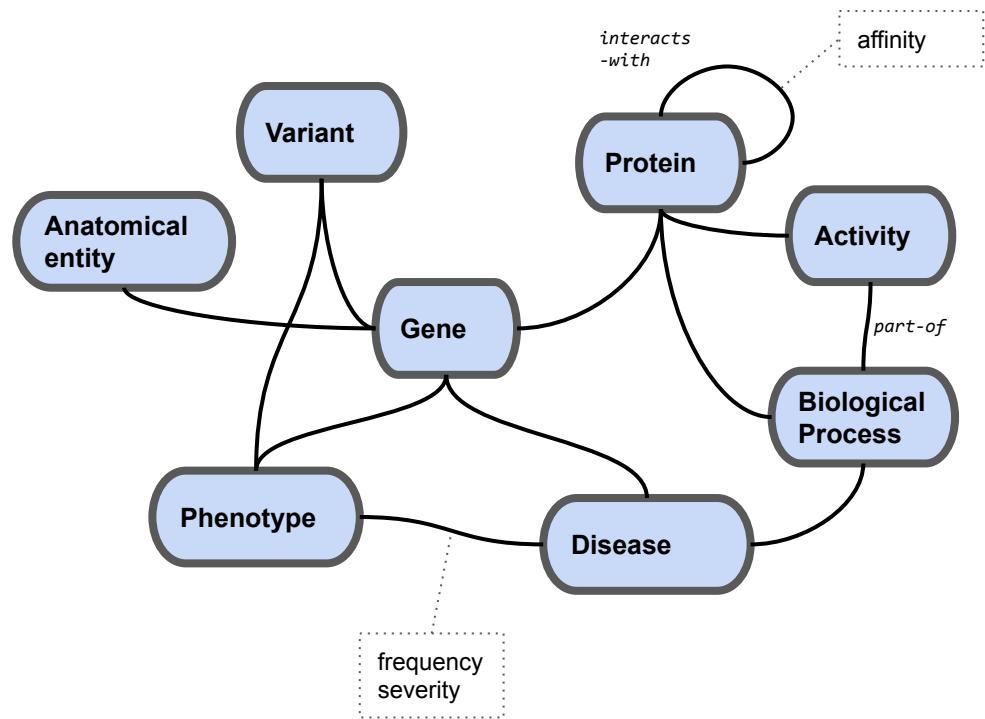
- Leverages GO and multiple OBOs
 - NEO
 - REACTO
- Heavy use of reasoning
- Instance-based representation

<http://geneontology.org/cam>

Knowledge Graphs

Shares many aspects of ontologies

- Edges can be more flexible than in OWL ontologies
- Suited for looser associations
- Probabilistic associations (e.g. disease to phenotype)



There are many bio KGs

Wikidata

NCATS Biomedical Data Translator

Monarch Knowledge Graph

Hetionet

SPOKE

Bio2rdf

Google “data commons”

Uniprot triplestore

EBI triplestore (RIP)

Most of these
incorporate OBO
ontologies

Most include automated
translations from
biological databases
⇒ more complete than
entity ontologies in OBO

Emerging ‘KG Hub’
efforts to coordinate
activities

What is the path forward for OBO treatment of DBs?

Not in

OBO is scoped to
“classic” ontologies

- We encourage a parallel ‘OBO analog’ for KGs and other integration efforts

Partially in

“Conceptual coat rack”

- Lightweight axiomatization
- Other knowledge in KGs and other structures

All in

OBO explicitly expands scope

- Many questions about how this works

May be a
hybrid of all
these

Must be
intentional!

Questions: Which entities

Which entities are a priority for

- OWL/semantic representation
- Inclusion in OBO

E.g. transcripts, variants, other genomic features, ...?

Questions: Construction

How should these be constructed?

- Full coverage
 - Direct transformation of a single resource?
 - Direct transformation of a multiple resources with entity normalization?
- Partial coverage
 - Curation de novo?
- Hybrid?

Questions: IDs and URIs

What should the IDs look like?

- Mint new ones de-novo
- Reuse IDs from a single chosen resource/database for each domain
- Reuse IDs from a mix of chosen resource/database for each domain

What about URIs?

- OBO PURLs
- Identifiers.org
- Existing semantic web PURLs if appropriate (e.g. purl.uniprot.org)
- Web portal URLs, e.g.
http://www.genenames.org/cgi-bin/gene_symbol_report?hgnc_id=10848

Questions: Ownership and Governance

Who should be in charge of these ontologies?

- The first person to write a Python script to convert a database dump to OWL
- OBO technical group
- The database resource itself
- Hybrid

Proposed Next Steps

- Further discussion will be required
 - Engage resource providers and wider non-ontologist user base
 - Start with loose coordination on OBO slack for now?
- White paper?
 - Set of recommendations
- Funding will be a challenge...
- Discuss!

Questions to discuss

Survey results:

- <https://docs.google.com/forms/d/1uPUNj26wUACWw-X4v9eCtapdKPa5ZXeTCMhdI8zMNE/edit#responses>

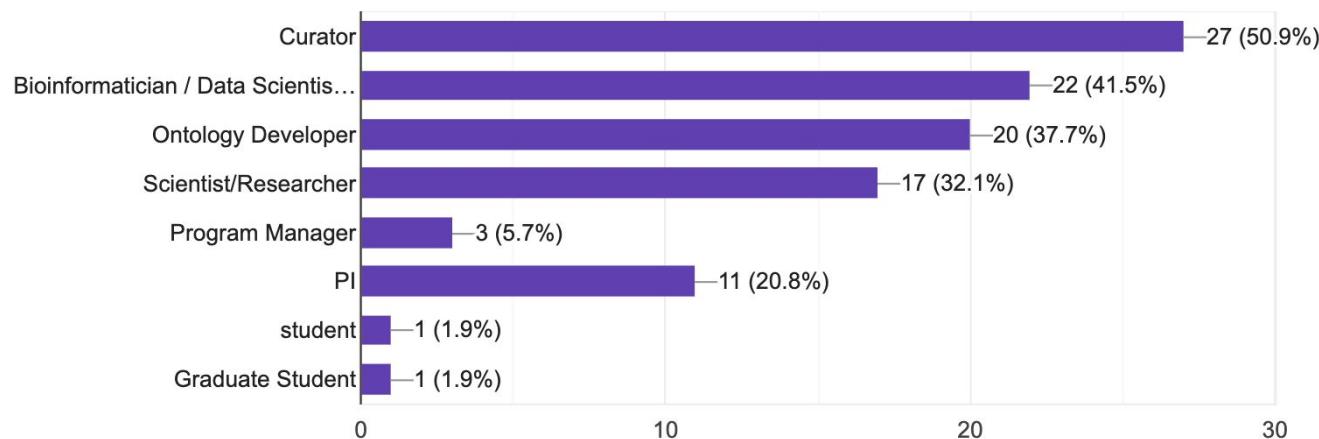
Responses 54

Mixture of respondents

Tell us about yourself

What best describes your role(s)

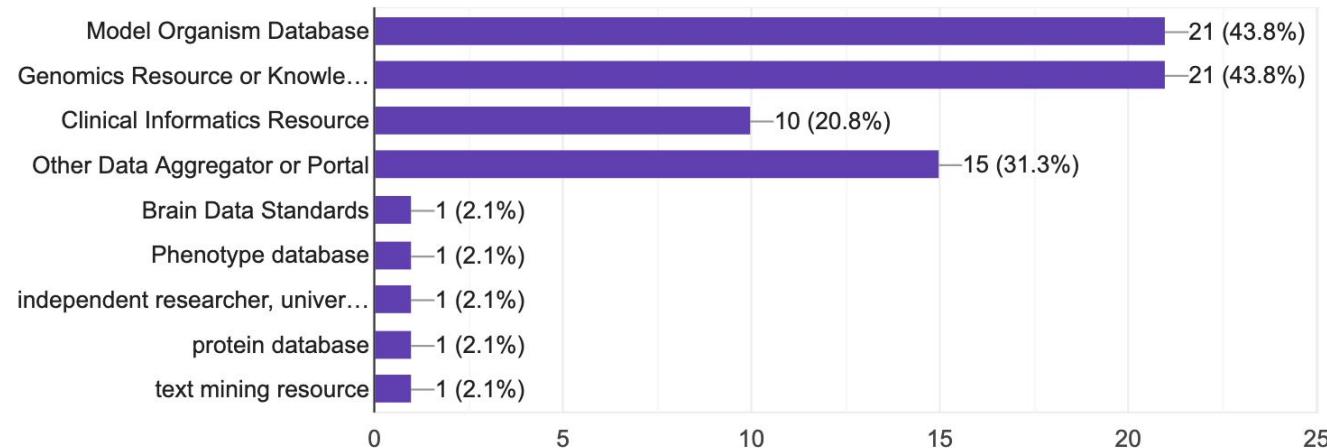
53 responses



Good representation from resources

Do you work for any of the following

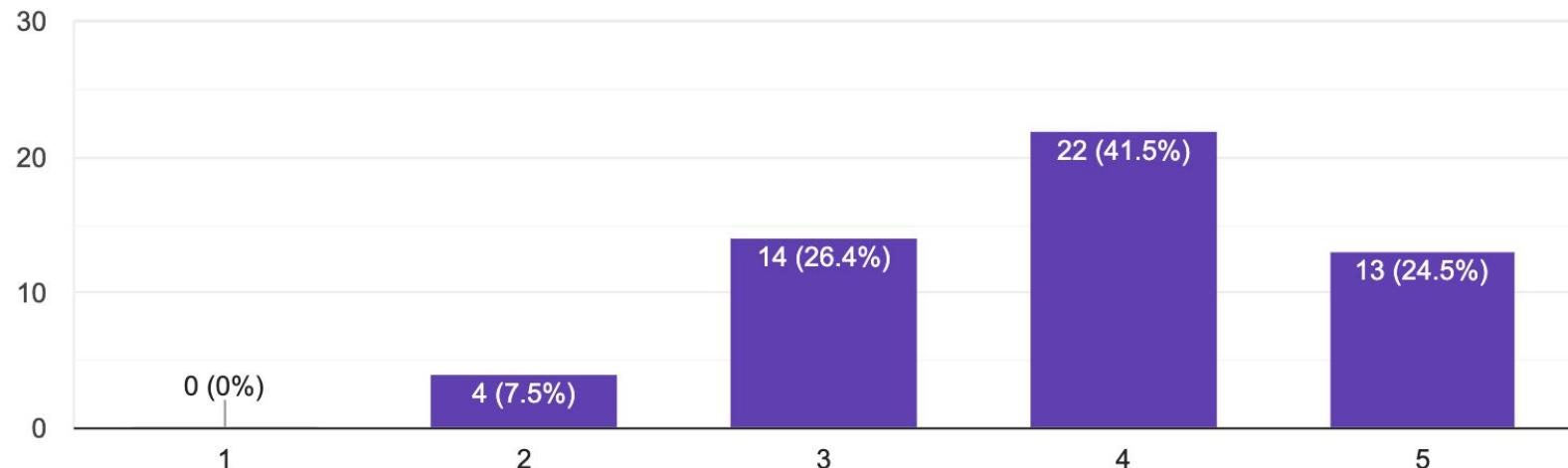
48 responses



Most people familiar with ontologies (unsurprising given we targeted obo community)

How comfortable are you with ontologies?

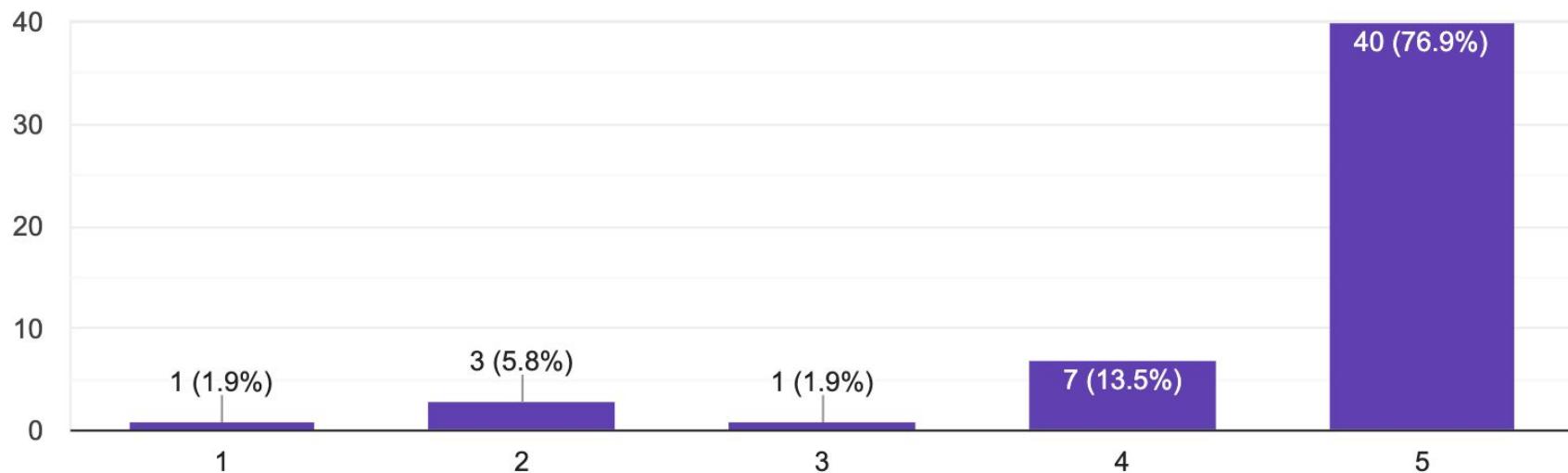
53 responses



ditto

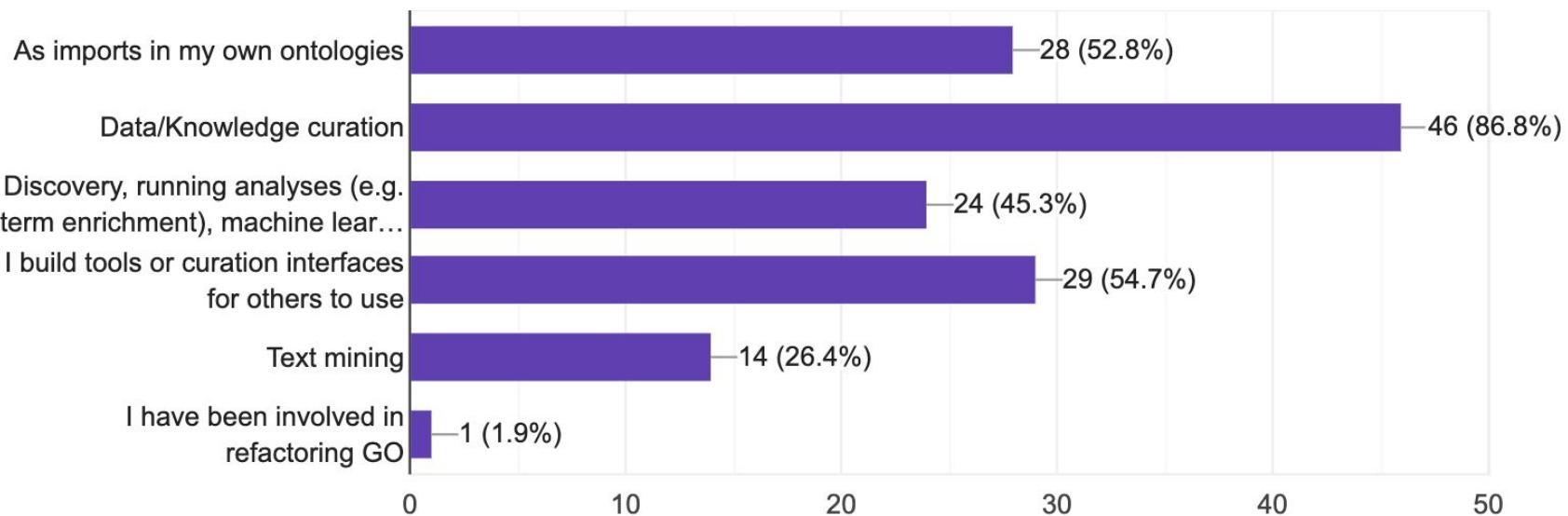
Do you use any of the ontologies in OBO?

52 responses



How do you use ontologies?

53 responses



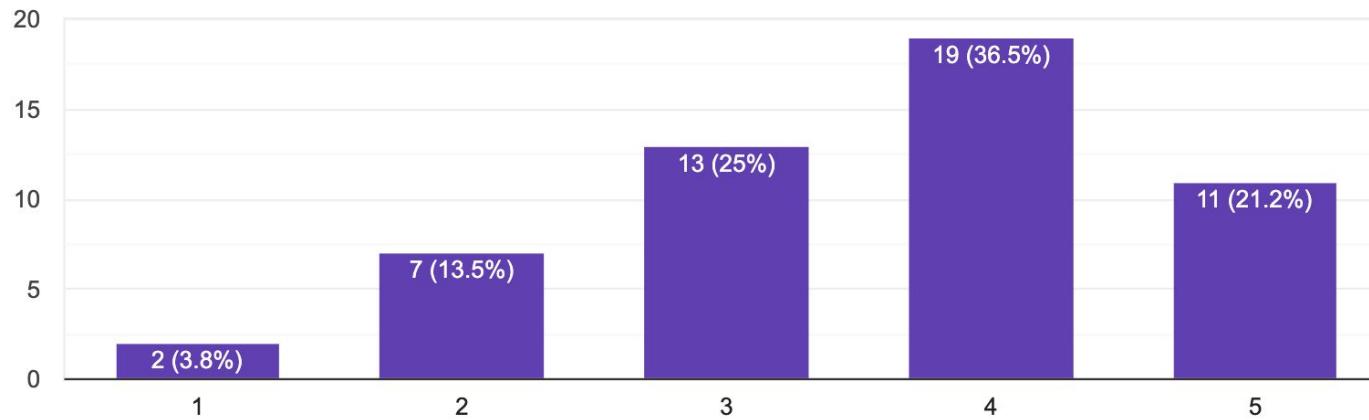
General need for gene terms

Rank the utility of some hypothetical ontologies

Rank the utility of a hypothetical ontology that includes specific genes (e.g. mouse SHH) as terms



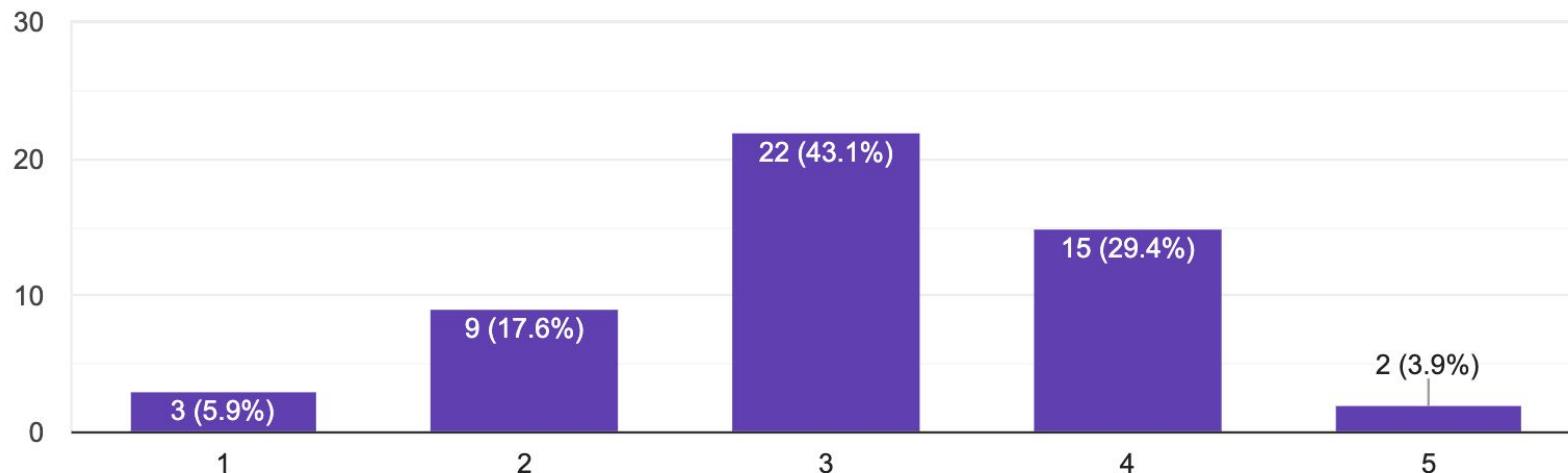
52 responses



Slightly less for transcripts

Rank the utility of an ontology that includes specific transcripts (e.g. ENST00000297261) as terms

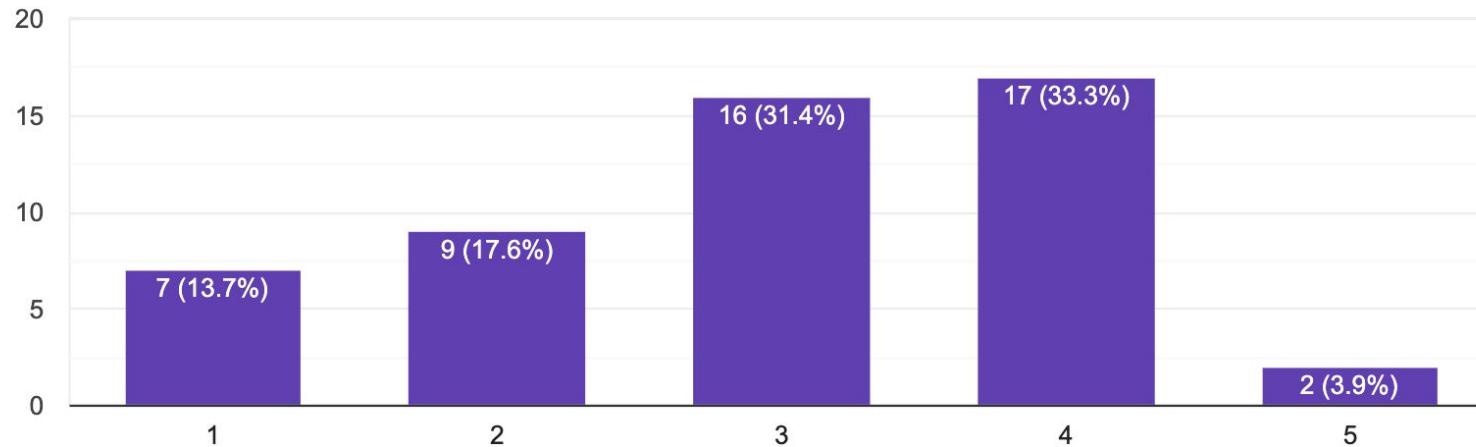
51 responses



Similar for other genome features

Rank the utility of an ontology that includes specific genomic features of other categories (e.g terms for exons such as ENSE00001086614, terms for specific introns, regulatory regions)

51 responses

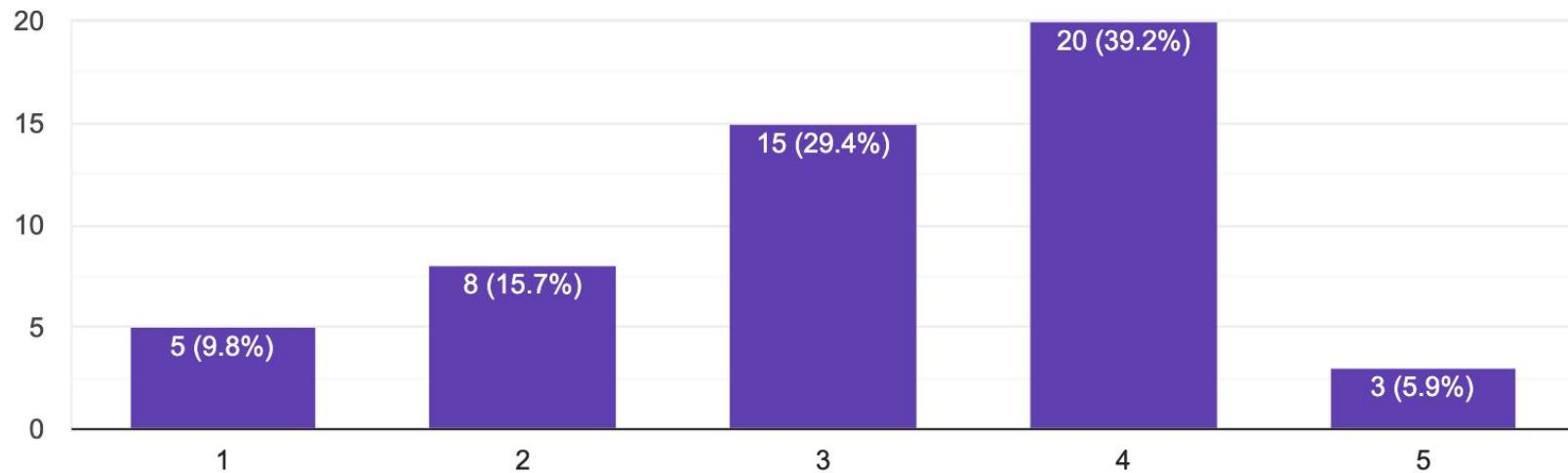


variants

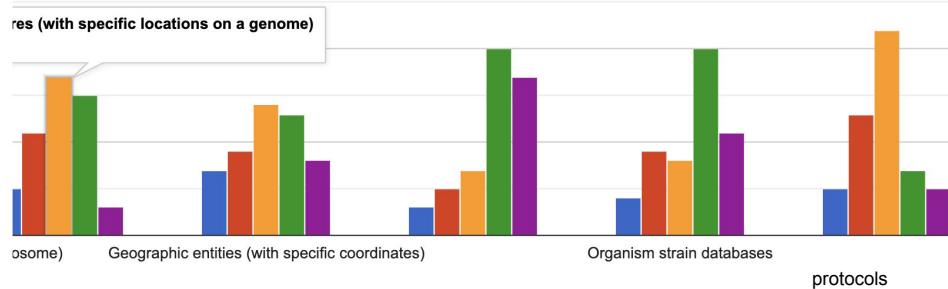
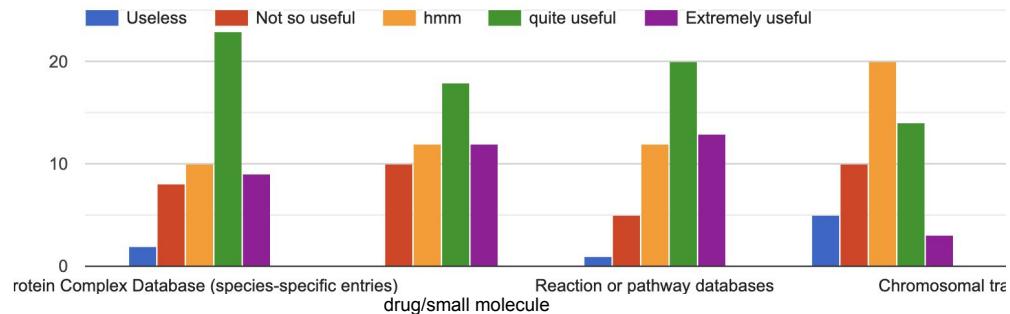
Rank the utility of an ontology that includes specific variants (e.g. Rs1234; BRAF V600E) as terms



51 responses



Rank the utility of hypothetical ontologies mirroring the following databases (rank the utility of the ontology version of the resource rather than the resource itself)



What other kinds of database entities do you think it would be useful to ontologize?

17 responses

co-occurring conditions

Xrefs

product descriptions

Reagents targeting specific genes

Plant diseases

patient response to drug terms (SD, PD, PR, etc)

Do you want to make a dictionary of dictionaries?

material sample types an digitization as far as not covered in obi/iao

nothing else biomedicine-specific, but need generic and conflated classes (which I'm creating for my work)

There is a danger here of trying to over-ontologize. Would like to hear more about this idea and what benefits it would have.

protein domains

Clinical documentation

Codes, terms, menu selectors; resources of any type

Ideal world, a legion of ontologists and infinite resources, I guess all of them.

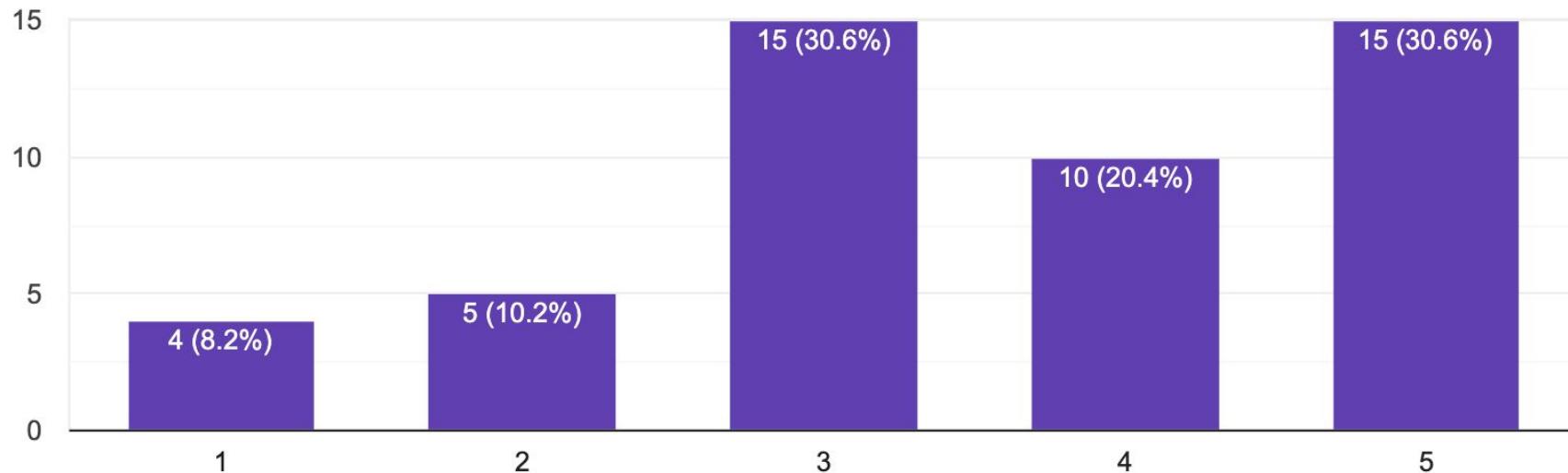
Genes is probably the most important

PDB, genebank, ensembl, ...

Social determinants of health

If such ontologies were to exist, do you think these belong in OBO

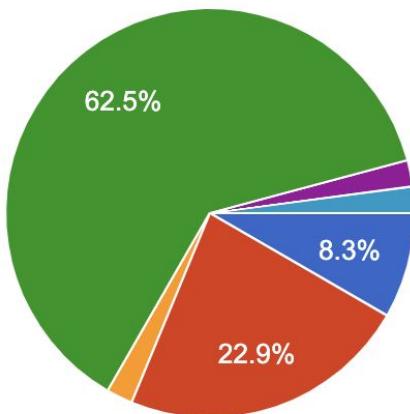
49 responses



Ontology construction and identifiers

How should ontologies be constructed where authoritative databases already exist?

48 responses

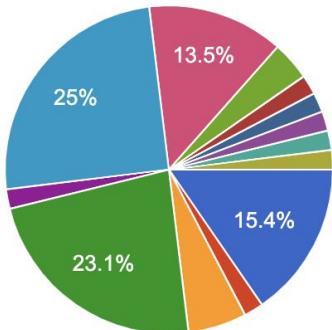


- Automatic translation from a single authoritative reference database (e.g....)
- Automatic translation and merging from multiple authoritative reference databa...
- By manual curation
- By a mixture of manual curation and automatic translation
- Still not convinced such ontologies are needed or would be value added
- All the above, on a case by case basis...

For an ontology that has terms corresponding to specific genes (e.g. human SHH), what should the identifiers look like? (in OWL, this is the last part of the class PURL)



52 responses



- Entirely new IDs should be created
- ENSEMBL gene IDs
- NCBI gene IDs
- Authoritative model organism databases
- Uniprot IDs
- Some mix of the above as appropriate
- I have no strong opinions or don't understand
- Gene symbols

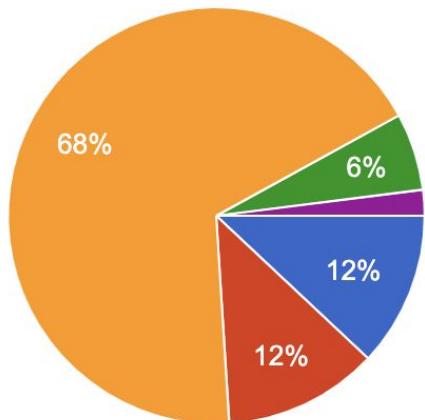
▲ 1/2 ▼

- Definitely a pull between most-widely used and most-rigorously defined. Aut...
- I believe eit should be the nucleic acid forms (DNA for DNA genomes and RN...
- avoid new IDs at all cost - plenty of good options listed here. MOD or HGNC wo...
- My mind is boggling at what could go wrong.....
- This is a meta-system design question...

▲ 2/2 ▼

Should authoritative reference databases be responsible for producing the ontology representation? (this question is about an ontology representation of the core entities in the database, not about terms used to classify these entities)

50 responses

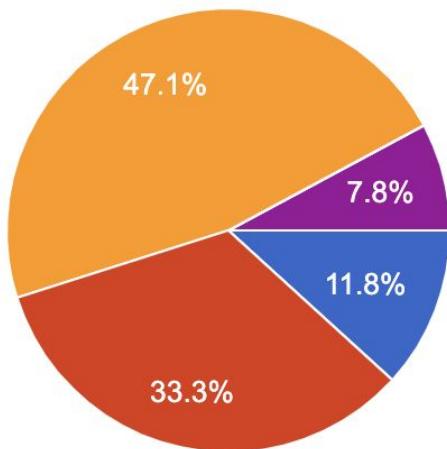


- Yes, authoritative databases should provide the OWL
- I don't think there should be an OWL/ontology representation of most of these databases
- The OWL should be produced as a collaboration between the reference d...
- I don't understand the question
- The genomes represent the information in the most compact way. I wonder wh...



Where more than one authoritative database exists for an entity type, what is the best way to proceed for producing an ontology/ies?

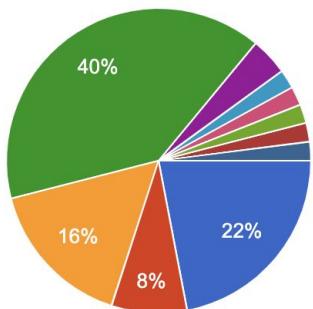
51 responses



- Select one as the canonical source
- Have ontology translations of each one and coordinate mappings
- Perform integration and entity normalization as part of ontology build process
- I don't understand the question
- I disagree with the premise, e.g. I don't think these ontologies should exist

Do you have a preference for what the OWL class PURLs should look like for these ontologies

50 responses



- OBO PURLs
- Identifiers.org
- Using the authoritative databases do..
- I don't understand or don't care
- A new system designed specifically for this
- I think the PID for the alignment is more appropriate
- Don't think there is necessarily a good answer
- Can't say with any authority

▲ 1/2 ▼

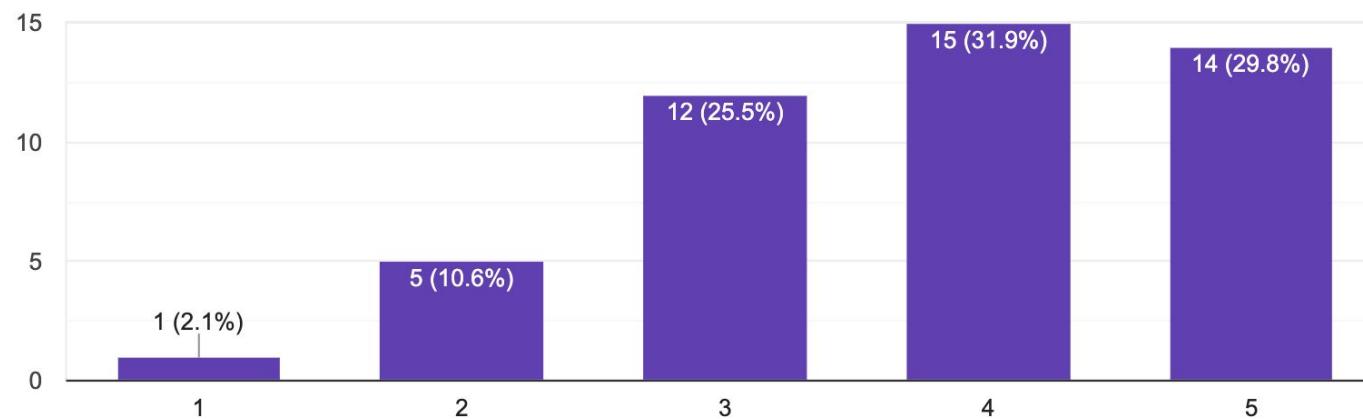
- Taking your question to mean 'generic' PURLs (because any identifier format is appropriate), the important thing is that the identifiers be FAIR, public, and well curated; I prefer plain IRIs
- Any purl that redirects to the authoritative database to get more info on term

Ontology Content

How important is breadth of coverage of the ontology? (compared to reference authoritative databases)

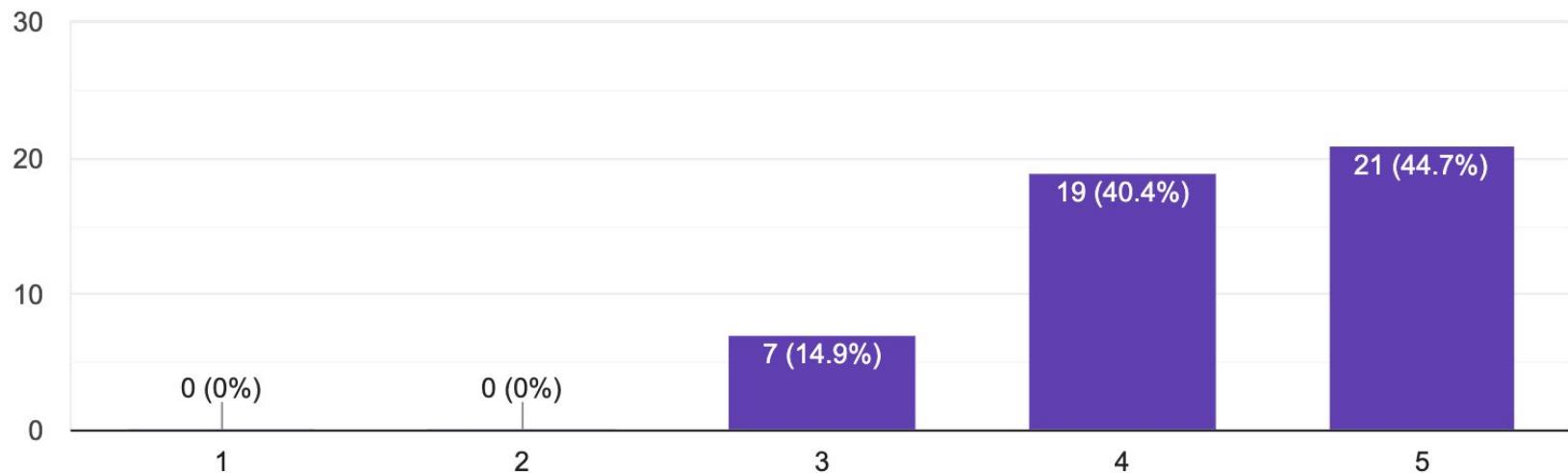


47 responses



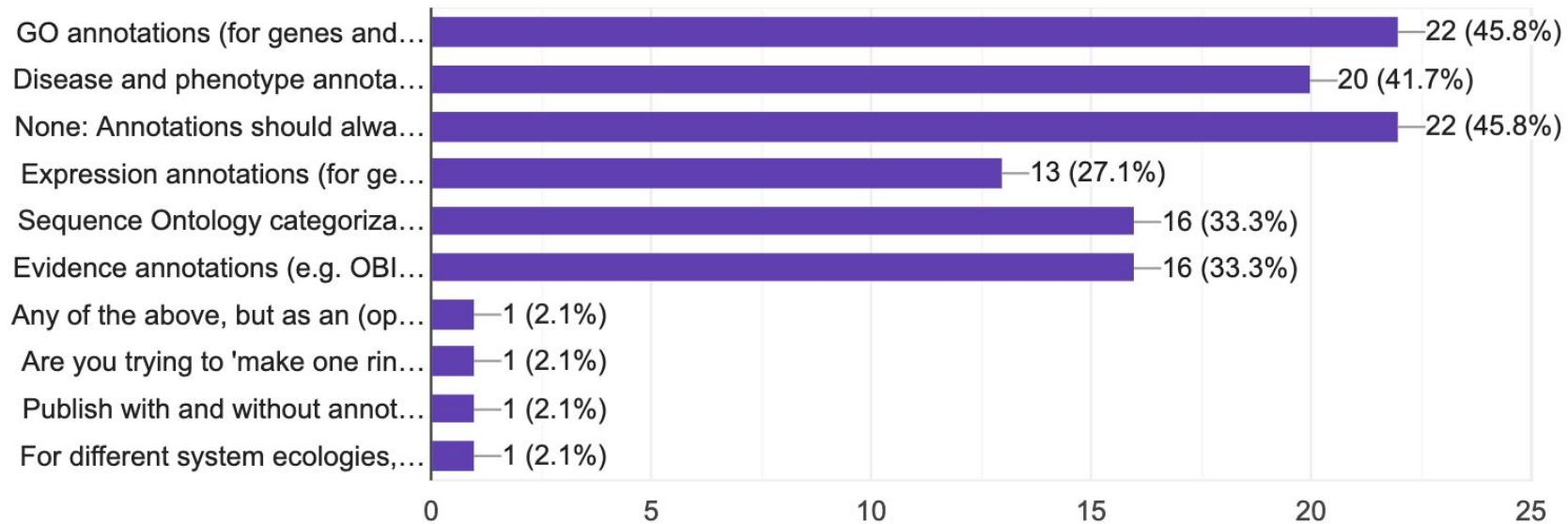
How important is it for terms in the ontology to be interlinked (e.g. SHH gene linked to its transcripts, and the transcripts to exons, variants to genes)?

47 responses



Which ontology annotations should be included in the ontology?

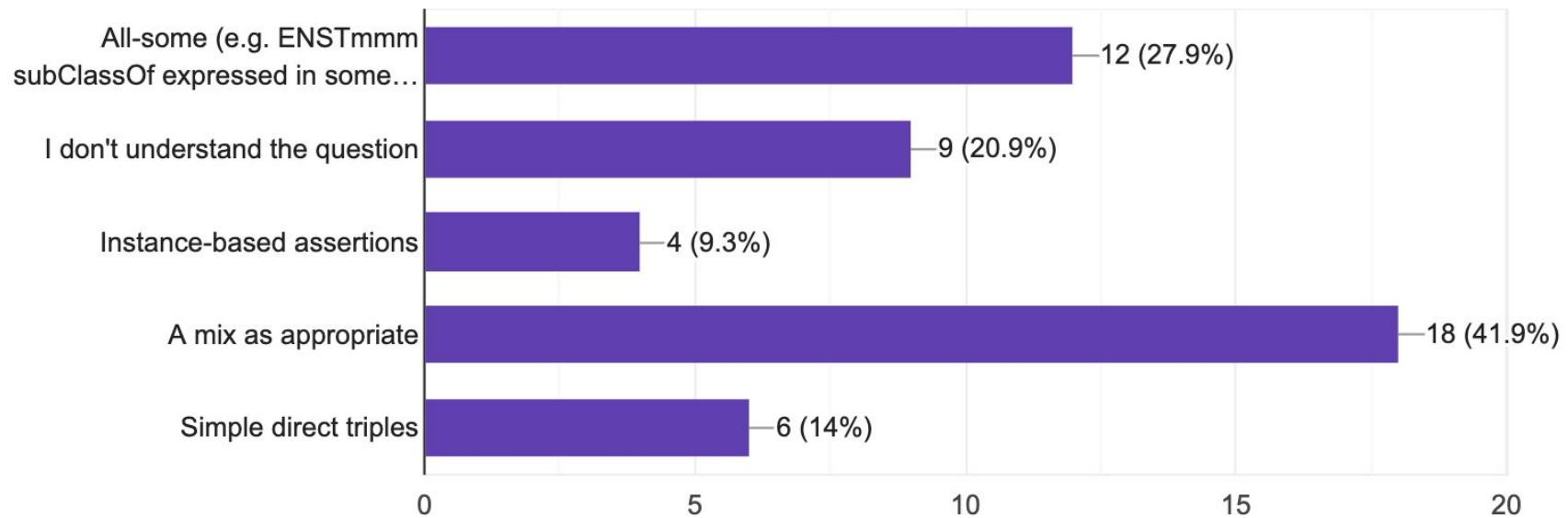
48 responses





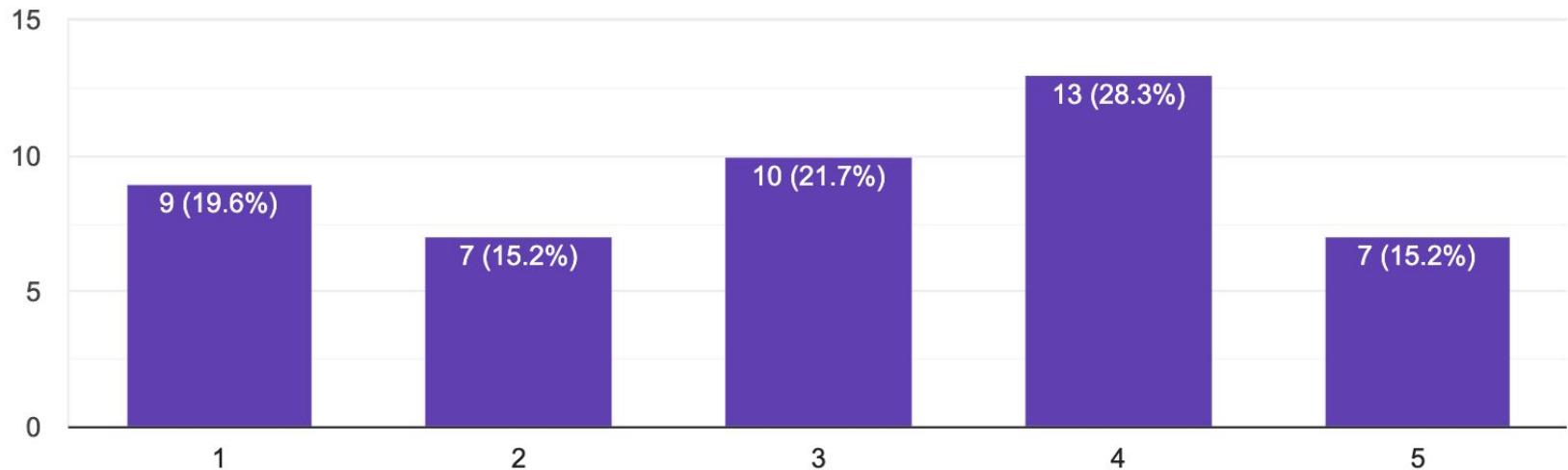
For linked entities and annotations, what kind of axiom should be used?

43 responses



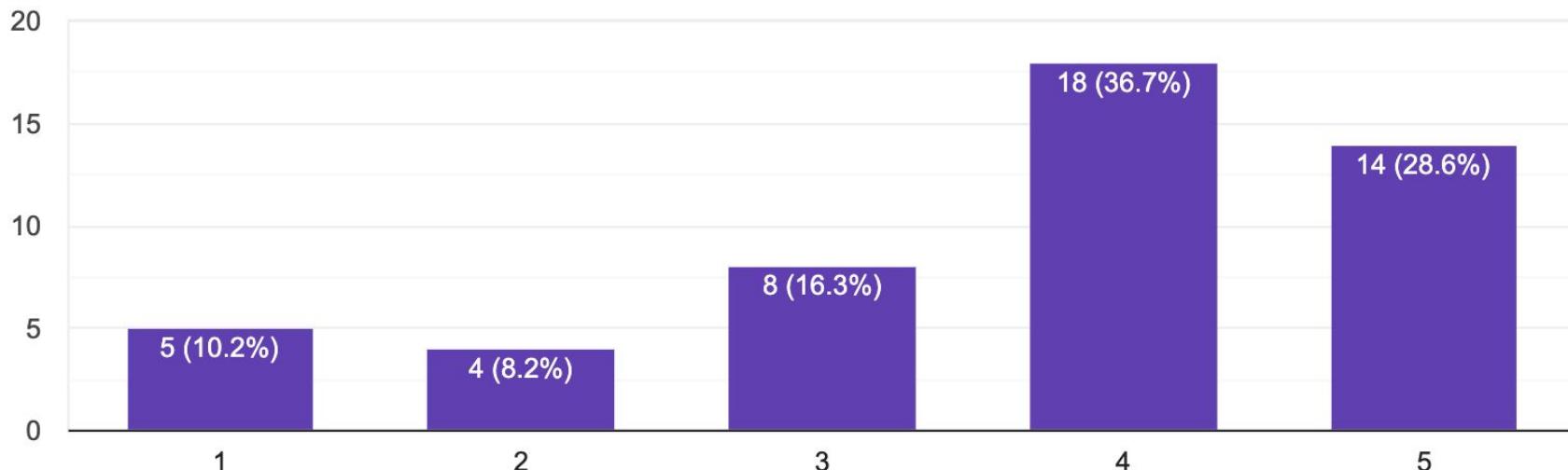
How important is it that the structure of the ontology follows formal upper ontologies (e.g. BFO)

46 responses



How important is having a hierarchical structure in such an ontology?

49 responses



If you can, provide an example of what kinds of hierarchical groupings would be useful (e.g. classifying genes into gene families)

18 responses

gene families, paralog groups, ortholog groups for cross species ontologies

HLA class I alpha chains.

by species. I am concerned that families are hypotheses and that unless the evidence for their creation is included, it may not make sense to ontologize.

by encoded protein domains and by function

transcription factors

gene families, sequence similarity

I am interested in annotation of proteins that bind to gene promoters and enhancers. Annotating these will need new annotation guidelines. We are now ripe to formulate those. Ultimately, the DNA sequence itself rules the process. The hierarchy has been proposed in the Sequence Ontology and now GO terms for binding to these SO DNA elements will be used in GO. I wonder how the splicing community is dealing with the question of hierarchical groupings of RNA sequence elements involved in splicing.

multi rooted hierarchy along different interests

not sure if binning is always appropriate, maybe use similarity scores

groupings are arbitrary; better to use a flat list and apply appropriate attributes to each entity

gene/gene product families

Recursions are the most useful for genome features.

classifying cell lines by tissue of origin

classifying genes into families, classifying protein domains into families and a hierarchy, classifying variants (potentially huge numbers of items that would benefit from classification and queries on subgroups)

Groupings that follow natural understanding and meaning, if an unambiguous primary organization exists

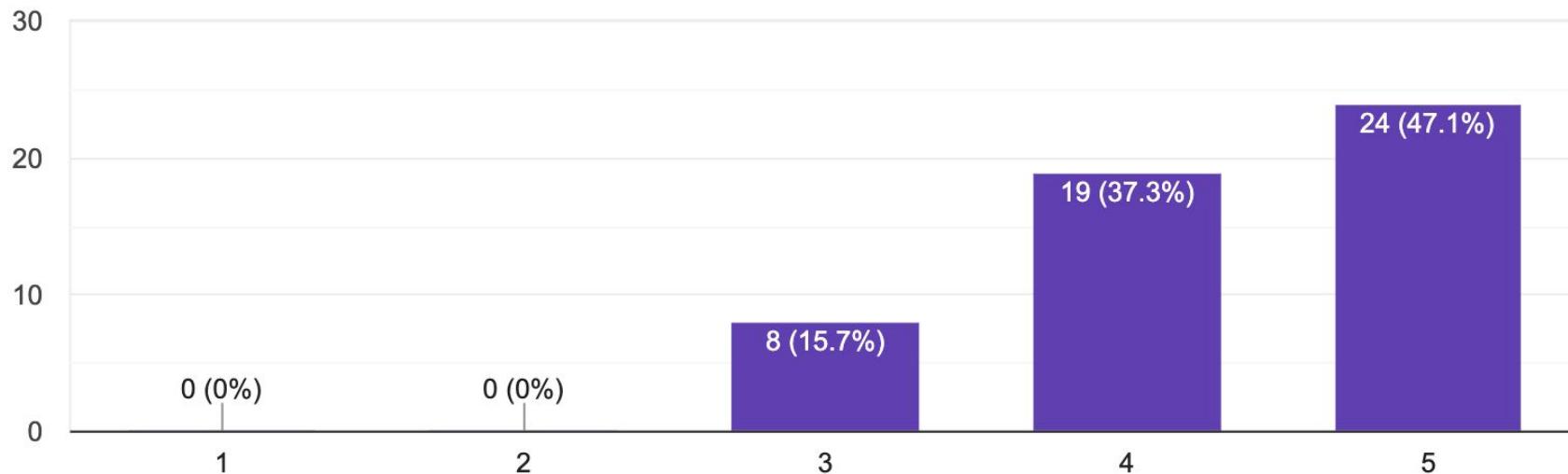
gene classes into upper gene classes up to a species-neutral uncommittting class (like PRO 1:1 homologues)

Protein families, functions, gene families

functional (paralogy vs. orthology) and structural (homology)

How important is it that information in the ontology is current with respect to any upstream authoritative databases

51 responses



Additional comments, final thoughts, things we could have been clearer about in this survey, etc?

18 responses

In my opinion, the most important thing is that biological databases **are** available as ontologies (ideally as both OWL and OBO), not exactly **how** they are represented (providing the ontology is sensibly designed). For example, on my project, we have been repeatedly stymied by the fact that no ontology exists to describe named plant diseases (meaning we cannot curate this data using an established controlled vocabulary). For ontologies to be truly useful, I think their coverage has to be as broad (and ideally deep) as possible, and their relations to other data sources should be emphasised.

With regards to the survey: it sometimes wasn't clear whether the questions were referring to OWL or OBO versions of the hypothetical ontology, particularly when the question concerned features that are beyond the scope of OBO (e.g. ontology imports).

I think the scope of what it means to ontologize a database is a bit unclear. I am concerned about dependencies on one resolver, and also that identifier provisioning by the source DBs does not always follow best practices. DBs could follow a new standard for data release that is "born interoperable," we've discussed this and written grants about it for years. Maybe we are ready now :-).

It's a very good thing to do. I am glad to collaborate. Thanks!

Already in your work with GO, you have made efforts to study and revise the GO knowledge graphs that consist of nodes and edges. For most biologists, the nodes are proteins (for me also DNA elements are such) and the edges are relations. For GO itself, the nodes are terms and edges are the hierarchical relations between the nodes and this is the basis for annotation and also GO-CAM based reasoning. I remember you making edges into nodes at your talk in April 2018 at EB for GREEKC. This amounts to reification. This is in my opinion an essential process of computation, but as largely not done, because of its pitfalls. Nodes should be material entities, few people see material entities as relations. However, in some situations ontology terms/classes are nodes and in others they are edges that link nodes. With your proposal to make any knowledge base an ontology, I believe you empower this. But, I wonder how many black holes will occur, where we lose sight of the horizon we wished to attain because it is computationally not possible to perform the operation 'make an ontology of this'. I would be glad to talk about this generally, but over and above that, specifically, for biochemical regulatory networks that govern gene transcription, and how to describe them, as they are ultimately the basis for cell type specification.

I only use OBO things in their OWL form

I like the way the survey forces users to address difficult choices, but I think it confuses the issue of ontology building with being able to refer to genetic entities defined in DBs in ontologies. We can't get around that people want/need to refer to these entities as part of the logical structure of ontologies, but maybe all we need to support this are a set of guidelines and recipes (or perhaps a service) for doing so in a more standardised manner than we do currently, rather than building ontology artefacts (the exception here is NCBI taxonomy, which is essential for lots of ontology building and needs structure).

Imagine an endpoint that I could use to generate standardised representations when provided with a list of gene, protein or transcript IDs, maybe combined with some configuration to specify what gross type of entity I wish to refer to (gene, protein, transcript). This could be backed by some standard mapping service like MyGene. It could return an entity using identifiers.org for the URL along with annotation of version, synonyms + a set of xref (SSOM?) mappings to the entity in other databases. Cross-species terms could work as long as tied to a single species. They could have a derived ID and, as PRO does, a statement that the term refers to orthologs of the gene X in species Y. Resolution of derived IDs would be more of a problem, but could be solved by a service endpoint.

I tried to answer quickly based on my intuitions, but a lot of these questions deserve careful analysis. There were some forced choices that didn't capture exactly what I'm thinking. Overall, it's a great set of questions covering a lot of terrain.

very comprehensive survey :-)

thanks for putting in the "I don't understand this" option for those of us who need to learn more about this area!

These are my opinions today. I may change my mind after your talk. I could see a usefulness for such ontologies potentially in the future once knowledge and gene structures are more stable. Even then this would be a combinatorial nightmare. Curation enables the association of variants of entities with phenotypes, and modified forms with GO terms. A knowledge graph with entities alone would not be useful, it would need to be all or nothing. A good database (i.e using Chado) allows the organization of objects in a hierarchical way anyway, for querying, so what would be advantage of instantiating this in a graph? The churn would be enormous . Ontologies change a lot, objects and annotations change more. This would seem to be another maintenance overhead when resources are scarce. That said, there are emerging knowledge graphs like Knetminer, that seem as though they might be useful for datamining. This could be ontologised, but I don't really see the point right now.....

These are my initial gut reactions -- open to being convinced!

Many of the questions assumed support for ontologizing entities so made it difficult to answer them. I am still unclear what is driving this idea and what value it adds. It is very interesting but I just don't see it as adding much value.

I think you should have started with a more definitive definition for the kind of ontology you are talking about - I'm inferring that you meant specifically an OWL-DL representation. While I think many 'database elements' discussed would benefit from OBO involvement, I don't think OWL provides a huge benefit in many cases (particularly those involving very large collections with relatively simple representations.). I think core elements of OBO are useful and applicable outside of the context of OWL-DL. These include standards for identifiers, relation types, annotations. e.g. OBO standards like these applied to simple graph structures would be very valuable. My recent experience suggests that the goal of OBO should be to deliver one interlinked knowledge graph covering each of the member OBO ontologies - represented as json text files.. In the context of this particular discussion, this knowledge graph should indeed contain nodes for the 'database entities', e.g. genes, in question. I think the OWL-DL aspect of these ontologies is most important as part of curation, testing, and build processes and as far as downstream consumers are concerned should be as hidden as a Jenkins script. (That makes me very sad BTW, but this is what I am seeing.).

For the purposes of answering these questions, it might be helpful to understand the intended use/use cases of the hypothetical ontologies to understand the scope

Many questions implied there was one correct answer, but the best answer would often be context-driven.
Good survey nonetheless.

Fun survey!

The "useless" category in some of the questions should be interpreted as "not relevant to my work"

I had some second choices, but not way to express (e.g. on annotation as external)