

Introduction to Ontologies

Nicole Vasilevsky

Programming for Biology

Cold Springs Harbor Laboratory

October 26, 2018

Hello!



- Lead Biocurator, Translational and Integrative Sciences Lab
- Research Assistant Professor, OCTRI and DMICE, Oregon Health & Science University



mondo
MONARCH DISEASE ONTOLOGY



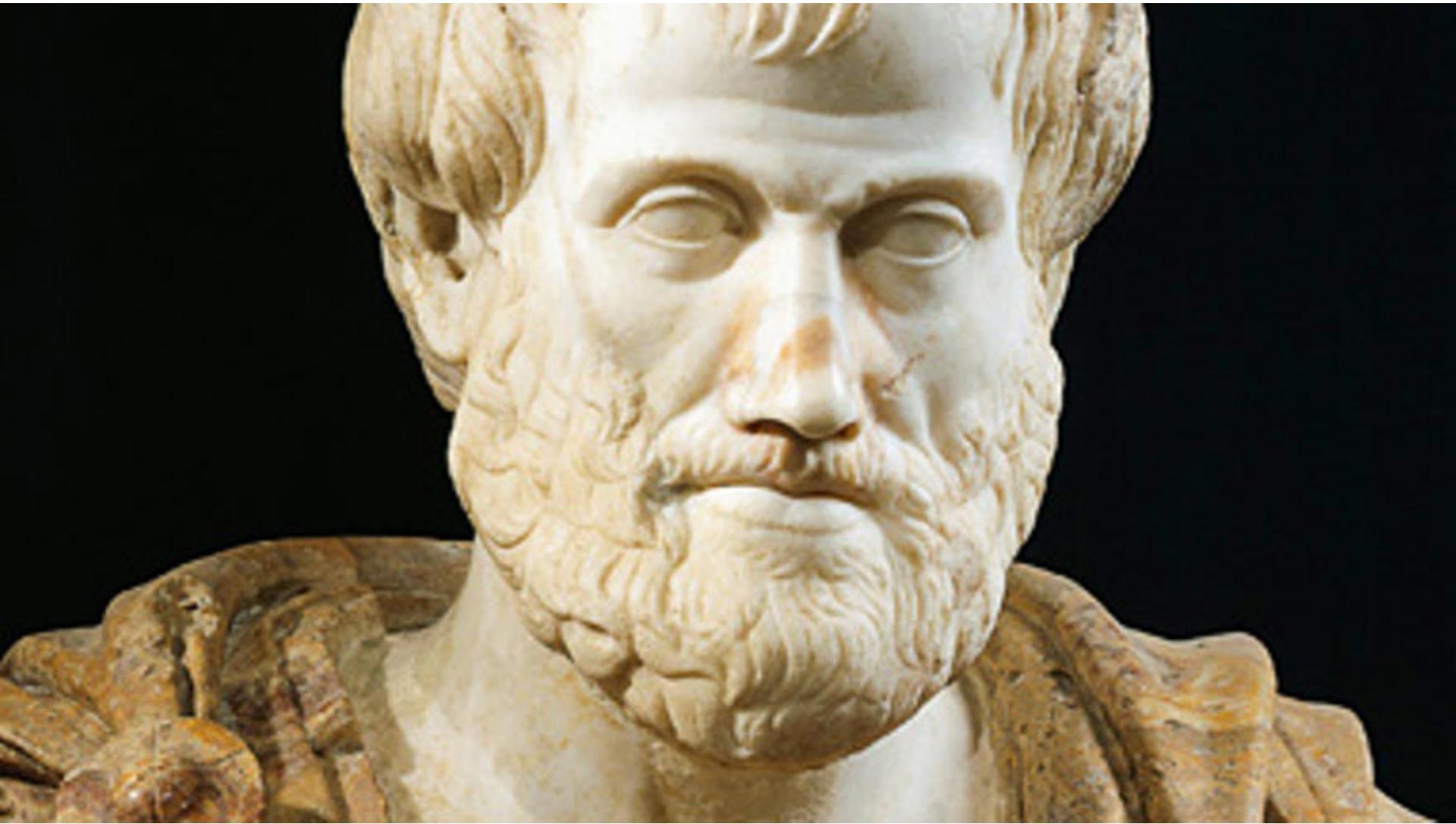
Learning objectives

- Describe what ontologies are
- Describe why ontologies are valuable
- Retrieve ontology terms using various tools
- Provide specific examples of how ontologies are used
- Contribute to ontology communities/submit tickets
- Describe methods for building ontologies
- Search for genes that have been annotated with GO terms

Part 1:

Introduction to Ontologies

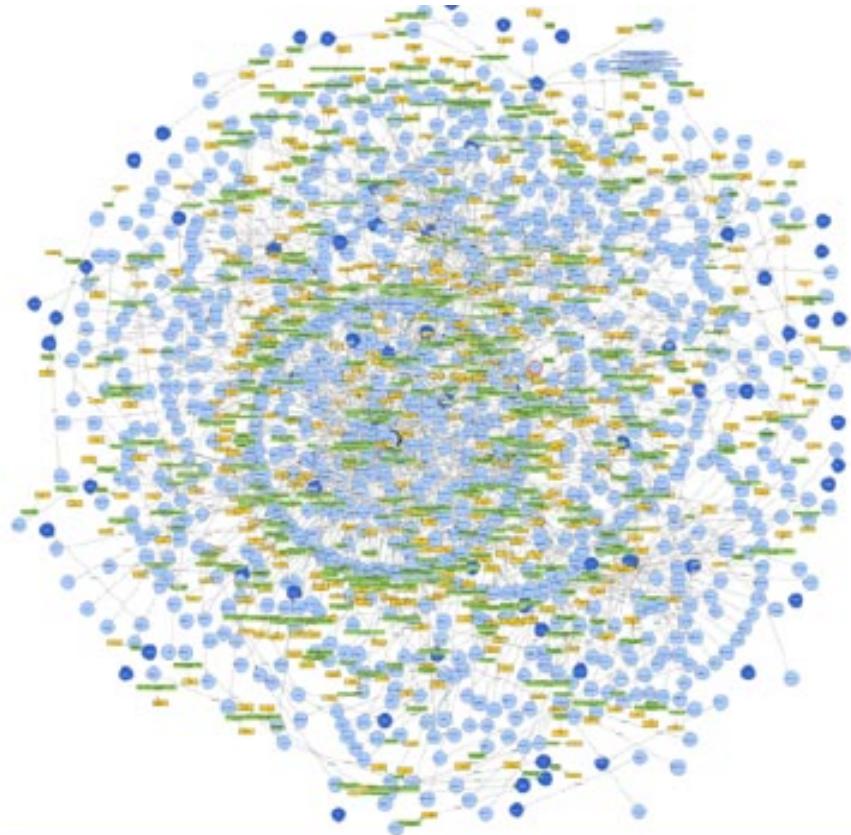
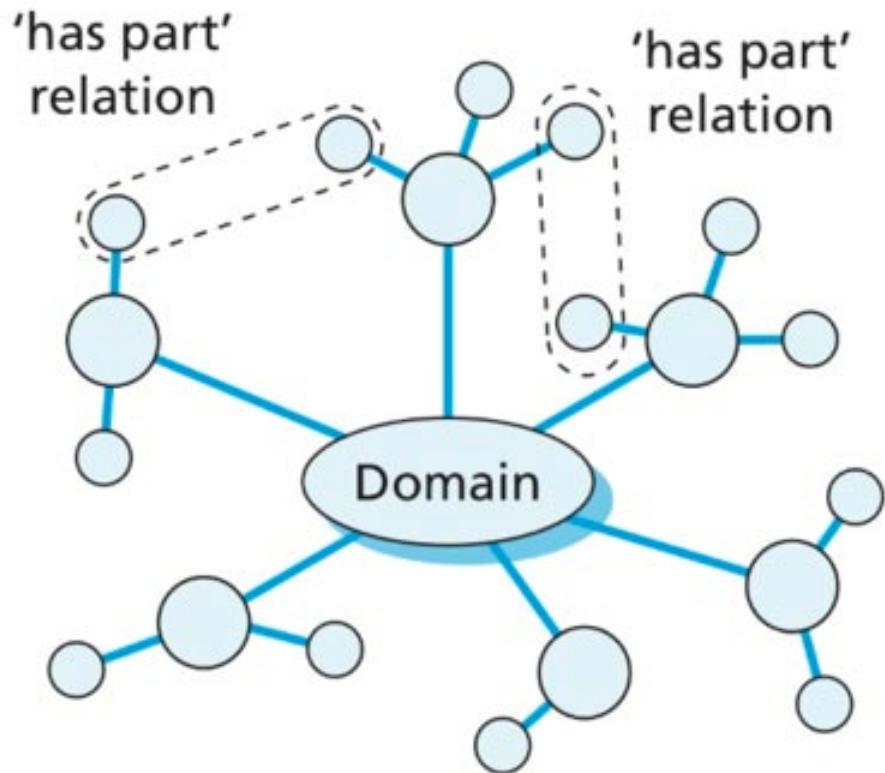
Aristotle (384 BCE–322 BCE)



“There is a science which studies Being qua Being, and the properties inherent in it in virtue of its own nature”

[Peter N. Robinson, Sebastian Bauer. Introduction to Bio-Ontologies]

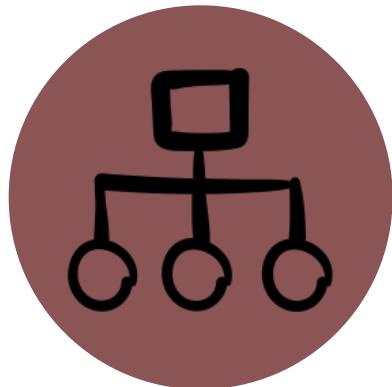
Ontologies are systematic representations of knowledge that can be used to integrate and analyze large amounts of heterogeneous data



Need to structure data



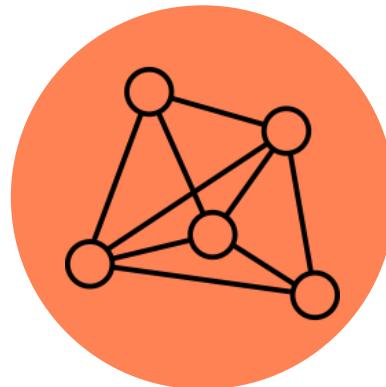
What can you do with an ontology?



Organizing data



Filtering data



Connecting data



Suggesting data

Ontology Basics

Controlled Vocabulary

(in its simplest form)



Definition:

Any closed,
prescribed list of
terms

Key Features:

- Terms are not usually defined
- Relationships between the terms are not usually defined
- Simplest form is a list



CV of wines:

- *Pinot noir*
- *Red*
- *Chardonnay*
- *Chianti*
- *Bordeaux*
- *Riesling*
- ...

These are all different types:

- color
- location
- varietal

Taxonomy

- **Definition:**
Any controlled vocabulary that is arranged in a hierarchy
- **Key Features:**
 - Terms are not usually defined
 - Relationships between the terms are not usually defined
 - Terms are arranged in a hierarchy
- **Taxonomy of wines:**
 - Red
 - Merlot
 - Zinfandel
 - Cabernet
 - Pinot Noir
 - White
 - Chardonnay
 - Pinot Gris
 - Riesling

Thesaurus

Definition:

A taxonomy that contains additional information about use of the terms

Key Features:

- Terms are not usually defined
- Relationships between the terms are not usually defined
- Terms are arranged in a hierarchy
- Scope notes or instructions for use are included

Examples:

- WorldNet
- NCI
- Cancer
- Thesaurus
- MeSH

Parkinson Disease MeSH Descriptor Data 2018

[Details](#)[Qualifiers](#)[MeSH Tree Structures](#)[Concepts](#)

Nervous System Diseases [C10]

Central Nervous System Diseases [C10.228]

Brain Diseases [C10.228.140]

Basal Ganglia Diseases [C10.228.140.079]

Parkinsonian Disorders [C10.228.140.079.862]

Lewy Body Disease [C10.228.140.079.862.400]

Parkinson Disease [C10.228.140.079.862.500]

Parkinson Disease, Secondary [C10.228.140.079.862.800] 

Nervous System Diseases [C10]

Central Nervous System Diseases [C10.228]

Movement Disorders [C10.228.662]

Parkinsonian Disorders [C10.228.662.600]

Lewy Body Disease [C10.228.662.600.200]

Parkinson Disease [C10.228.662.600.400]

Parkinson Disease, Secondary [C10.228.662.600.700] 

Nervous System Diseases [C10]

Neurodegenerative Diseases [C10.574]

Chronic Traumatic Encephalopathy [C10.574.250]

Heredodegenerative Disorders, Nervous System [C10.574.500] 

Lewy Body Disease [C10.574.531]

Motor Neuron Disease [C10.574.562] 

Multiple System Atrophy [C10.574.625] 

Paraneoplastic Syndromes, Nervous System [C10.574.781] 

Parkinson Disease [C10.574.812]

Postpoliomyelitis Syndrome [C10.574.827]

Prion Diseases [C10.574.843] 

Subacute Combined Degeneration [C10.574.910]

Tauopathies [C10.574.945] 

TDP-43 Proteinopathies [C10.574.950] 

Parkinson Disease MeSH Descriptor Data 2018

[Details](#)[Qualifiers](#)[MeSH Tree Structures](#)[Concepts](#)

MeSH Heading	Parkinson Disease
Tree Number(s)	C10.228.140.079.862.500 C10.228.662.600.400 C10.574.812
Unique ID	D010300
Annotation	drug ther: consider also ANTIPARKINSON AGENTS ; /chem ind = PARKINSON DISEASE, SECONDARY /chem ind
Scope Note	A progressive, degenerative neurologic disease characterized by a TREMOR that is maximal at rest, retropulsion (i.e. a tendency to fall backwards), rigidity, stooped posture, slowness of voluntary movements, and a masklike facial expression. Pathologic features include loss of melanin containing neurons in the substantia nigra and other pigmented nuclei of the brainstem. LEWY BODIES are present in the substantia nigra and locus coeruleus but may also be found in a related condition (LEWY BODY DISEASE, DIFFUSE) characterized by dementia in combination with varying degrees of parkinsonism. (Adams et al., Principles of Neurology, 6th ed, p1059, pp1067-75)
Entry Version	PARKINSON DIS
Entry Term(s)	Idiopathic Parkinson Disease Idiopathic Parkinson's Disease Lewy Body Parkinson Disease Lewy Body Parkinson's Disease Paralysis Agitans Parkinson Disease, Idiopathic Parkinson's Disease Parkinson's Disease, Idiopathic Parkinson's Disease, Lewy Body Primary Parkinsonism
See Also	Lewy Bodies
Public MeSH Note	1979; see PARKINONISM 1967-1978, see PARALYSIS AGITANS 1963-1966
History Note	1979(1963)
Entry Combination	chemically induced:Parkinson Disease, Secondary /chemically induced
Date Established	1967/01/01
Date of Entry	1999/01/01
Revision Date	2013/07/08

What is an ontology?

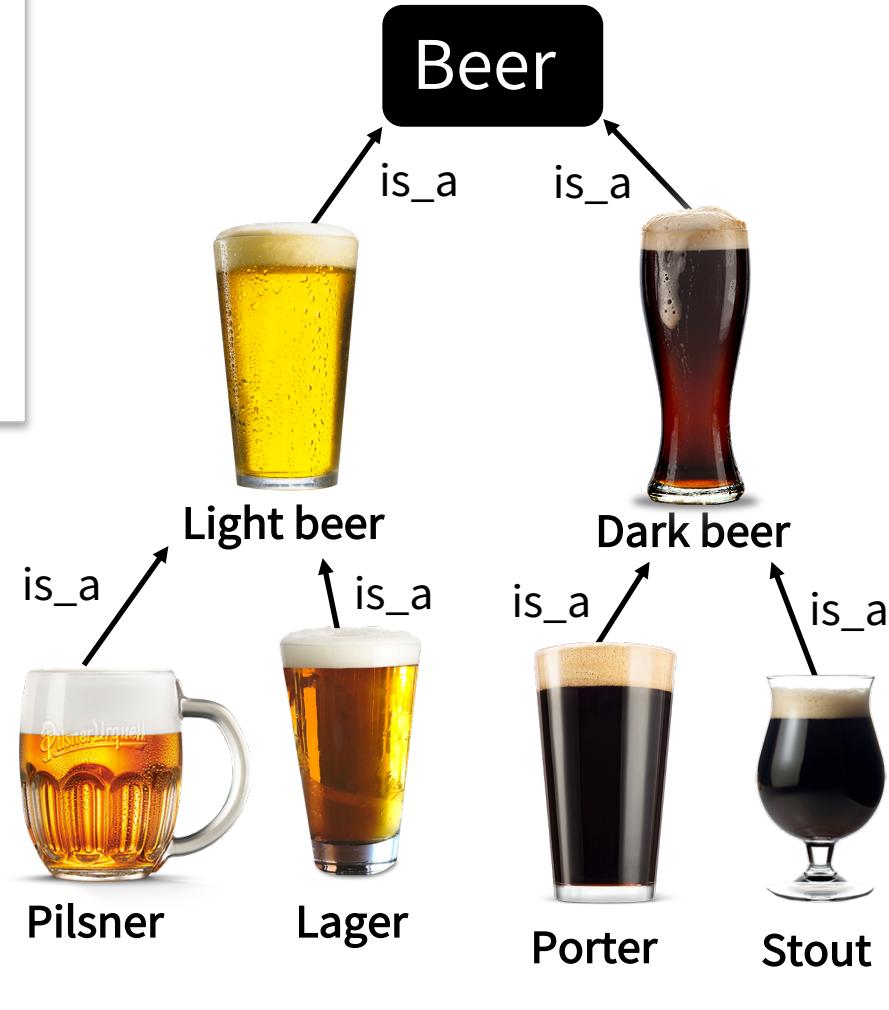
on·tol·o·gy

/æn'täləjē/ 🔍

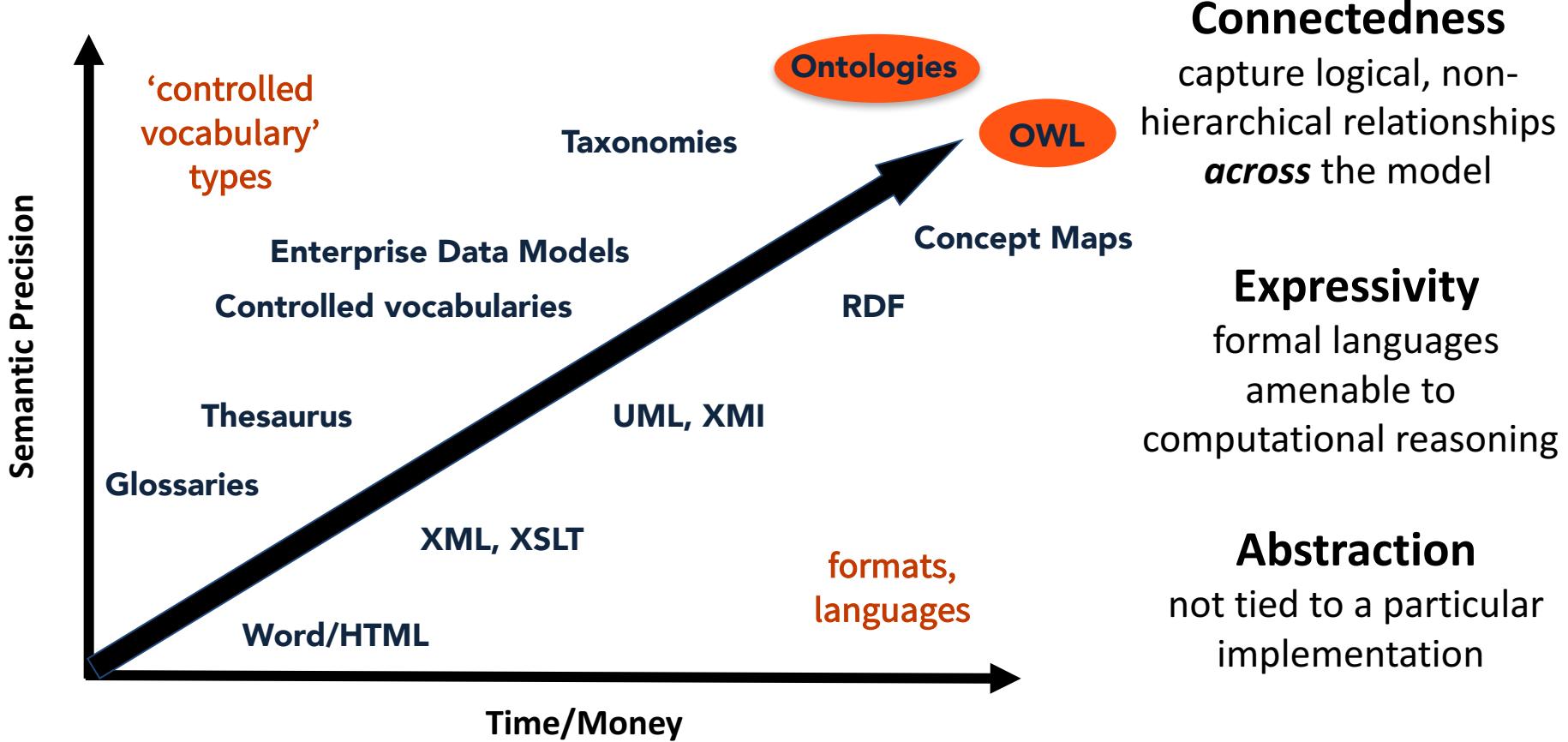
A knowledge classification of a domain, where the relationships between concepts are formally defined and logically related, which allows for computational reasoning

Key Features:

- Terms are defined
- Semantics - relationships between terms are defined, allowing logical inference and sophisticated data queries
- Terms are arranged in a hierarchy
- Expressed in a knowledge representation language such as RDFS, OBO, or OWL



Complexity of Vocabulary Types



Who has heard of ontologies?

Name some specific examples



mondo
MONARCH DISEASE ONTOLOGY



GENEONTOLOGY
Unifying Biology

SNOMED CT
The global language of healthcare

The Sequence Ontology

DISEASE
ONTOLOGY

The Experimental Factor Ontology (efo) logo features a blue bar with the text "Experimental Factor Ontology" and a small blue sphere icon to the left.

The EnvO logo features the word "Env" in a large, black, sans-serif font. To the right of the "o" is a circular graphic containing a stylized Earth.

Ontologies in Everyday Life



Show results for

Movies & TV

- Prime Video
- Blu-ray
- DVD
- Movies
- Kids & Family

▽ See more

Toys & Games

- Action & Toy Figures
- Toy Building Sets

Novelty & More

- Movie & TV Fan T-Shirts
- Women's Novelty Tops & Tees
- Men's Novelty Shirts
- Movie & TV Fan Accessories

Books

- Marvel Comics & Graphic Novels
- Superhero Comics & Graphic Novels
- Comics & Graphic Novels
- Teen & Young Adult Books
- Teen & Young Adult Literature & Fiction

▽ See more

CDs & Vinyl

- Soundtracks
- Pop
- Movie Scores
- Movie Soundtracks

Prime Video

- Movies
- TV

Digital Music

- Soundtracks

▽ See All 24 Departments



Sponsored ⓘ
Guardians Of The Galaxy

\$19⁶⁴ Blu-ray ✓prime
FREE Delivery by Fri, Oct 12
Or FREE One-Day Pickup
More Buying Choices
\$13.69 (18 used & new offers)



Guardians of the Galaxy

Prime Video
\$3⁹⁹ – \$17⁹⁹ Rent or Buy



Guardians of the Galaxy

A group of intergalactic criminals are at control of the universe.

Cast



Chris Pratt
Peter Quill



Vin Diesel
Groot



Braeden Roc

Play trailer



Chris Pratt

American actor



facebook.com/PrattPrattPratt

Christopher Michael Pratt is an American actor. Pratt came to prominence with his television roles, particularly for his role as Andy Dwyer in the NBC sitcom Parks and Recreation, for which he received ...
[Wikipedia](#)

Born: June 21, 1979 (age 39 years), [Virginia, MN](#)

Height: 6' 2"

Spouse: Anna Faris (m. 2009–2018)

TV shows: [Parks and Recreation](#), [The O.C.](#), [Everwood](#), MORE

Children: [Jack Pratt](#)

Movies

[View 25+ more](#)



Guardians of the Galaxy
2014



Jurassic World:
Fallen Ki...
2018



Avengers:
Infinity War
2018

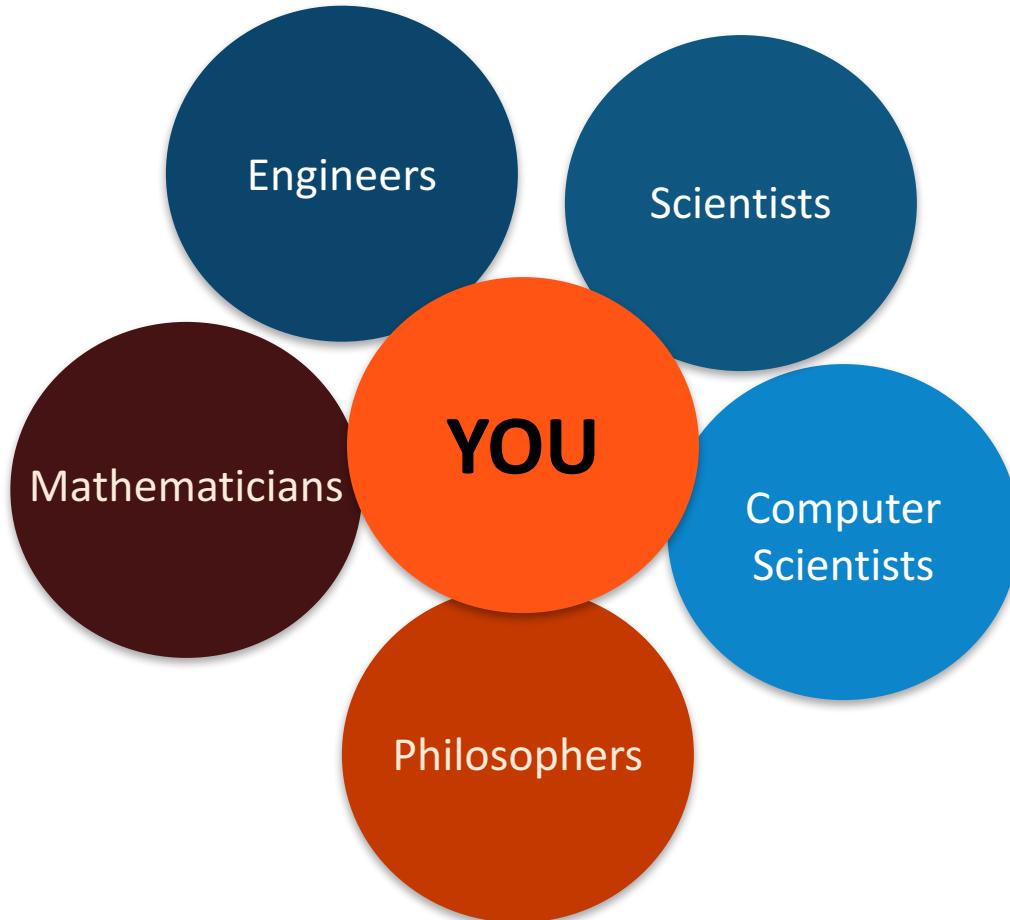


Passengers
2016



Guardians of the Galaxy V...
2017

Who are ontologists?



Let's
review
some
concepts

- What is a definition of an ontology?
- How are ontologies useful?

Ontology Structure

Elements of an Ontology

Classes

Concepts in the ontology,
also called terms

Relationships

Properties, how the classes
relate to each other

Instances

Individuals



Ontologies are expressed in formal language like Web Ontology Language (OWL)

OWL is a semantic web computational logic-based language, designed to represent rich and complex knowledge about things and the relations between them. It also provides detailed, consistent and meaningful distinctions between classes, properties and relationships.

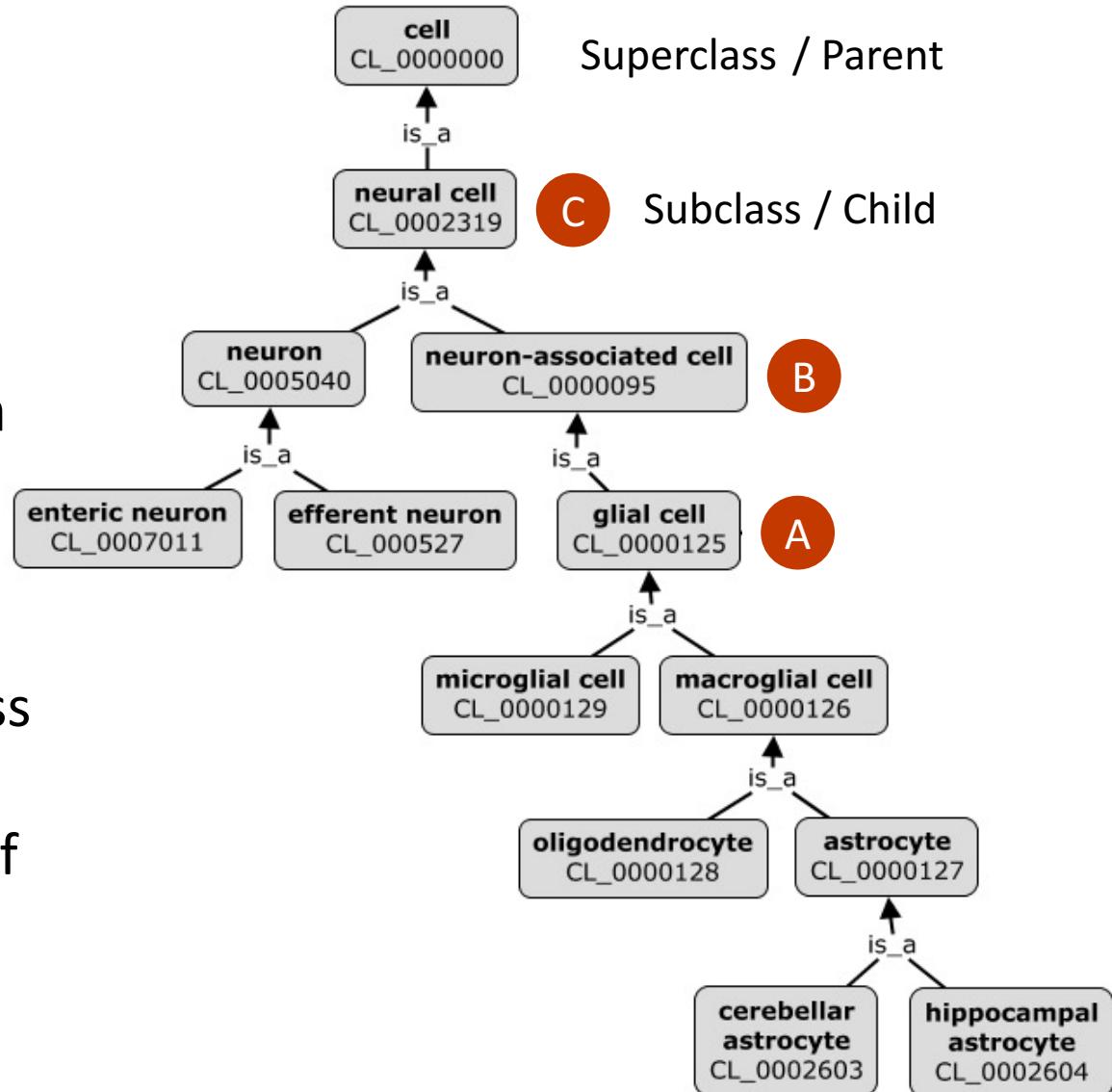
“Why not be inconsistent in at least one aspect of a language which is all about consistency?”

- *Guus Schreiber, Why OWL and not WOL?*



The is_a relationship

- Terms (classes) in an ontology are often classified via an ‘is_a’ relationship
- What is a subsumption hierarchy?
- Which class is a superclass of B?
- Which class is a subclass of C?
- Is an astrocyte a type of glial cell?

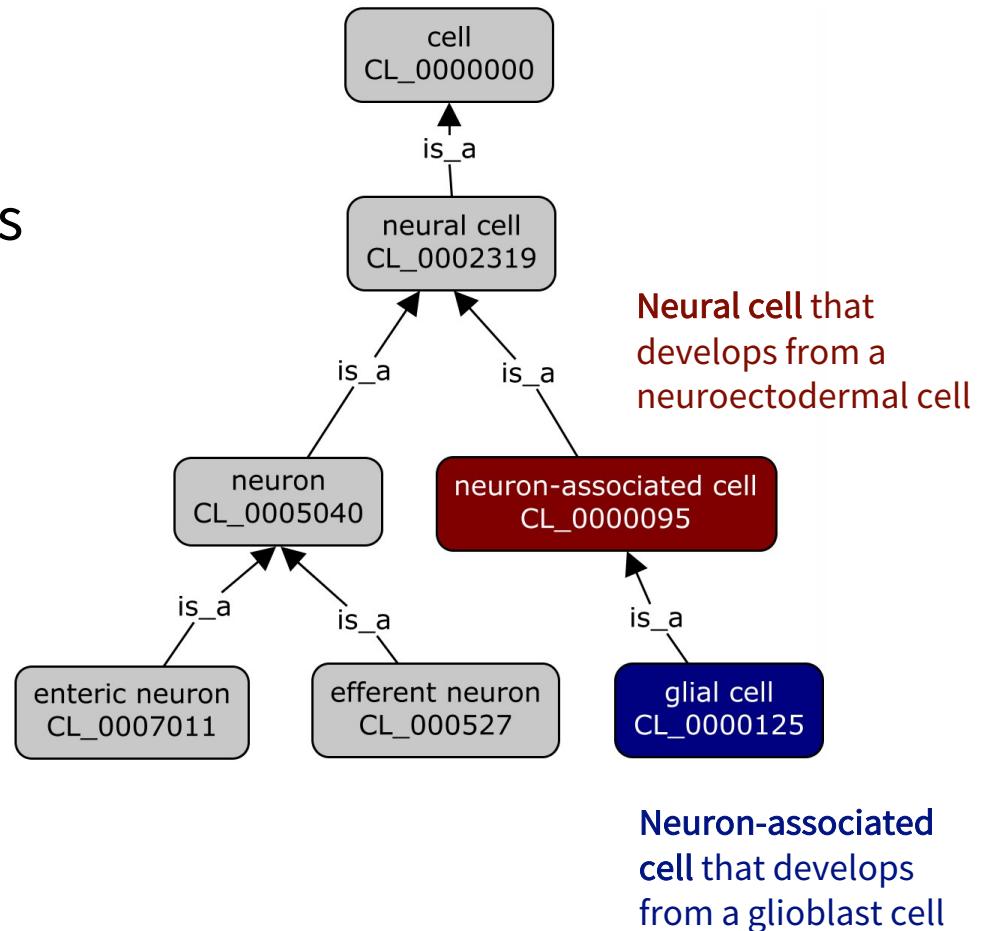


The is_a relationship

- The **is_a** relation is like inheritance
- Children terms inherit the properties and relationships of the parent term

Glial cell

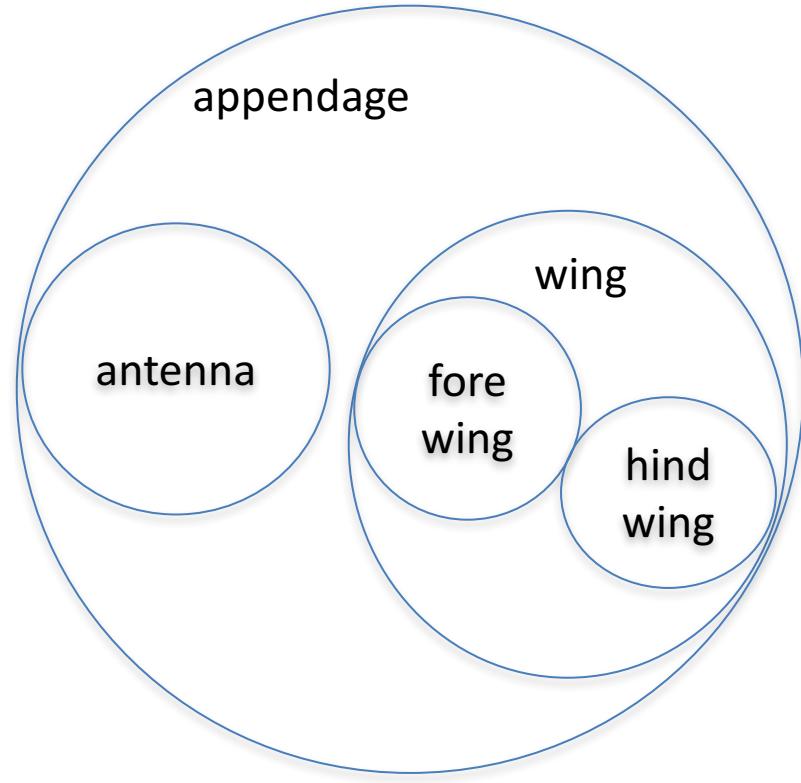
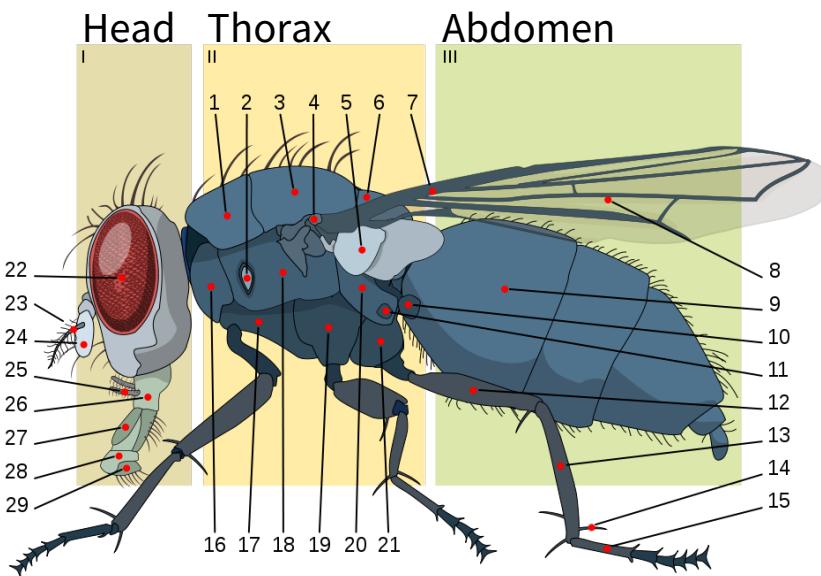
- Is a neuron-associated cell
- Has all the features of a neuron-associated cell with some sort of differentiating features



The is_a relationship – Another example

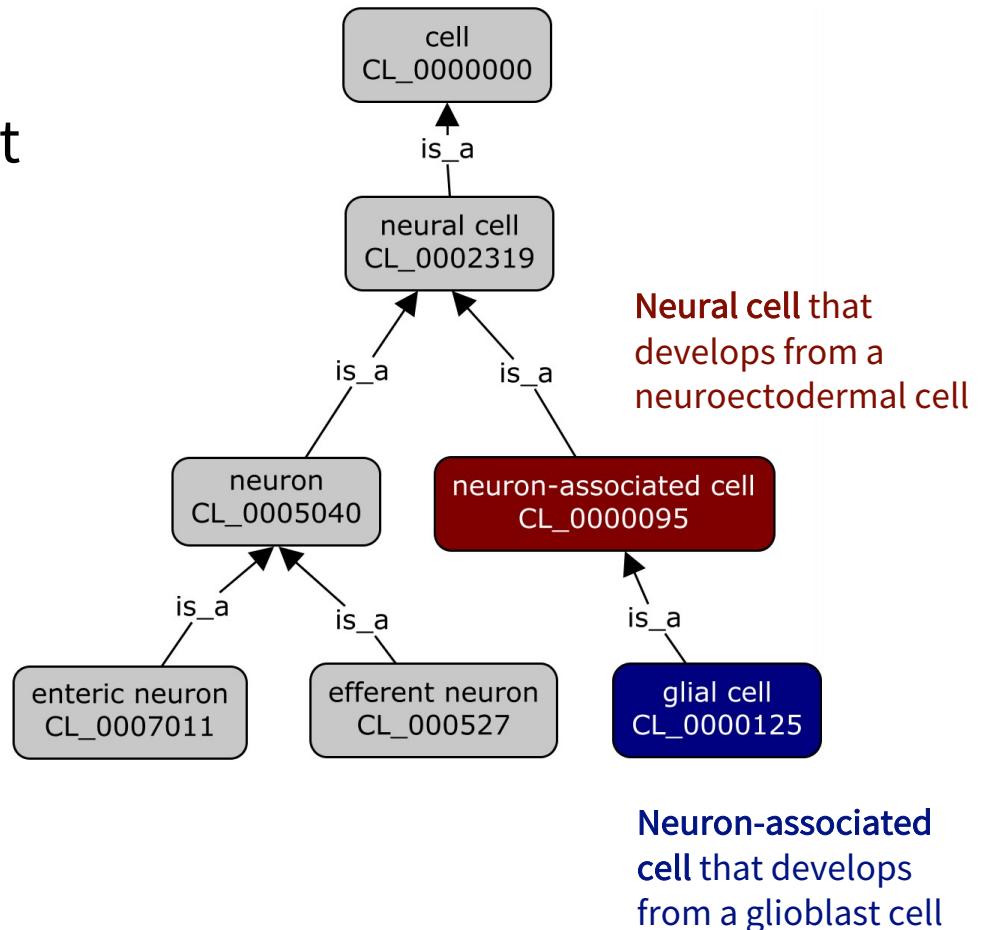
- Classes

- appendage
 - I antenna
 - I leg
 - I wing
 - I forewing
 - I hindwing



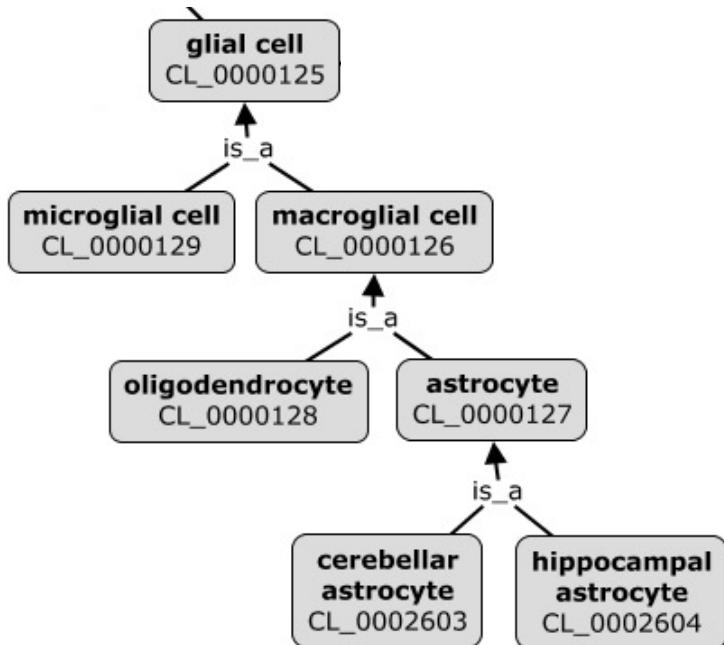
Necessary but not sufficient

Subsumption relation states that a glial cell is a type of neuron-associated cell but not all cells that develop from glioblasts are neuron-associated cells



Hierarchical Query Expansion

Genes annotated as being expressed in different types of glial cells



Glial cell 20

Macrophagel cell 15

Microglial cell 10

Query 'glial cell' without hierarchy
Results = 20

How many results do you think would be returned with the hierarchy?

Glial cell 20

└─ **Macrophagel cell** 15

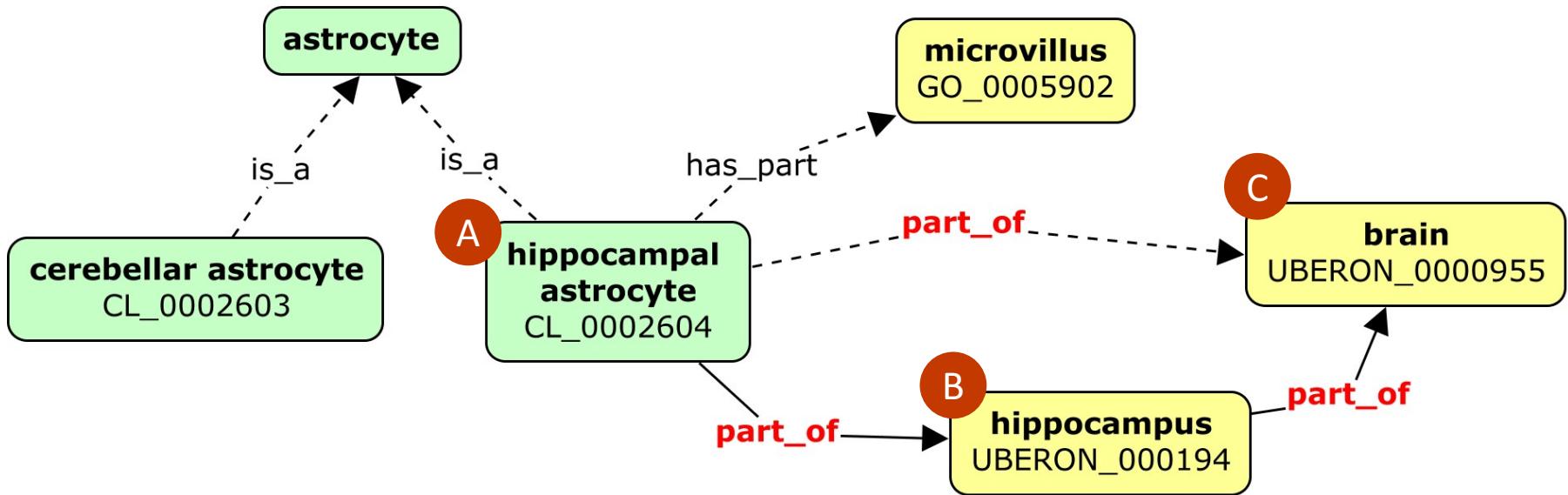
└─ **Microglial cell** 10

Query 'glial cell' with hierarchy
Results = 45

The part_of relationship

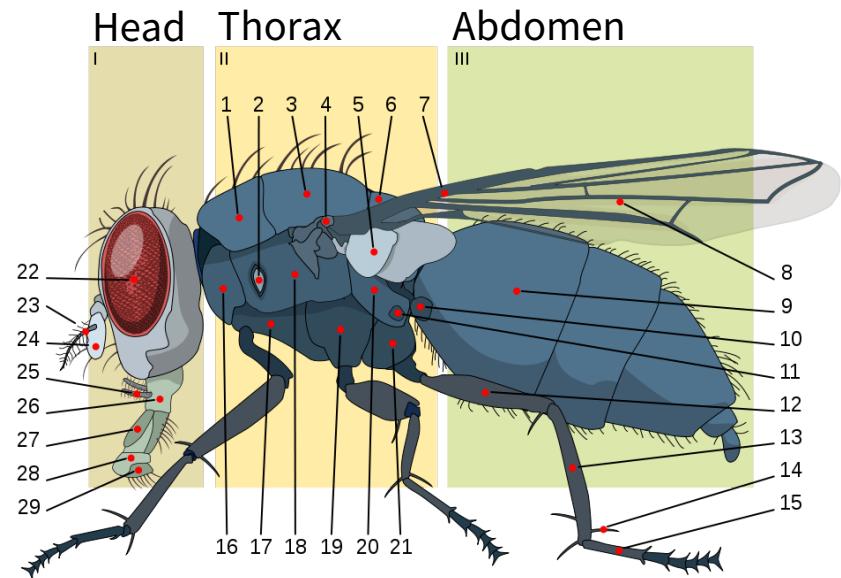
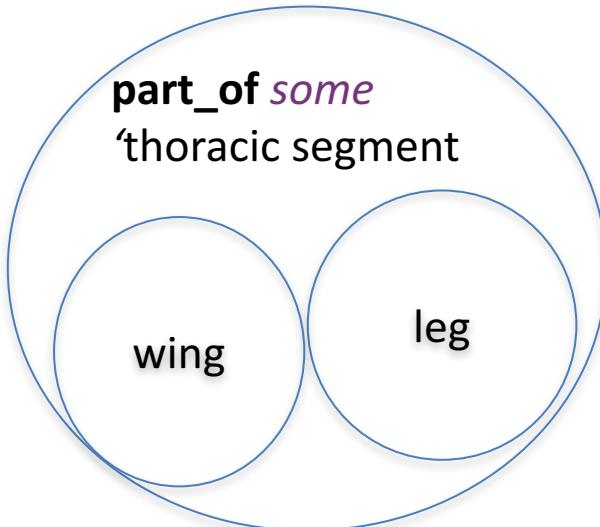
The rules of being a part:

- Nothing is a part of itself
- If A is a part of B then the B is not a part of A
- If A is a part of B and B is a part of C then A is a part of C
- The relationship is **asymmetrical** and **transitive**
- Parts do not inherit properties from the whole

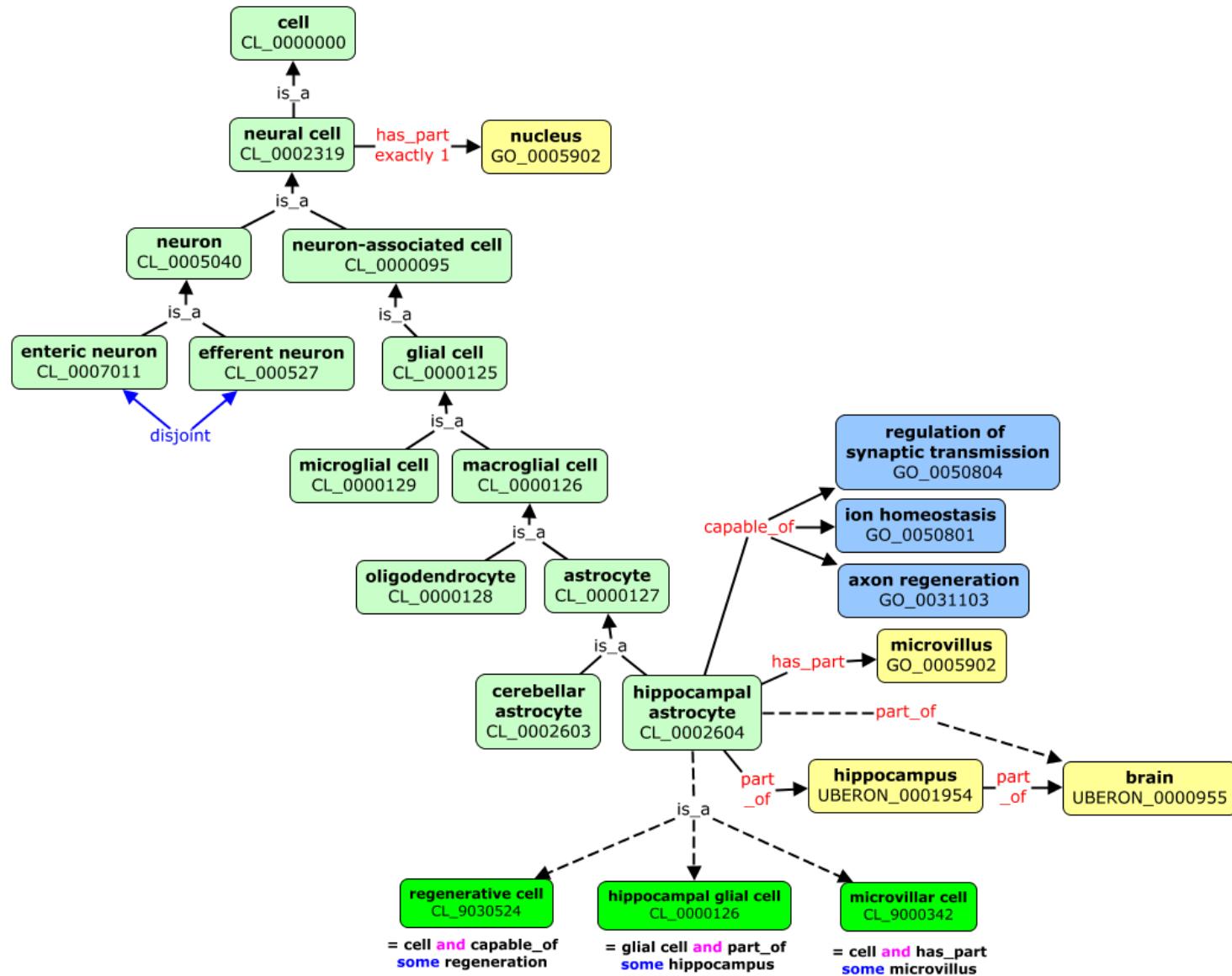


The part_of relationship – Another example

'leg' *SubClassOf* part_of *some* thoracic segment

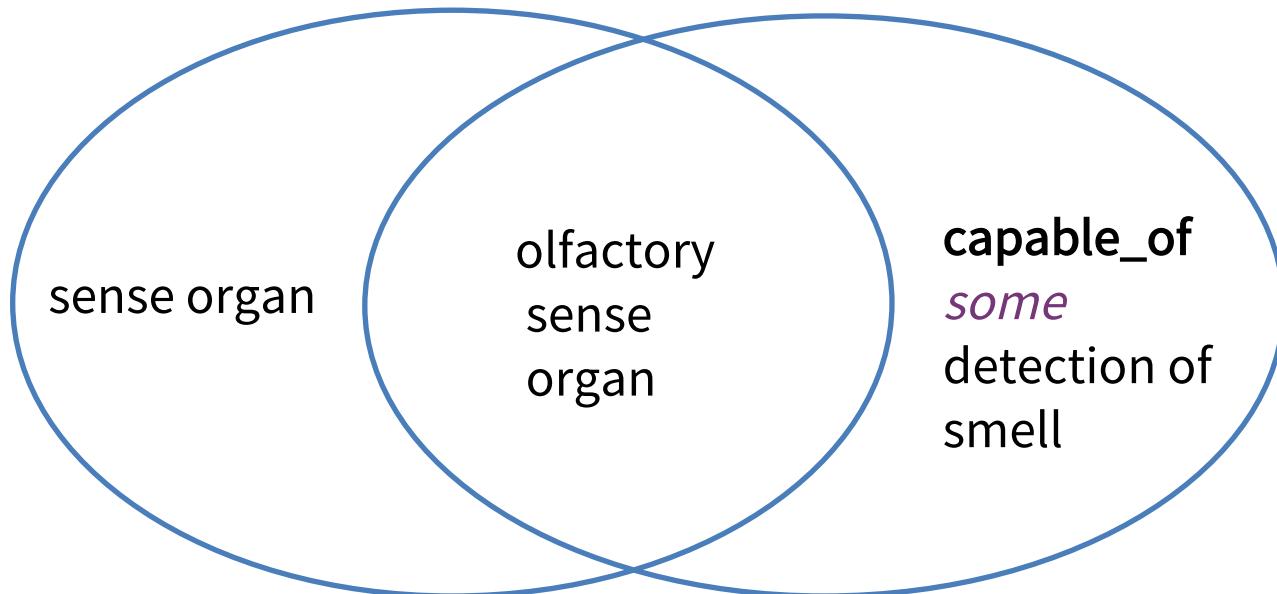


Many associative relationships between classes create network structure

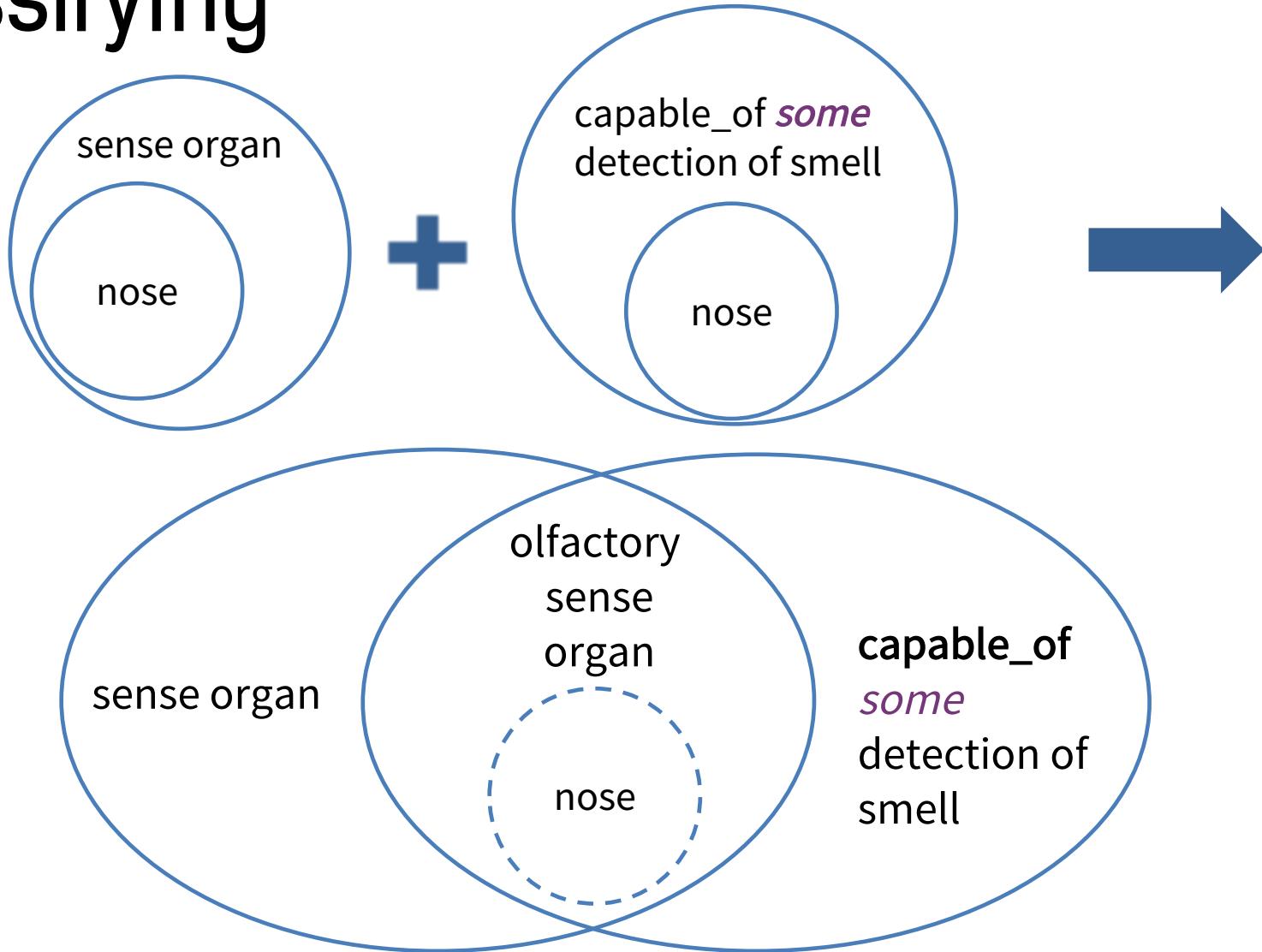


Necessary and sufficient conditions

Any sense organ that functions in the detection of smell is an olfactory sense organ

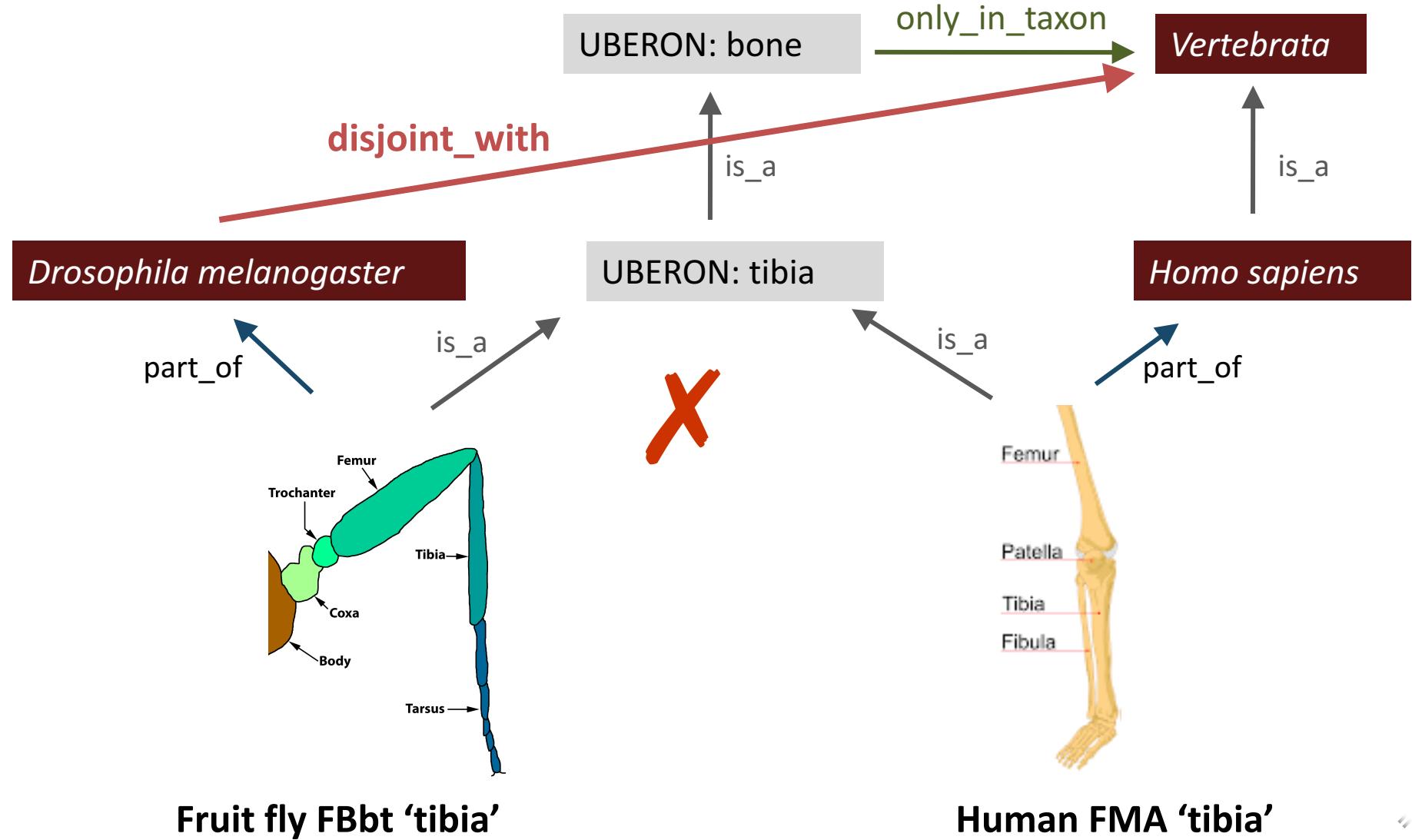


Classifying



These are necessary and sufficient conditions, also called an equivalent class axiom

Using reasoners to detect errors



The OBO Foundry

The OBO Foundry

- A collaborative community of ontology developers
- Adhere to a shared set of principles for ontology development and best practices



Open

Interoperable

Definitions

Common syntax

Logically well formed

OBO Relations

Unique identifier
space

Scientifically
accurate

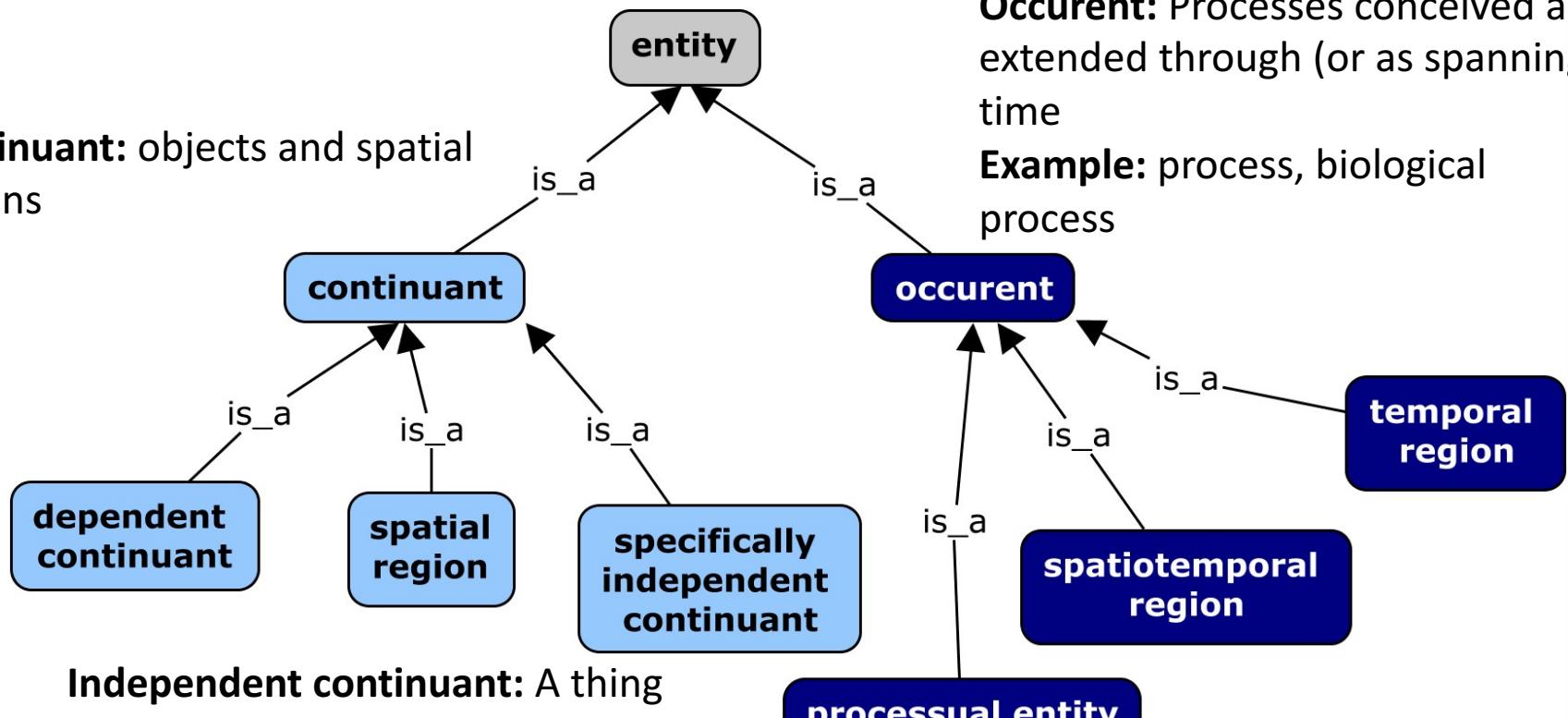
Provide
documentation

Basic Formal Ontology (BFO)

“Narrowly focused on the task of providing a genuine upper ontology which can be used in support of domain ontologies developed for scientific research, as for example in biomedicine within the framework of the [OBO Foundry](#).”

Top Level Terms in BFO

Continuant: objects and spatial regions



Independent continuant: A thing that bears qualities

Example: anatomical entity

Dependent continuant: A thing that inheres in other entities

Example: phenotype

Occurrent: Processes conceived as extended through (or as spanning) time

Example: process, biological process

?

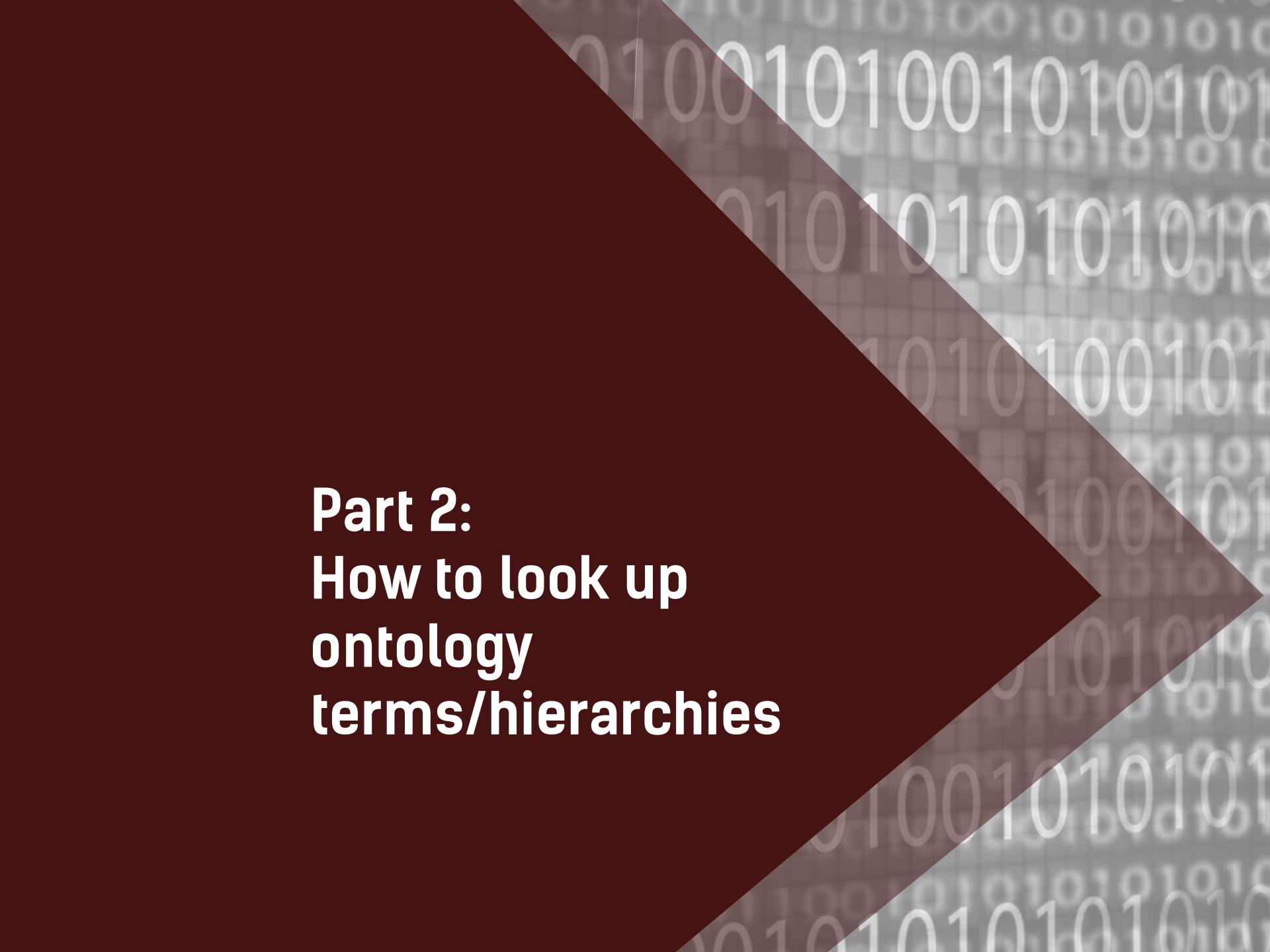
Is the class 'DNA sequencing' a continuant or occurrent?

?

What BFO class should be a superclass of 'blue eyes'?

Let's review some concepts

- What are the three main elements of an ontology?
- What is the difference between the is_a relationship and the part_of relationship?



Part 2:

How to look up ontology terms/hierarchies

The OBO Foundry

The OBO Foundry



About ▾ Principles ▾ Ontologies ▾ Participate ▾ FAQ ▾ Legacy ▾ Search Ontobee Submit

The OBO Foundry

Welcome to the new OBO website! See the Announcement for more info.

Download table as: [[YAML](#) | [JSON-LD](#) | [RDF/Turtle](#)]

chebi	Chemical Entities of Biological Interest	A structured classification of molecular entities of biological interest focusing on 'small' chemical compounds. Detail							
doid	Human Disease Ontology	An ontology for describing the classification of human diseases organized by etiology. Detail							
(cc) BY	Gene Ontology	An ontology for describing the function of genes and gene products Detail							
obi	Ontology for Biomedical Investigations	An integrated ontology for the description of life-science and clinical investigations Detail							
(cc) BY	Phenotypic quality	An ontology of phenotypic qualities (properties, attributes or characteristics) Detail							

More than just a website, it's a community of ontology developers dedicated to working together using common principles

<http://www.obofoundry.org/>

National Center for Biomedical Ontologies BioPortal



Welcome to BioPortal, the world's most comprehensive repository of biomedical ontologies

Search for a class

🔍

[Advanced Search](#)



Find an ontology

🔍

[Browse Ontologies](#)

BioPortal Statistics

Ontologies	729
Classes	9,594,071
Resources Indexed	48
Indexed Records	39,537,360
Direct Annotations	95,468,433,792

A library of ontologies and ontology related services

<http://bioportal.bioontology.org/>

Ontology Lookup Service

The screenshot shows the OLS homepage with a dark header bar containing the EMBL-EBI logo and navigation links for Services, Research, Training, and About us. Below the header is a light blue banner with the OLS logo and the text "Ontology Lookup Service". The main content area has a light gray background. It features a search bar with placeholder text "Search OLS..." and a magnifying glass icon. Below the search bar is an example query "Examples: diabetes, GO:0098743". To the right of the search bar is a link "Looking for a particular ontology?". On the left side, there's a section titled "About OLS" with a brief description of the service and its development team. In the center, there's a "Related Tools" section listing OxO, Zooma, and Webulous services. On the right, there's a "Contact Us" section with information about feedback, GitHub issues, and mailing lists, along with a Twitter feed from @EBIOLS.

EMBL-EBI

Services | Research | Training | About us

Ontology Lookup Service

Home | Ontologies | Documentation | About | Contact Us

Welcome to the EMBL-EBI Ontology Lookup Service.

Search OLS...

Examples: diabetes, GO:0098743

Looking for a particular ontology?

About OLS

The Ontology Lookup Service (OLS) is a repository for biomedical ontologies that aims to provide a single point of access to the latest ontology versions. You can browse the ontologies through the website as well as programmatically via the OLS API. OLS is developed and maintained by the [Samples, Phenotypes and Ontologies Team \(SPOT\)](#) at EMBL-EBI.

Related Tools

In addition to OLS the SPOT team also provides the OxO, Zooma and Webulous services. [OxO](#) provides cross-ontology mappings between terms from different ontologies. [Zooma](#) is a service to assist in mapping data to ontologies in OLS and [Webulous](#) is a tool for building ontologies from spreadsheets.

Contact Us

For feedback, enquiries or suggestion about OLS or to request a new ontology please contact [ols-support @ ebi.ac.uk](mailto:ols-support@ebi.ac.uk). For bugs or problems with the code or API please report on [GitHub issue](#) For announcements relating to OLS, such as new releases and new features sign up to the [OLS announce mailing list](#)

EBISPORT OLS Retweeted

Simon Jupp @simonjupp
We're looking to hire a senior developer to join us at @emblebi working on @EBIOLS and our full stack of #ontology services. DM me for more info embl.de/jobs/searchjob...

Aug 13, 2018

<https://www.ebi.ac.uk/ols/index>



Welcome to Ontobee!

Ontobee: A [linked data](#) server designed for ontologies. Ontobee is aimed to facilitate ontology data sharing, visualization, query, integration, and analysis. Ontobee dynamically [dereferences](#) and presents individual ontology term URIs to (i) *HTML web pages* for user-friendly web browsing and navigation, and to (ii) *RDF source code* for [Semantic Web](#) applications. Ontobee is the default linked data server for most [OBO Foundry library ontologies](#). Ontobee has also been used for many non-OBO ontologies.

Please select an ontology (optional)

Keywords: Search terms Batch Search

Jump to <http://purl.obolibrary.org/obo/> Go

Currently Ontobee has been applied for the following ontologies:

No.	Ontology Prefix	Ontology Full Name	ODO	List of Terms
1	AEO	Anatomical Entity Ontology	L	
2	AGRO	Agronomy Ontology	L	
3	APO	Ascomycete phenotype ontology	L	
4	APOLLO_SV	Apollo Structured Vocabulary	N	
5	ARO	Antibiotic Resistance Ontology	L	
6	BCGO	Beta Cell Genomics Ontology	L	
7	BCO	Biological Collections Ontology	L	

Query OBO ontologies, and provides RDF supporting remote query of each ontology term and the Semantic Web

<http://www.ontobee.org/>

AmiGO: Gene Ontology Viewer



AmiGO 2

Home

Search ▾

Browse

Tools & Resources

Help

Feedback

About

AmiGO 1.8

AmiGO 2

More information on quick search

Quick search

Search

Search Templates



Use predefined **templates** to explore Gene Ontology data.

[Go »](#)

Advanced Search



Interactively **search** the Gene Ontology data for annotations, gene products, and terms using a powerful search syntax and filters.

[Search ▾](#)

Browse the Ontology



Use the drill-down **browser** to view the ontology structure with annotation counts.

[Go »](#)

GOOSE



Use **GOOSE** to query the legacy GO database with **SQL**.

[Go »](#)

Term Enrichment Service



Your genes here...

Statistics



View the most recent **statistics** about the Gene Ontology data in AmiGO.

[Go »](#)

And Much More...

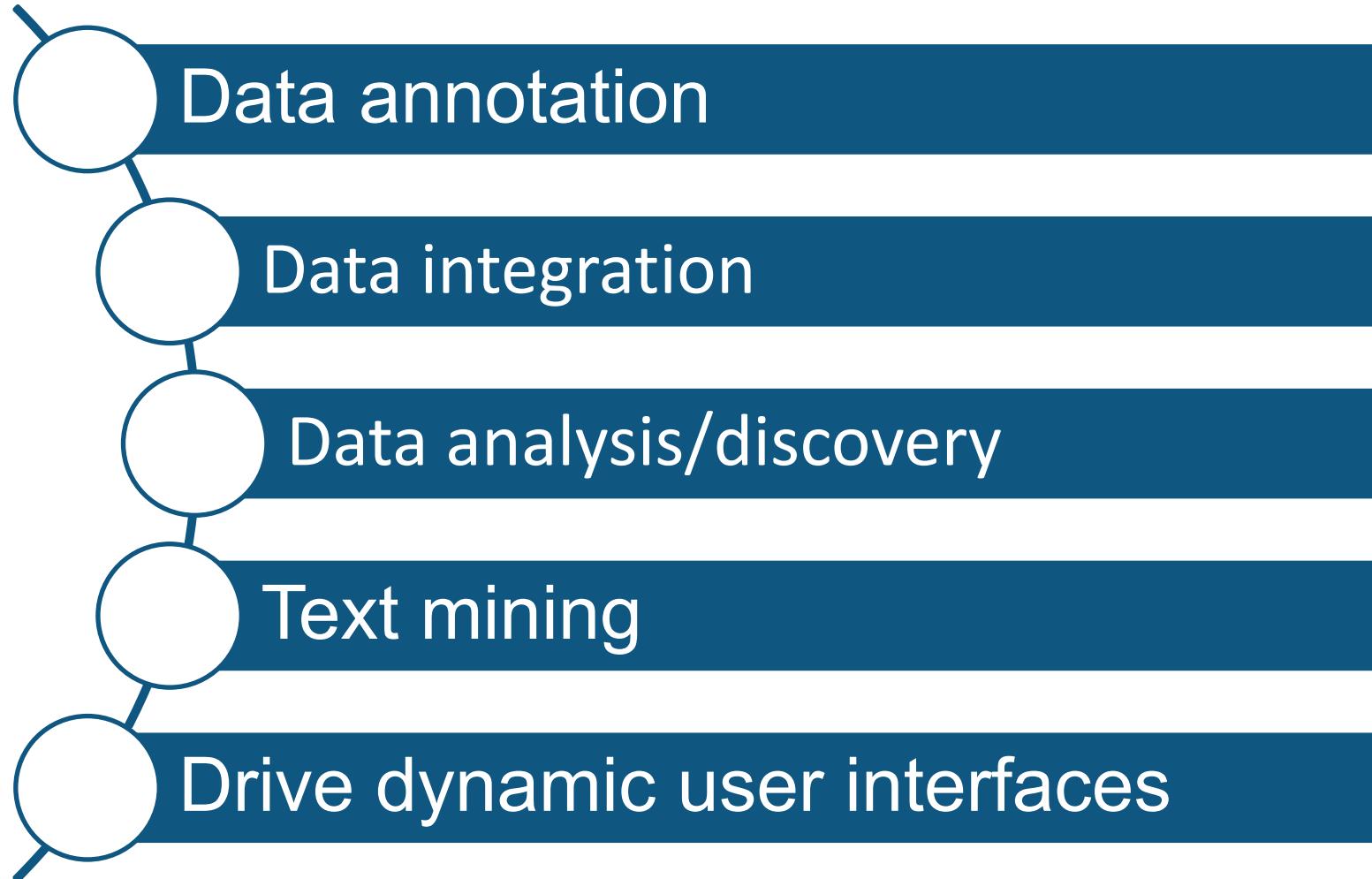


Many **more tools** are available from the software list, such as alternate searching modes, Visualize, non-JavaScript pages.

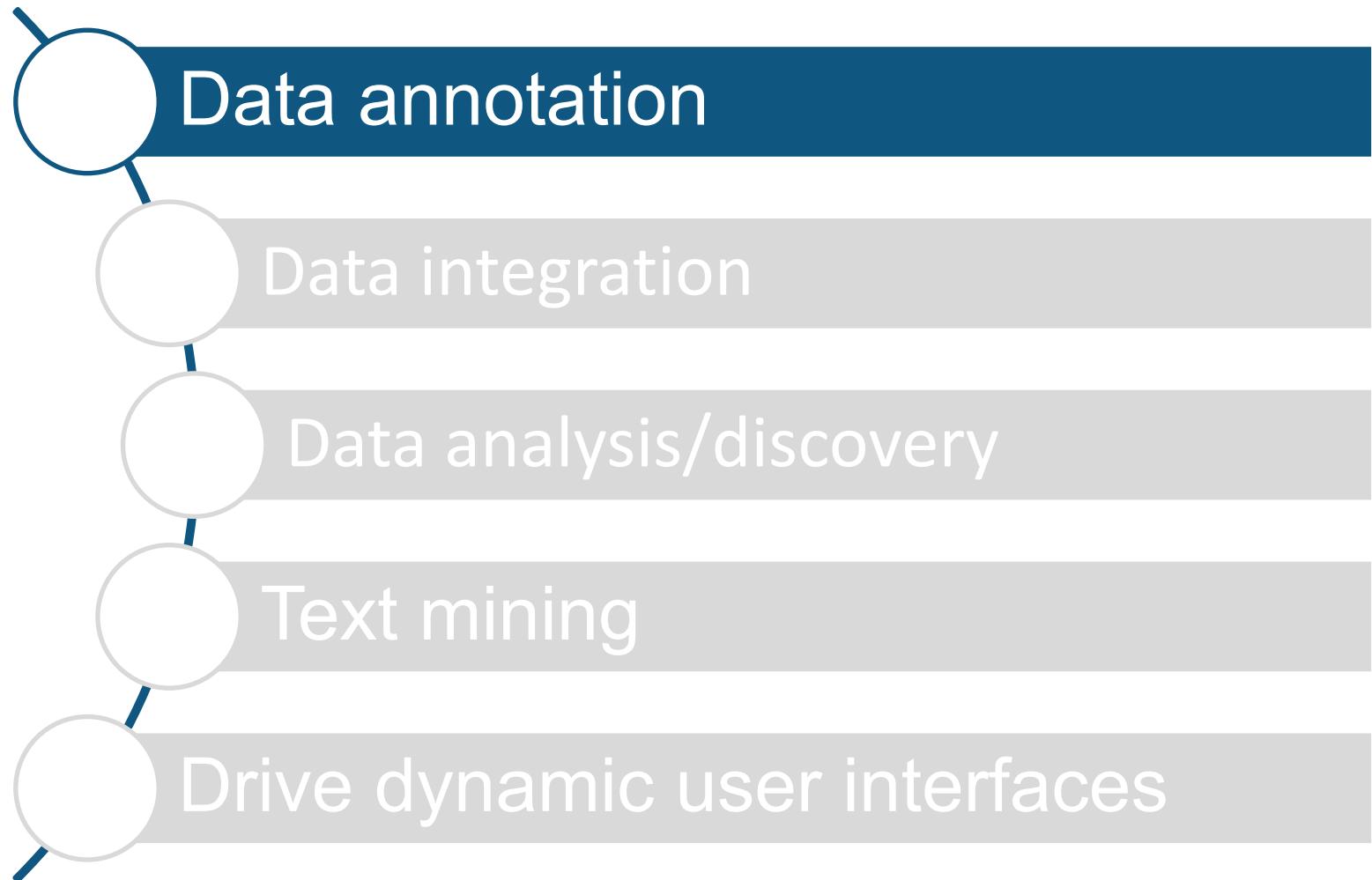
[Go »](#)

Part 3: Applications of Ontologies/ Use Cases

Applications of ontologies



Applications of ontologies



Gene Ontology

The Problem

A lot of sequence databases. No common vocabulary to talk about genes. No way to search sequences databases other than by similarity.

The Solution

Created in 1998



GENEONTOLOGY
Unifying Biology

Describe function, process and location, by creating an ontology and unifying vocabulary to provide structure to aid annotation and querying.



Provides computable knowledge regarding the functions of genes and gene products

Gene Ontology

Biological processes

apoptosis

Cellular components

organelle

Molecular functions

ligase activity

GO annotations

Evidence-based statements relating a specific gene or gene product to a specific ontology term

QuickGO at the EBI: Gene Ontology Viewer



Gene Ontology and GO Annotations

Help

Contact

API

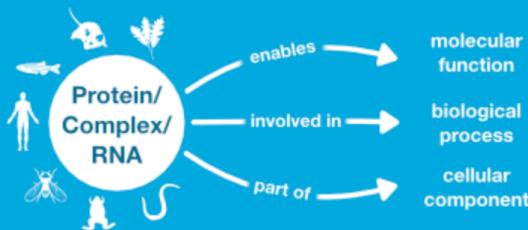
Basket

Search

Search

e.g apoptosis; GO:0006915; ECO:0000314; tropomyosin

View GO Annotations



Explore biology

Use sets of GO terms (**slims**) that describe your area of interest



<https://www.ebi.ac.uk/QuickGO/>

Biocuration (noun)

Biocuration is the extraction of knowledge from unstructured biological data into a structured, computable form.

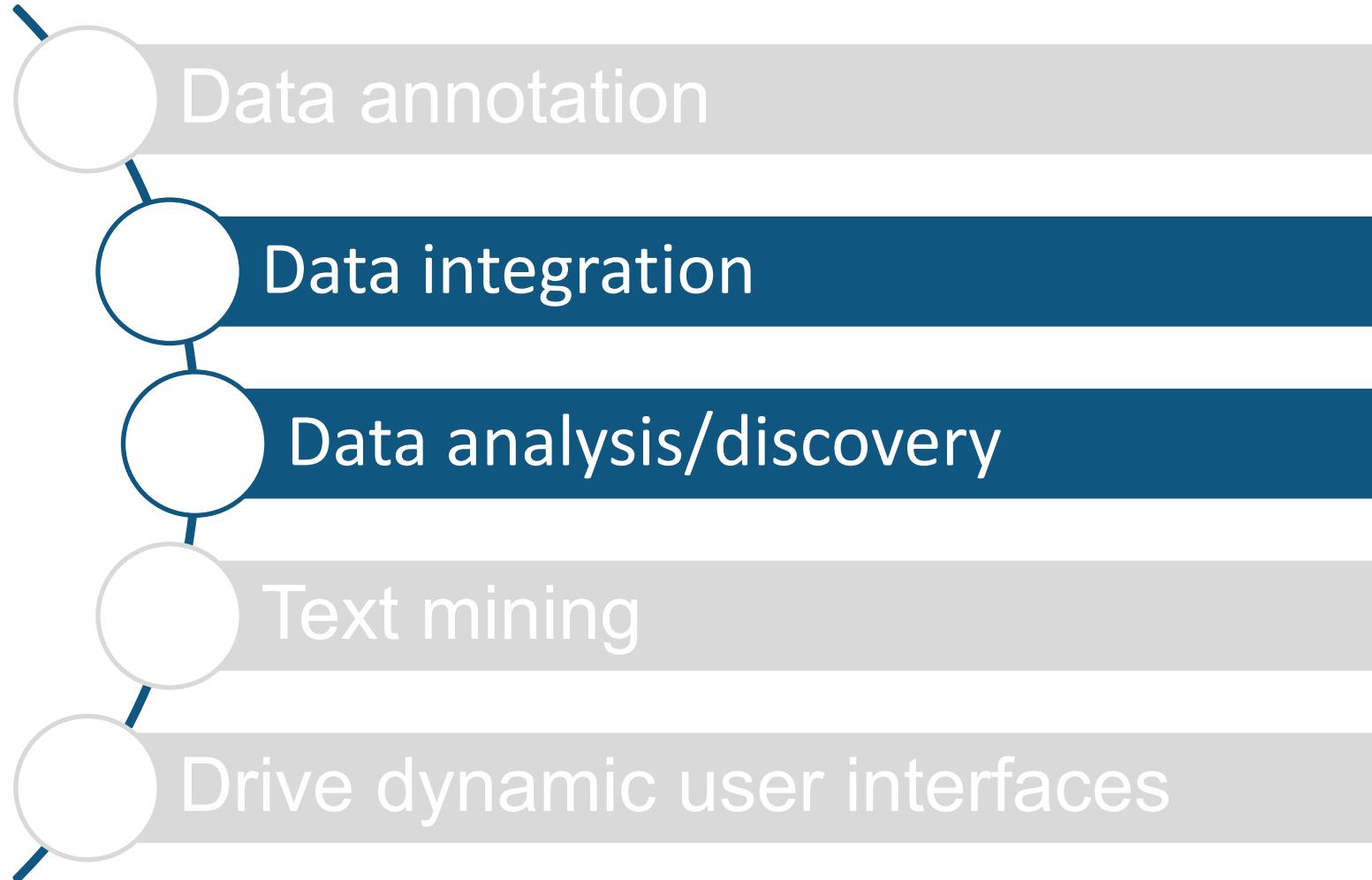


bit.ly/biocurationpaper

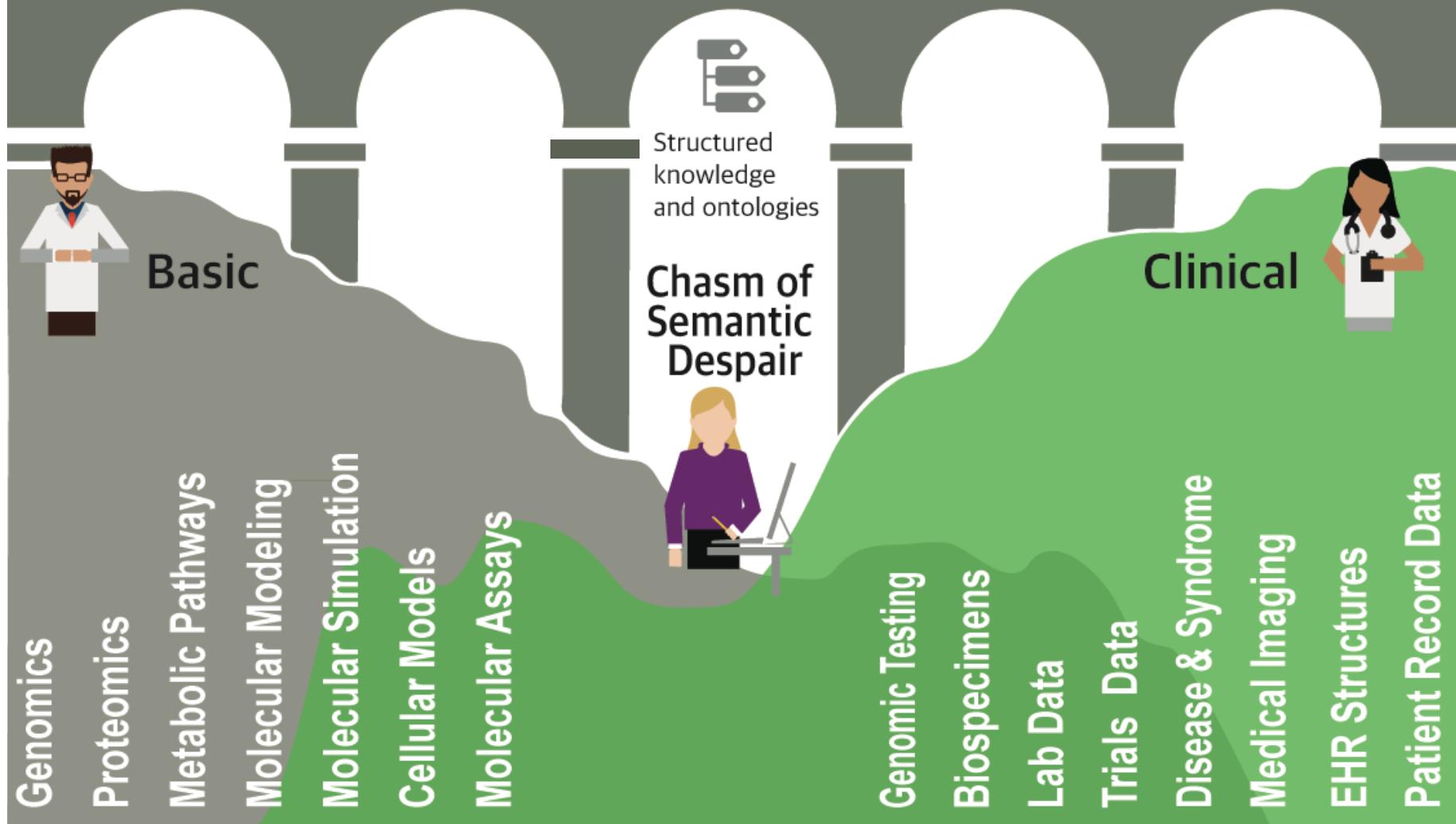
Model Organism Databases (MODs)



Applications of ontologies

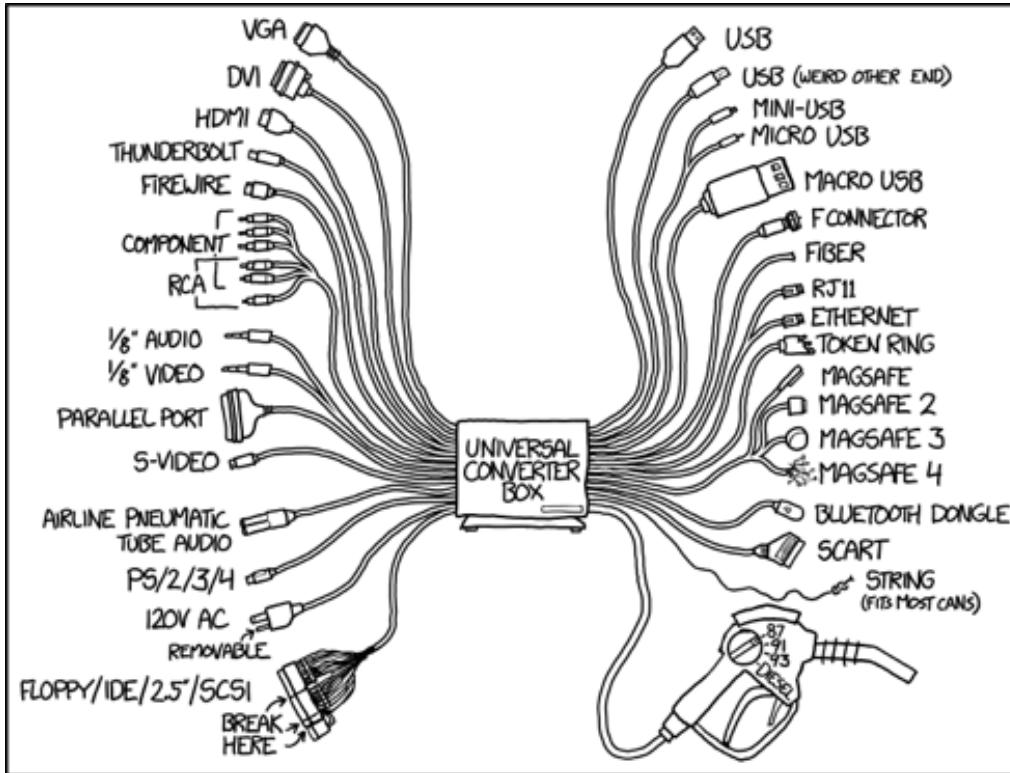


Translational Bridge



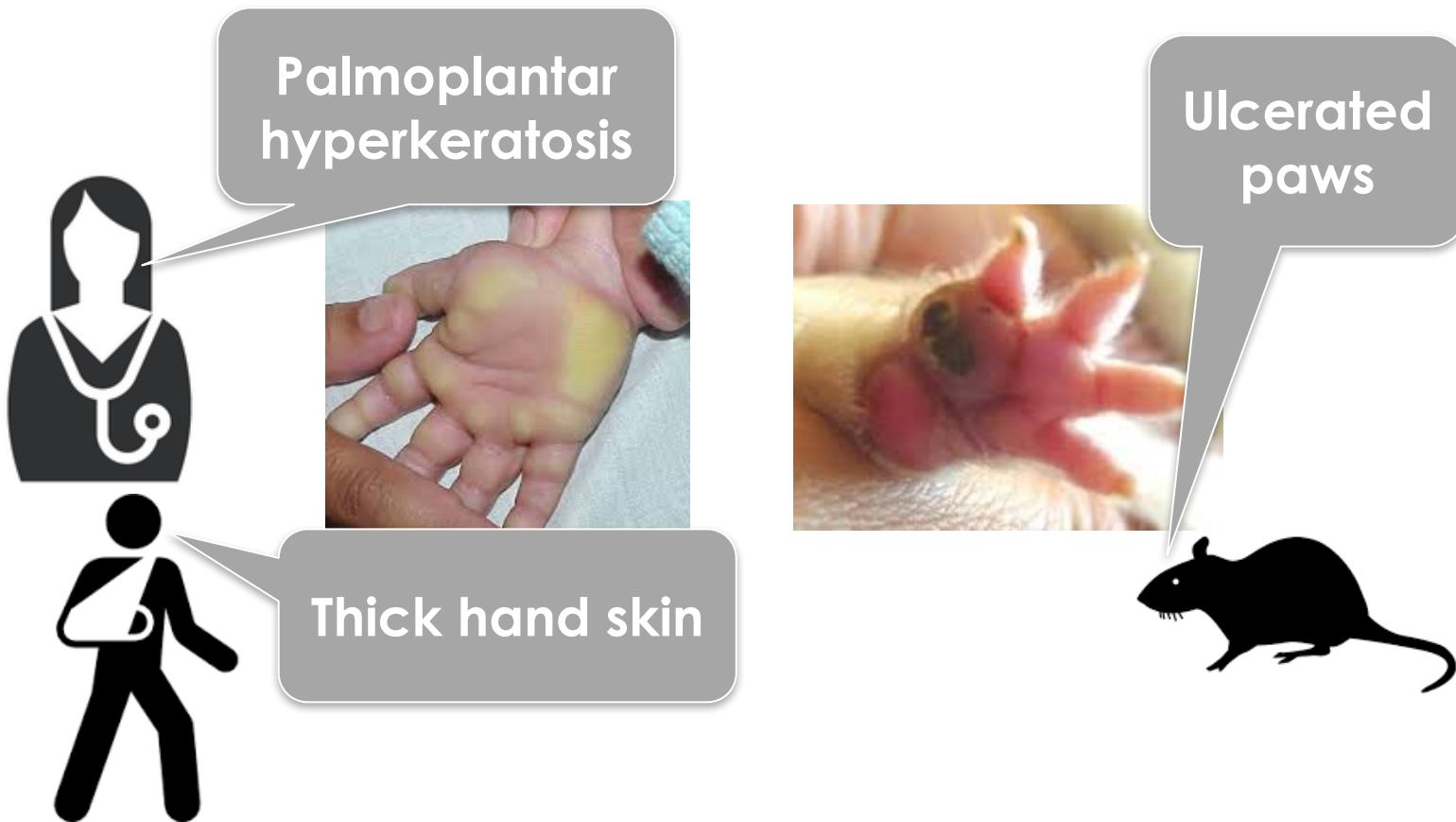
Goal: Align and structure disparate data types to make them more interoperable, discoverable and reusable by building ontologies to structure these data

Ontologies can serve as a bridge for data harmonization

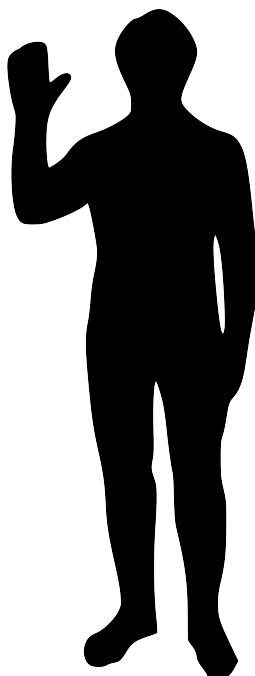


<https://xkcd.com/1406/>

Use case: Phenotypes (observable characteristics) are described differently by different groups

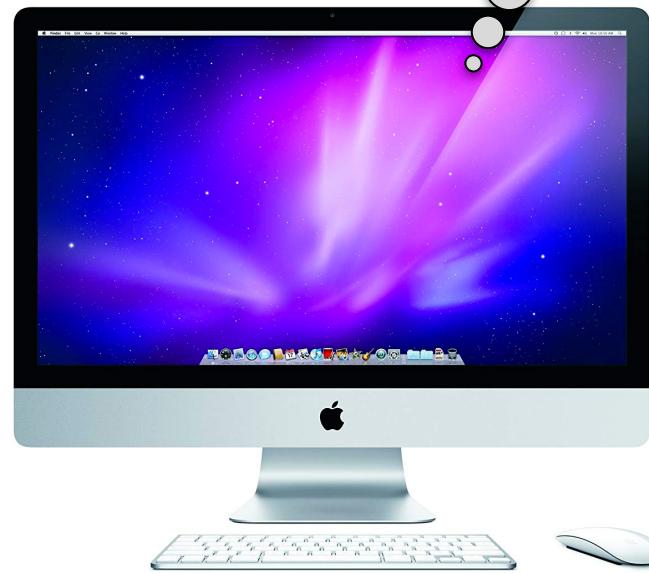


How can we help computers understand phenotypic features?



Human phenotypic feature

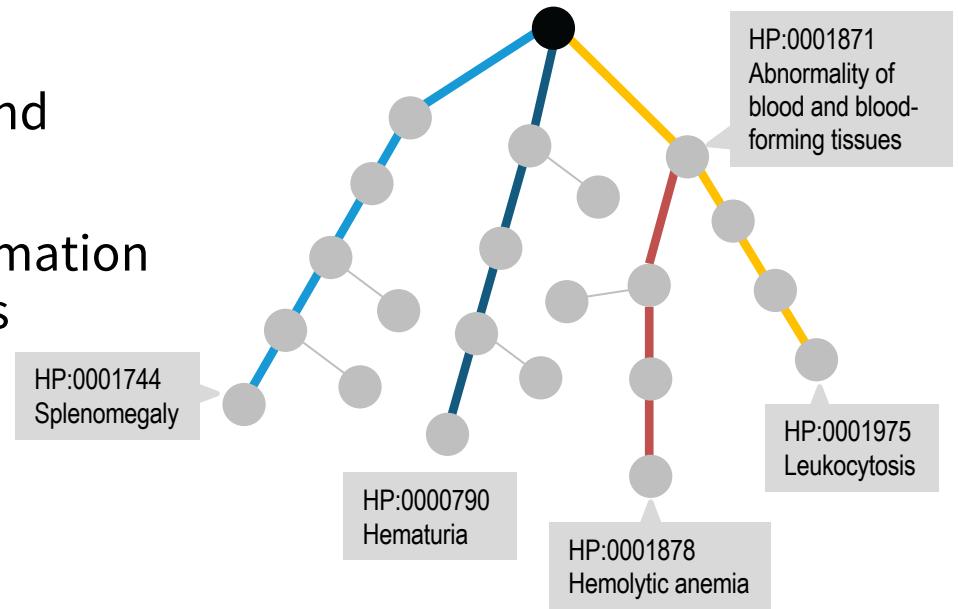
“Palmoplantar hyperkeratosis”



Human Phenotype Ontology (HPO)



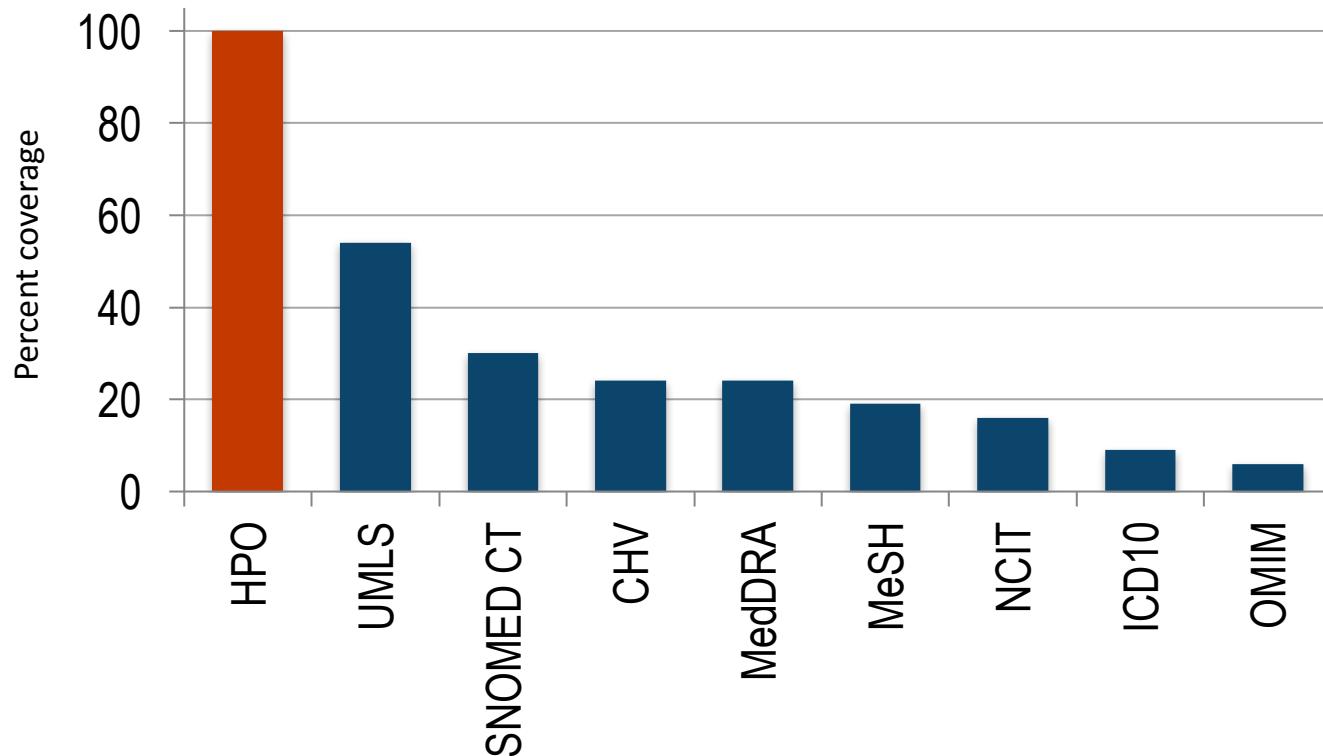
- A standardized, machine readable vocabulary of phenotypic abnormalities encountered in human disease
- Created in 2008 by Peter Robinson and Sebastian Köhler at Charité, Berlin
- Goal is to integrate phenotypic information across scientific fields and databases
- Developed using:
 - Medical literature
 - Orphanet
 - DECIPHER
 - OMIM
 - User requests/contributions





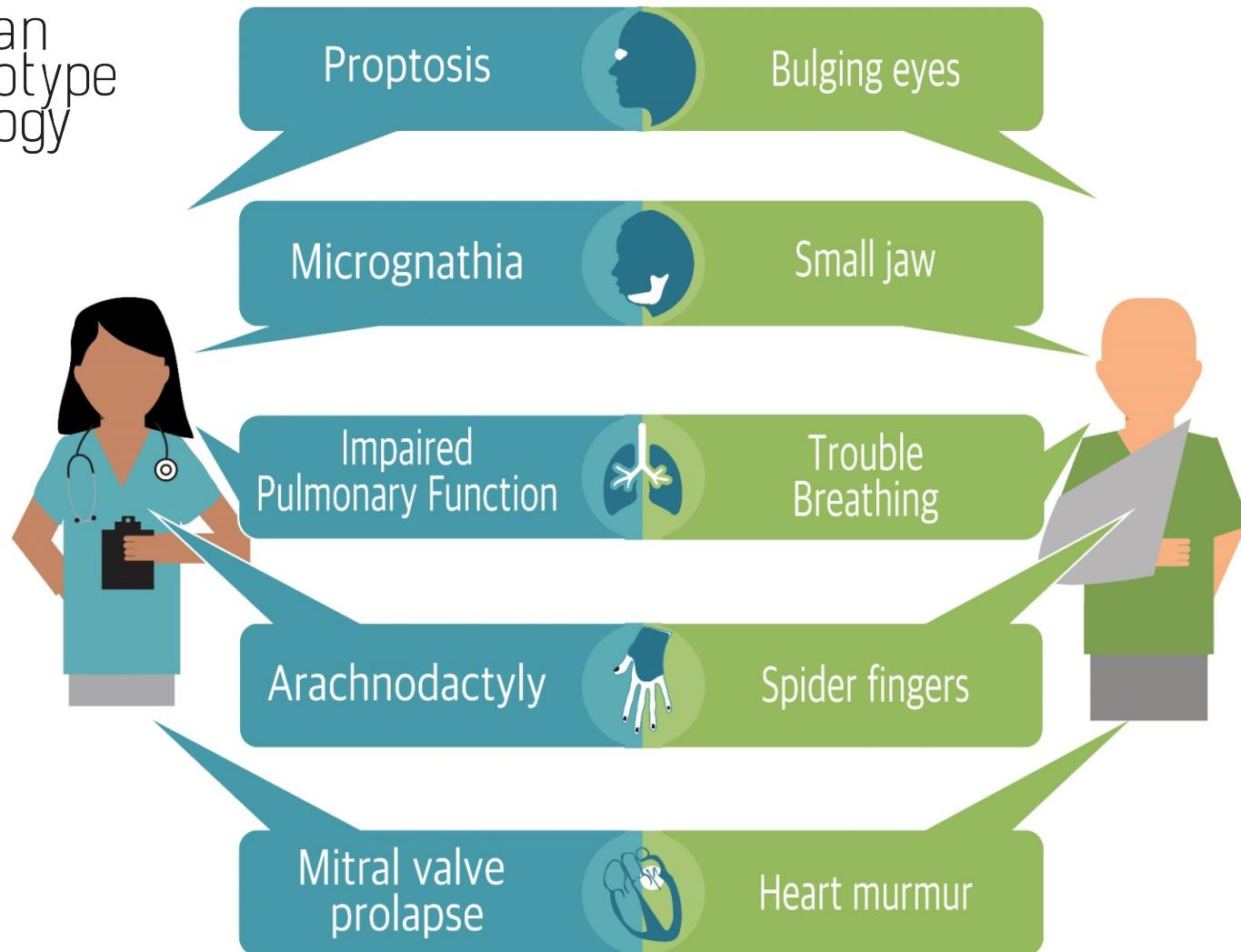
Why the HPO?

Other clinical vocabularies don't adequately cover phenotypic descriptions



=> HPO is now in the UMLS

HPO Layperson Terminology





Patient self-phenotyping tools driven by layperson HPO

PHENOTYPR

CONTACT US

P H E N O T Y P R

Patient Provided Phenotyping

[Let's get started](#)

WHAT IS PHENOTYPR?

Phenotypr is an educational tool for parents and families who believe they may have a child with a genetic intellectual disorder. Using

WHAT IS HPO?

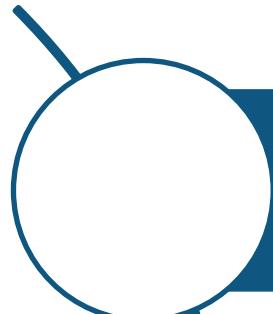
The Human Phenotype Ontology (HPO) aims to provide a standardized vocabulary of phenotypic abnormalities

WHY USE HPO?

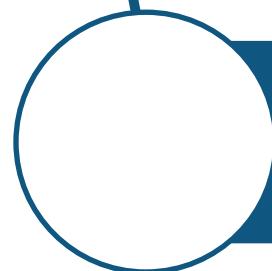
Knowing the HPO mapping of your child can provide your clinicians access to quality, structured phenotyping data. Providing this



Patient driven self-phenotyping by layperson HPO



4,555 terms in the HPO are annotated with at least one lay person synonym (35.4% coverage)



7,607 number of lay person synonyms total



60% of all disease annotations (73,932 of 122,120) are referring to HPO terms with lay translations

HPO Language Translations

Translation of labels, synonyms, and text definitions



Dutch

1%



Portuguese

1%



Russian

12%



German

13%



Italian

13%



Spanish

16%



French

20%



Japanese

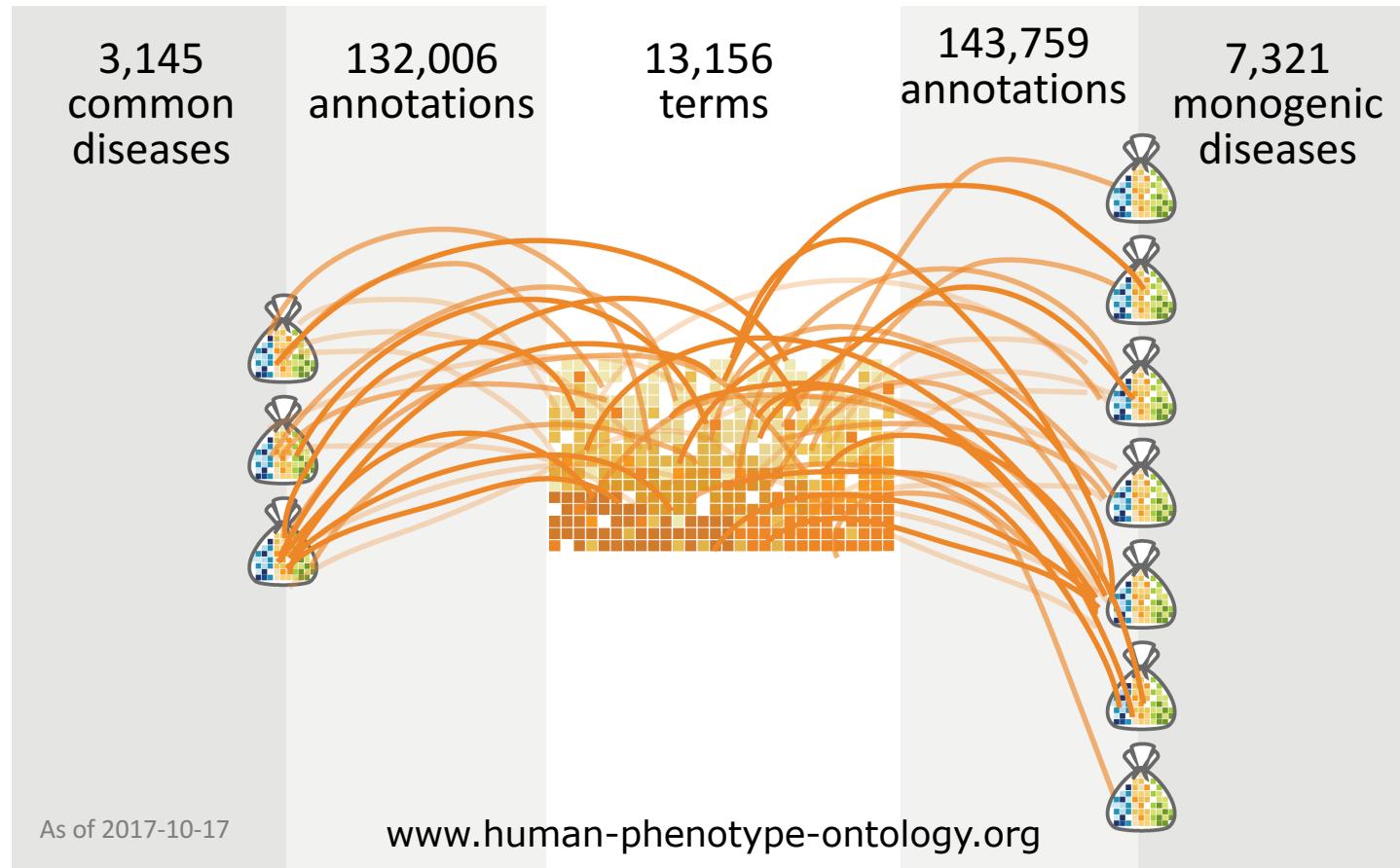
100%



Chinese

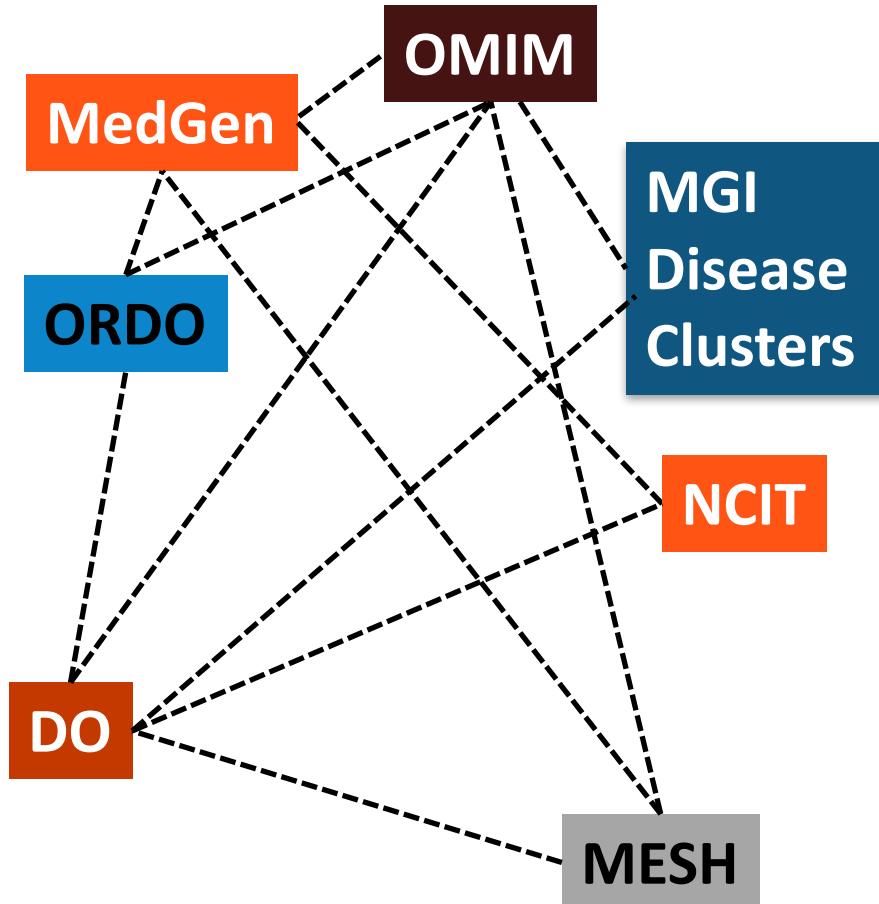
near 100%

HPO and Disease Annotations



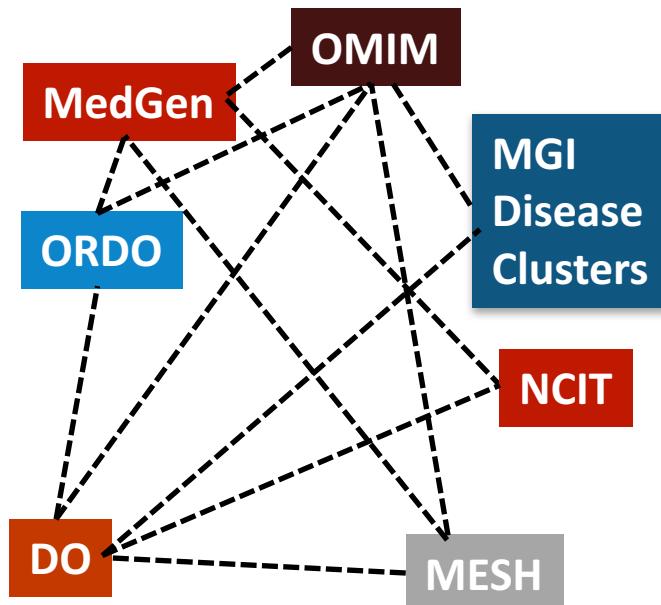
A number of clinical databases use the HPO to annotate patient phenotype data, to help inform disease-gene associations

Disease Classification



- Numerous nosologies exist
- Biocurators use different ontologies to annotate datasets
- Combining ontologies is difficult because it is not clear which diseases are functionally equivalent
- Attempts to align with each other

Monarch Merged Disease Ontology (MONDO)



Developed by semi-automatic methods, using kBOOM algorithm
<http://bit.ly/k-BOOM>

MONDO is available for disease annotations
Ontology:
<http://purl.obolibrary.org/obo/mondo.owl>



U.S. Department of Health & Human Services



National Institutes of Health



NCATS



National Center
for Advancing
Translational Sciences

GARD

Genetic and Rare Diseases
Information Center

1-888-205-2311

Diseases

Guides

News

About GARD

En Español



Search for Diseases, Organizations, News and More...

GO



HOME > DISEASES > SICKLE CELL ANEMIA



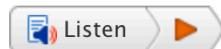
Sickle cell anemia

Other Names: HbS disease; Hemoglobin S Disease; Sickling disorder due to hemoglobin S

Categories: [Blood Diseases](#); [Congenital and Genetic Diseases](#); [Endocrine Diseases](#); [See More](#)

<https://rarediseases.info.nih.gov/diseases/8614/sickle-cell-anemia>

Symptoms



This table lists symptoms that people with this disease may have. For most diseases, symptoms will vary from person to person. People with the same disease may not have all the symptoms listed. This information comes from a database called the [Human Phenotype Ontology \(HPO\)](#). The HPO collects information on symptoms that have been described in medical resources. The HPO is updated regularly. Use the HPO ID to access more in-depth information about a symptom.

Showing 1-5 of 37 | [View All](#)

Medical Terms	Other Names	Learn More: HPO ID
100% of people have these symptoms		
Chronic hemolytic anemia		0004870
80%-99% of people have these symptoms		
Recurrent infections	Frequent infections [more ▾]	0002719
30%-79% of people have these symptoms		
Abnormality of the spleen		0001743
Aseptic necrosis		0010885
Chest pain		0100749

Showing 1-5 of 37 | [View All](#)

NCI Thesaurus (NCIt)

- Developed by the National Cancer Institute (NCI)
- Ontological representation of
 - clinical care
 - basic research
 - translational research
 - public and administrative activities
- OWL2 ontology with anatomical axioms that are linked to disease concepts
- Contains over 100,000 definitions

Abnormal cell

Anatomic Structure,
System or Substance

Biological process

Disease, Disorder
or Finding

↳ Disease or Disorder

↳ Neoplasm

Gene

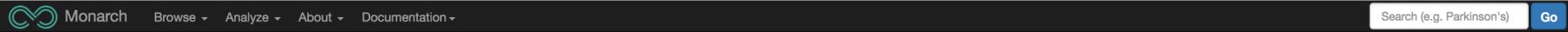
Organism

Monarch Initiative



Monarch uses semantic technologies to aggregate data to support disease diagnostics, where use of large scale integrated data can inform disease diagnosis and decisions

Monarch Initiative



Monarch Phenotype Profile Analysis

The Phenotype Analysis Tool enables you search our database using the [OwlSim](#) Semantic Similarity analysis engine to find phenotypically similar diseases or genes in a variety of organisms, then visualize their overlap. ([Read more](#) if you want to understand how this works.)

Instructions: [Show/Hide](#)

Step 1: Create A Profile of Phenotypes

Start typing a phenotype label (eg. Resting tremor); select from menu. Repeat as desired to build up a profile.

Empty: Add items using the input above.

Step 2: Select Comparables

Choose From Predefined Target Groups

Enter A Specific Set of Genes

Target Groups:

All genes and diseases

Analyze

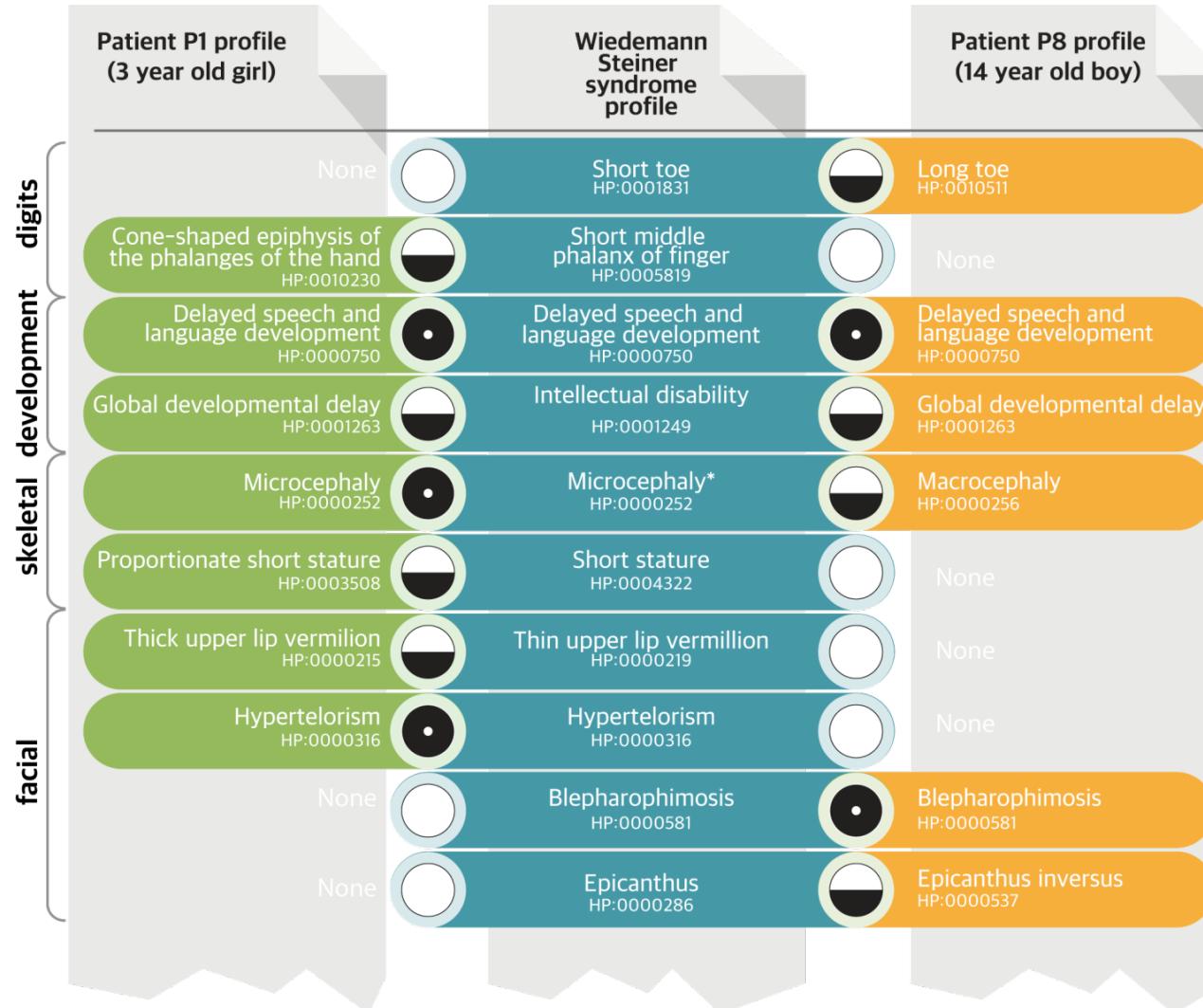
The Monarch Initiative is a collaboration between:

Oregon Health & Science University
Lawrence Berkeley National Laboratory
Garvan Institute of Medical Research
Barts & The London School of Medicine & Dentistry, Queen Mary University of London

Monarch is supported generously by a NIH Office of the Director Grant #5R24OD011883, as well as by NIH-UDP: HHSN268201300036C, HHSN268201400093P, NCI/Leidos #15X143. We are grateful to the many [original sources of our data](#) for allowing Monarch to integrate them in this way. Except where forbidden by the original sources, this work is licensed under a Creative Commons Attribution 3.0 License.

The Jackson Laboratory
Charité - Universitätsmedizin Berlin
William Harvey Research Institute
Renaissance Computing Institute, University of North Carolina at Chapel Hill

Differential diagnosis with similar but non-matching phenotypes is difficult



- Two patients with mutation in *KMT2A*
- Not same variant
- Phenotype similarity metric helped inform diagnosis of same disease



Legend



Perfect Match



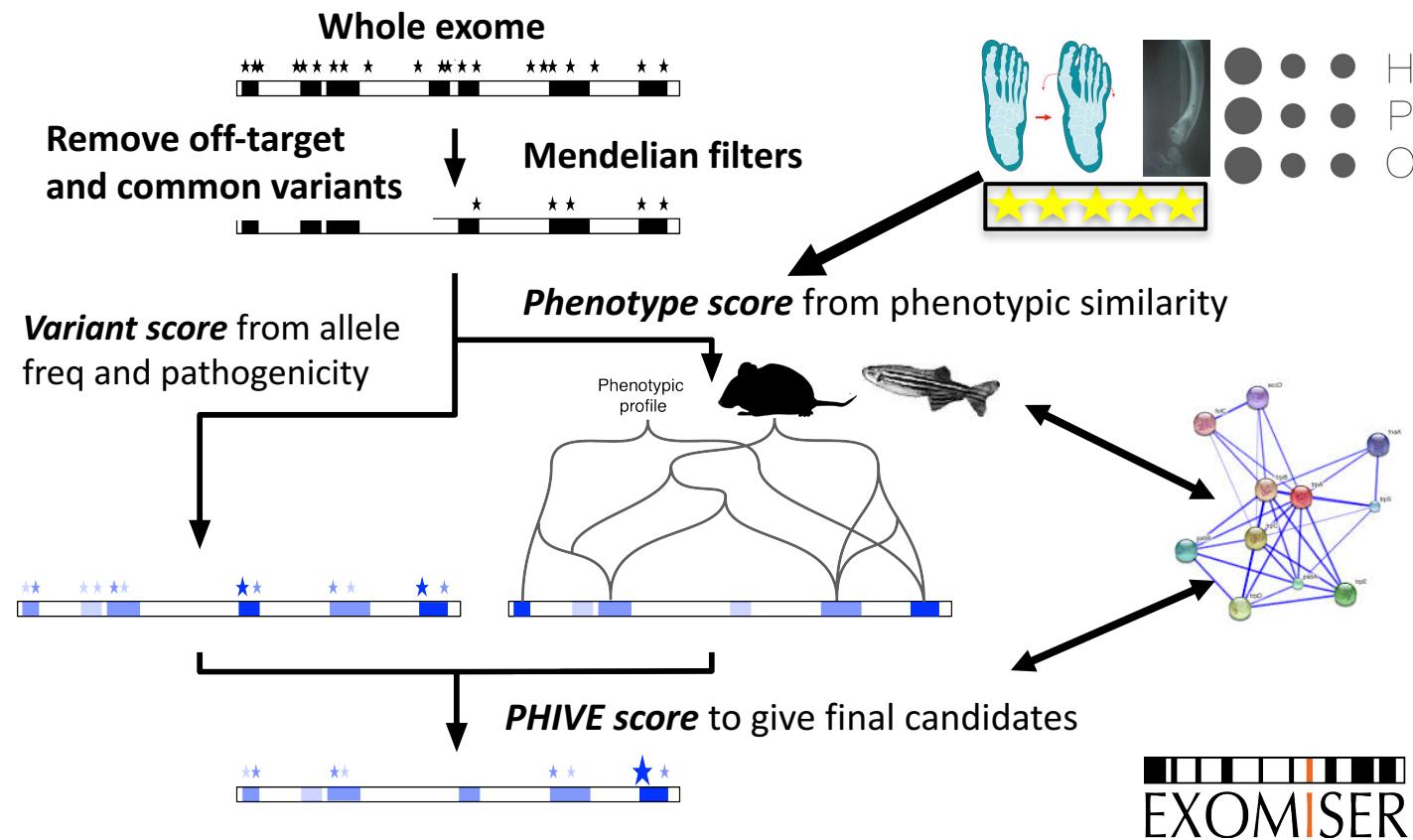
Fuzzy Match



No Match

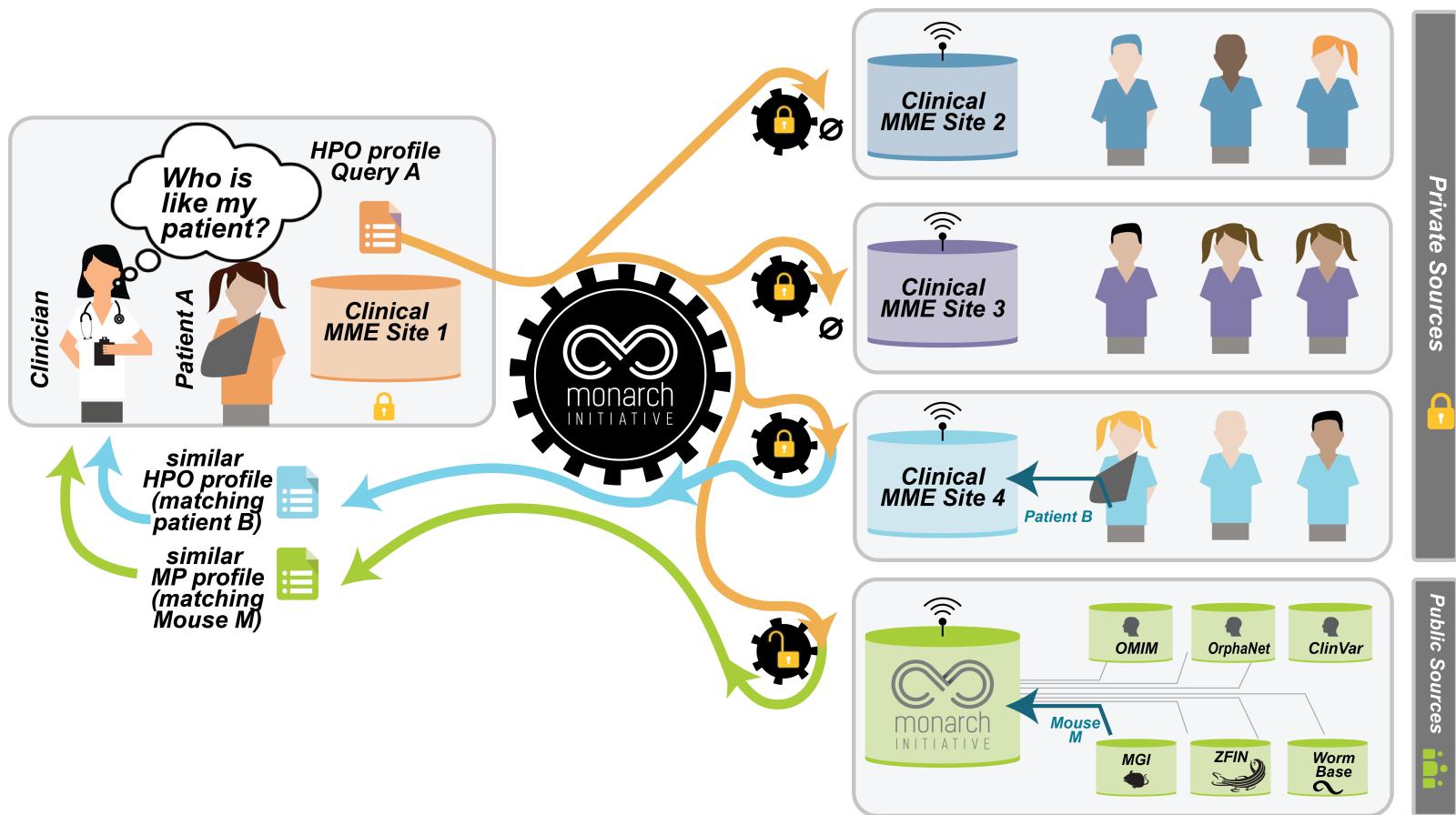
DOI: 10.1126/scitranslmed.3009262

Combining G2P data for variant prioritization



- HPO-based phenotypic profile matching used in combination with genome sequencing has allowed prioritization of candidate genes with predicted pathogenic variants.
- This approach helped establish diagnoses in 28% of children in the UDP/N

Matchmaker Exchange for patients, diseases, and model organisms to aid diagnosis and mechanistic discovery



Ontologists are part of team science efforts

Contributors and expertise needed for a genetic diagnosis

Clinical/care



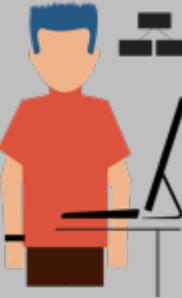
Thomas Markello
Dong Chen
Justin Y. Kwan
Iren Horkayne-Szakaly
Alan Morrison
Olga Simakova
Irina Maric
Jay Lozier
Andrew R. Cullinane
Tatjana Kilo
Lynn Meister
Kourosh Pakzad
Sanjay Chainani
Roxanne Fischer
Camilo Toro
James G. White
David Adams
Cornelius Boerkoel
William A. Gahl
Cynthia J. Tifft
Meral Gunay-Aygun

Pathology



Hans Goeble
Karen Balbach
Nadine Pfeifer
Sandra Werner
Christian Linden

Ontologist



Melissa Haendel
Peter Robinson
Chris Mungall
Sebastian Kohler
Cindy Smith
Nicole Vasilevsky
Sandra Dolken

CS/informatics



Elizabeth Lee
Amanda Links
Will Bone
Murat Sincan
Damian Smedley
Jules Jacobson
Nicole Washington
Elise Flynn
Sebastian Kohler
Orion Buske
Marta Girdea
Michael Brudno
Jeremy Band

Curator



Melissa Haendel
David Adams
David Draper
Bailey Gallinger
Joie Davis
Nicole Vasilevsky
Heather Trang
Rena Godfrey
Gretchen Golas
Catherine Groden
Michele Nehrebecky
Ariane Soldatos
Elise Valkanas,
Colleen Wahl
Lynne Wolfe

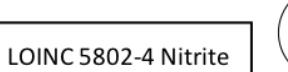
Basic research



Johannes Grosse
Attila Braun
David Varga-Szabo
Niklas Beyersdorf
Boris Schneider
Lutz Zeitlmann
Petra Hanke
Patricia Schropp
Silke Mühlstedt
Carolin Zorn
Michael Huber
Carolin Schmittwolf
Wolfgang Jagla
Philipp Yu
Thomas Kerkau
Harald Schulze
Michael Nehls
Bernhard Nieswandt

LOINC2HPO: Structuring EHR data

- EHR data are often encoded LOINC codes – a universal standard for coding clinical laboratory tests
- Multiple tests can be used for clinical findings
- We developed a curation tool that is used to convert EHR observations into HPO terms for use in clinical research

LOINC	Test outcome	HPO term
Qn	  	<p>HP:0002900 Hypokalemia</p> <p>NOT HP:0011042 NOT Abnormality of potassium homeostasis</p> <p>HP:0002153 Hyperkalemia</p>
Ord	 	<p>NOT HP:0031812 NOT Nitrituria</p> <p>HP:0031812 Nitrituria</p>
Nom	   	<p>HP:0040317 Blue urine</p> <p>HP:0040320 Red-brown urine</p> <p>HP:0032003 Green urine</p> <p>HP:0032002 Orange urine</p>



NATIONAL CENTER
FOR DATA TO HEALTH

Pilot study: UNC EHR Asthma Data

15k patients with multiple visits over 2 years

1/3 of patients with ICD9 Asthma diagnosis codes

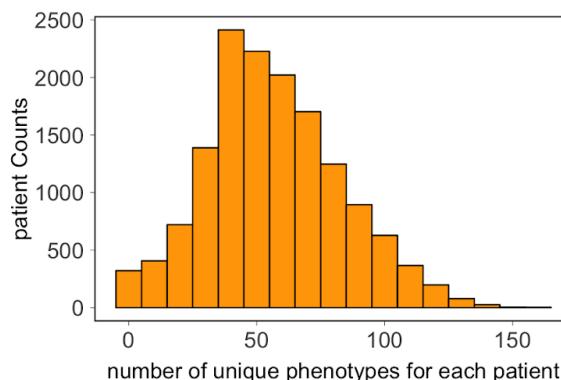
Mild
Prednisone $\leq 3x/2$ years

Moderate/severe
Prednisone $\geq 3x/2$ years



Laboratory tests total: 11 million mapped: 9.3

Converted 83% of LOINC tests to HPO codes



Each ind. pt had an average of 42 unique HPO terms

Many HPO terms significantly overrepresented in the “severe” group

Increased red blood cell count

Increased VLDL cholesterol concentration

Abnormal VLDL cholesterol concentration

Hypothenuria

Increased hematocrit

Abnormal eosinophil count

Eosinophilia



Identify rare disease patients and target diagnostics

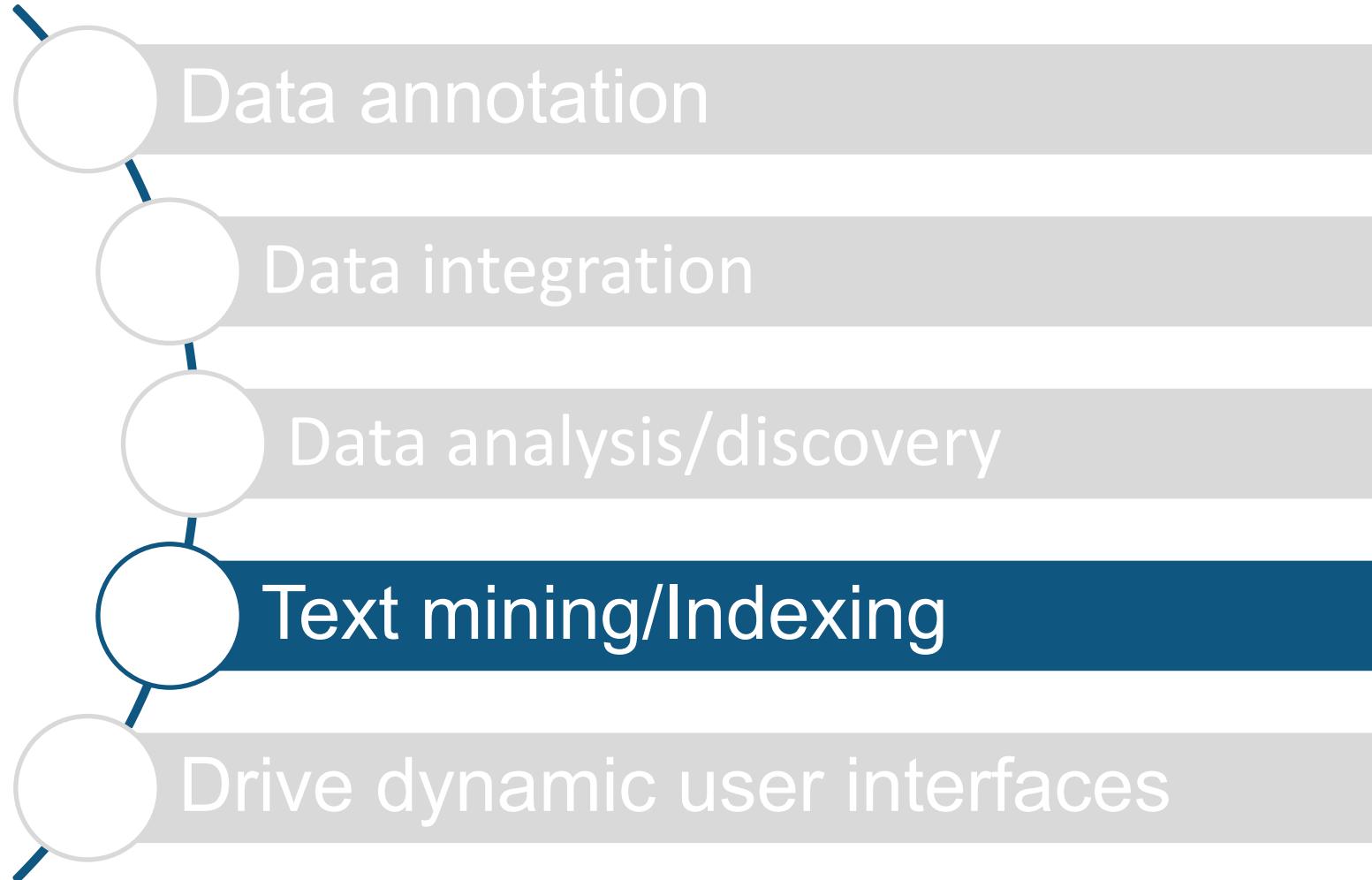


Support phenotype-driven diagnostics

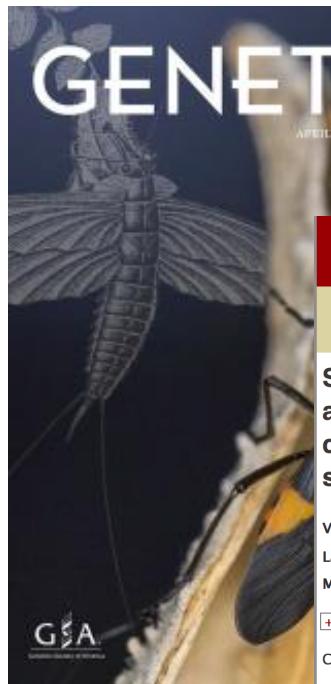


Discovery and mechanistic research

Applications of ontologies



Journals are now requiring HPO terms



OJRD ORPHANET JOURNAL
OF RARE DISEASES

Molecular Case Studies

HOME | ABOUT | ARCHIVE | SUBMIT | AUTHOR INFO | ADVERTISE | ALERTS / RSS | CONTACT | HELP

Sustained therapeutic response to riboflavin in a child with a progressive neurological condition, diagnosed by whole-exome sequencing

Vandana Shashi^{1,8}, Slavé Petrovski^{2,3,8}, Kelly Schoch¹, Rebecca Crimian¹, Laura E. Case⁴, Roha Khalid⁵, Maysantoine A. El-Dairi⁶, Yong-Hui Jiang^{1,7}, Mohamad A. Mikati⁵ and David B. Goldstein²

+ Author Affiliations

Corresponding authors: dg2875@cumc.columbia.edu, vandana.shashi@duke.edu

^{1,8} These authors contributed equally to this work.

Abstract

One of the most promising outcomes of whole-exome sequencing (WES) is the alteration of medical management following an accurate diagnosis in patients with previously unresolved disorders. Although case reports of targeted therapies resulting from WES have been published, there are few reports with long-term follow-up that confirm a sustained therapeutic response. Following a diagnosis by WES of Brown-Vialetto-Van Laere Syndrome 2 (BVVL2), high-dose riboflavin therapy was instituted in a 20-mo-old child. An immediate clinical response with stabilization of signs and symptoms was noted over the first 2-4 wk. Subsequent clinical follow-up over the following 8 mo demonstrates not just stabilization, but "Goldstein author:D.B." tained improvements in all manifestations of this usually fatal

doi: 10.1101/mcs.a000265
Cold Spring Harb Mol Case Stud 1: a000265
© 2015 Shashi et al.; Published by Cold Spring Harbor Laboratory Press

» Abstract
Full Text
Full Text (PDF)
Supplemental Material
Article Category
FOLLOW-UP REPORT
Services
Google Scholar
Articles by Shashi, V.
Articles by Goldstein, D. B.
PubMed/NCBI
Articles by Shashi, V.
Articles by Goldstein, D. B.
HPO Terms
drooling
gait imbalance
neurodegeneration
seesaw nystagmus
upper motor neuron abnormality

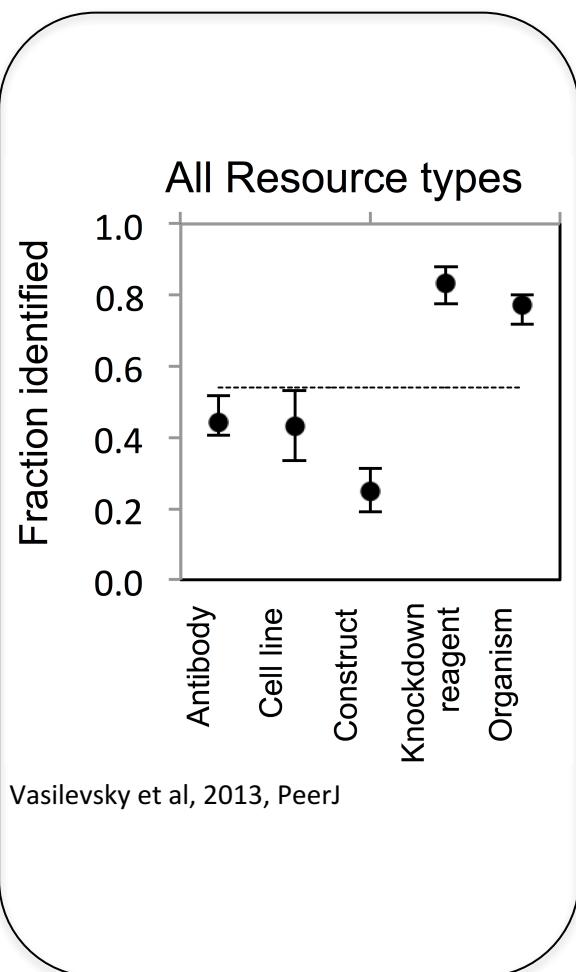
Explore HPO Ontology

HPO Terms

drooling
gait imbalance
neck muscle weakness
neurodegeneration
seesaw nystagmus
upper motor neuron abnormality

Resource Identification Initiative

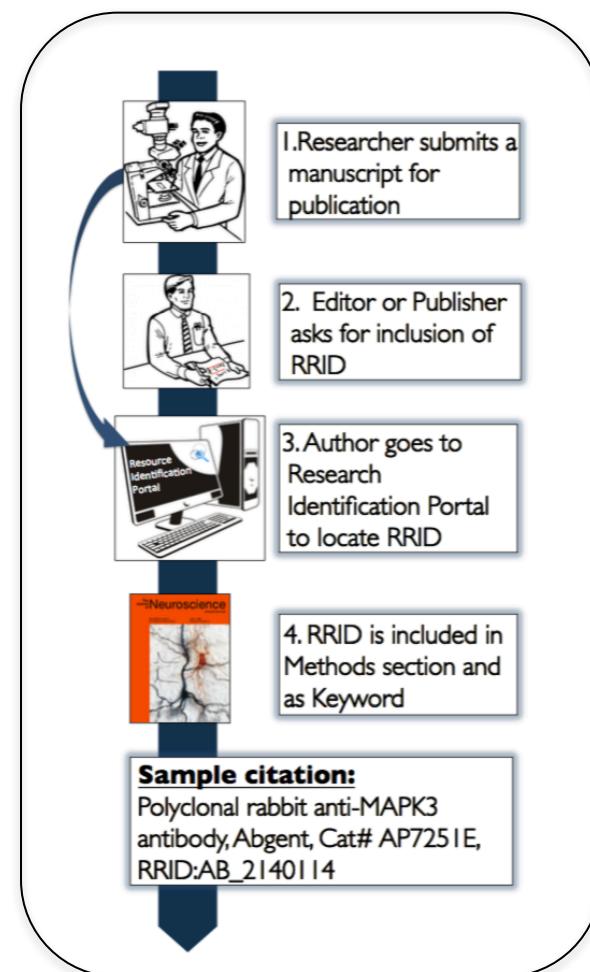
The problem



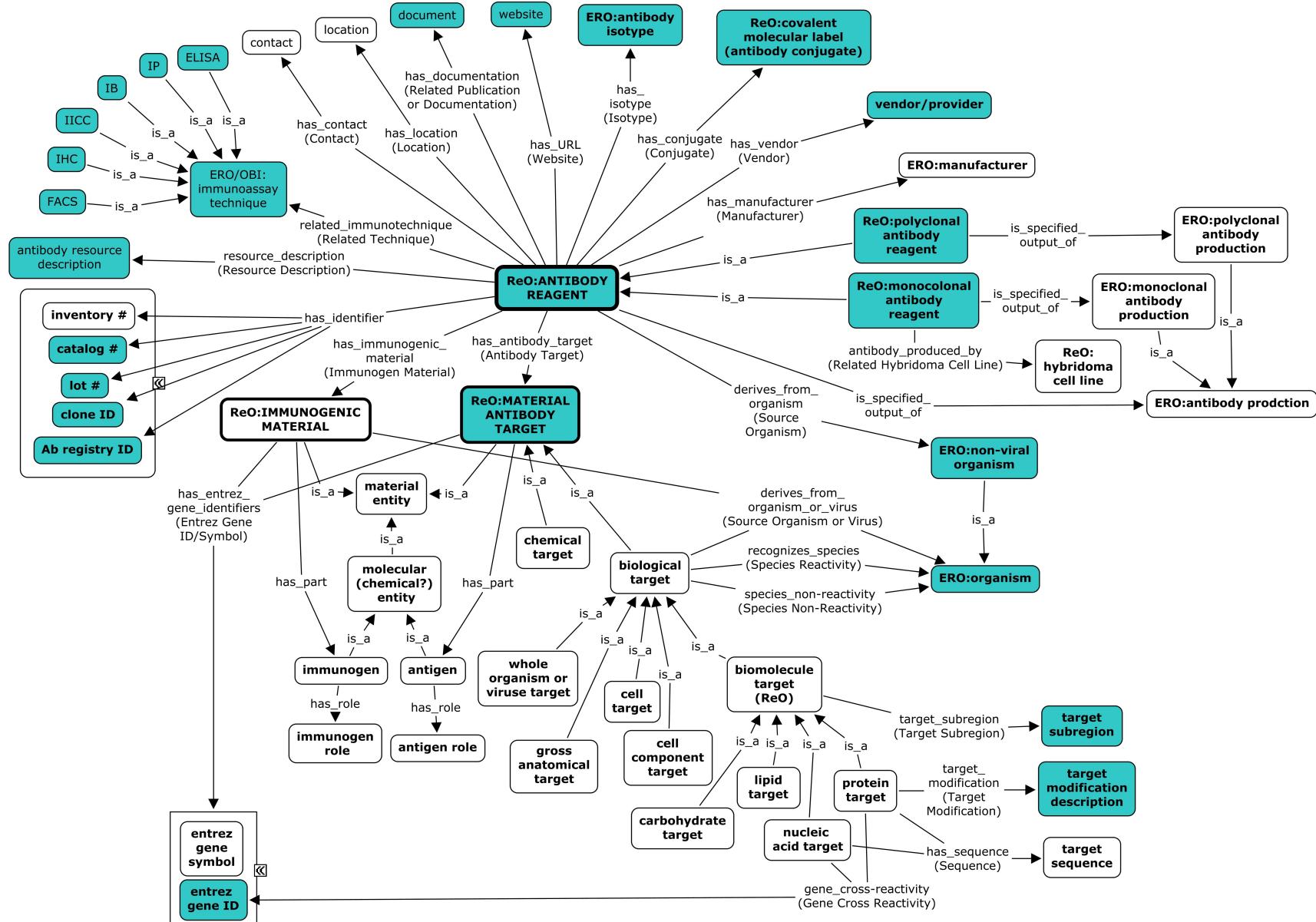
A solution



Workflow



Antibody Model

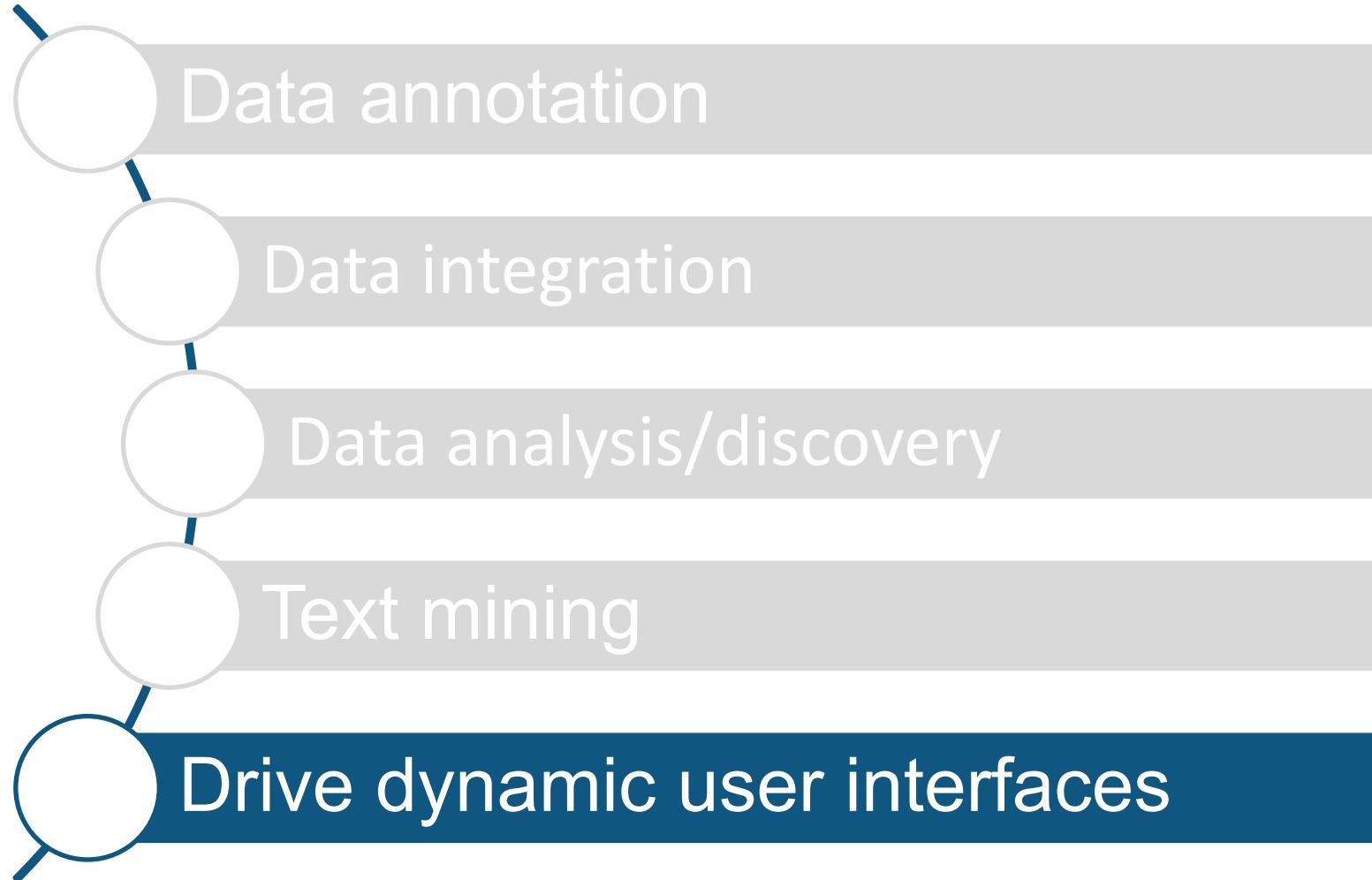


CRAFT Corpus

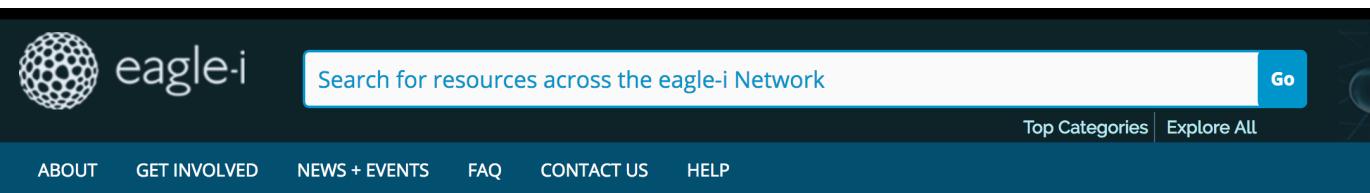
- Highly curated ‘Gold-standard’ corpora are needed for training and evaluation of NLP systems
- **97** full-length, open access biomedical journal articles

	# Total Annotations	Avg # Annotations per Article
ChEBI	8,137	121
Cell Ontology	5,760	86
Gene Ontology: biological processes	16,184	241
Gene Ontology: cellular components	8,354	125
Gene Ontology: Molecular Function	4,062	61
NCBITaxon	7,449	111
Protein Ontology	15,594	233
Sequence Ontology	22,090	330
Molecular Process Ontology	293	4
Uberon Anatomy Ontology	12,238	183

Applications of ontologies



Driving a user interface for quality data capture



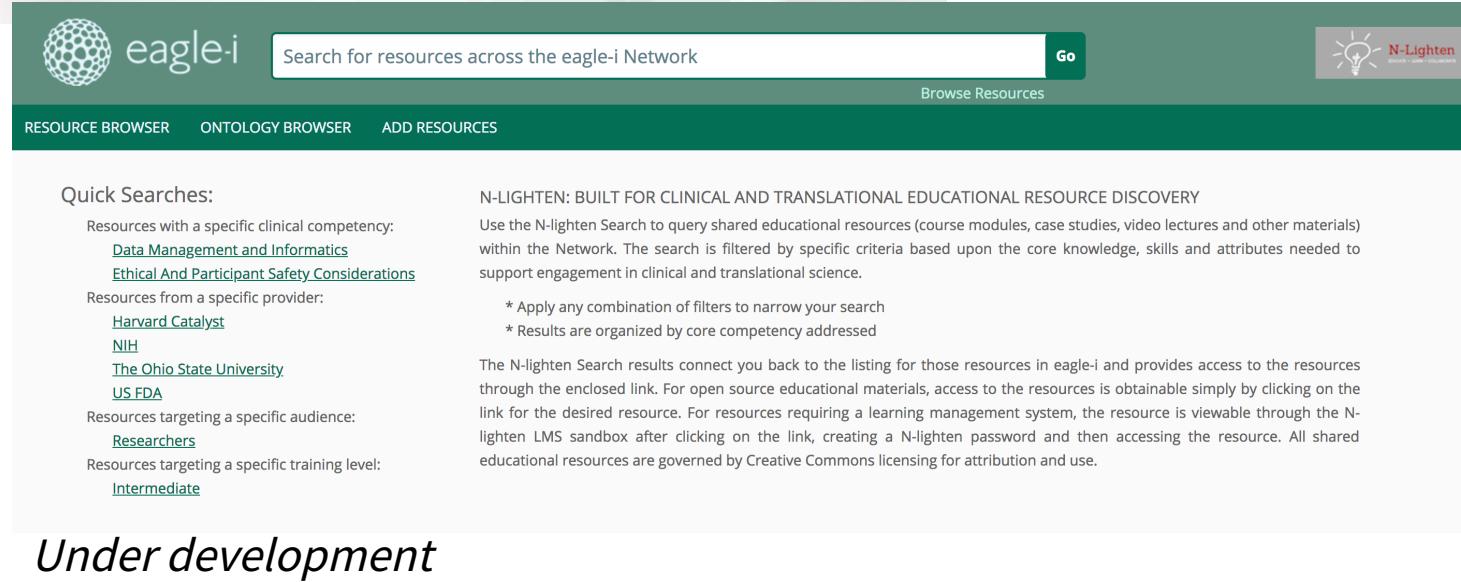
The screenshot shows the eagle-i Network search interface. At the top left is the eagle-i logo, which consists of a stylized globe icon followed by the text "eagle-i". To the right is a search bar with the placeholder text "Search for resources across the eagle-i Network". To the right of the search bar is a blue "Go" button. Below the search bar are two links: "Top Categories" and "Explore All". At the bottom of the header are navigation links: "ABOUT", "GET INVOLVED", "NEWS + EVENTS", "FAQ", "CONTACT US", and "HELP".

Making open science happen,
one resource at a time.

Search Now



www.eagle-i.net



The screenshot shows the eagle-i Network search interface integrated with N-lighten. At the top left is the eagle-i logo. To the right is a search bar with the placeholder text "Search for resources across the eagle-i Network" and a blue "Go" button. To the right of the search bar is a "Browse Resources" link. At the bottom of the header are three links: "RESOURCE BROWSER", "ONTOLOGY BROWSER", and "ADD RESOURCES".

Quick Searches:

Resources with a specific clinical competency:
[Data Management and Informatics](#)
[Ethical And Participant Safety Considerations](#)

Resources from a specific provider:
[Harvard Catalyst](#)
[NIH](#)
[The Ohio State University](#)
[US FDA](#)

Resources targeting a specific audience:
[Researchers](#)

Resources targeting a specific training level:
[Intermediate](#)

N-LIGHTEN: BUILT FOR CLINICAL AND TRANSLATIONAL EDUCATIONAL RESOURCE DISCOVERY

Use the N-lighten Search to query shared educational resources (course modules, case studies, video lectures and other materials) within the Network. The search is filtered by specific criteria based upon the core knowledge, skills and attributes needed to support engagement in clinical and translational science.

* Apply any combination of filters to narrow your search
* Results are organized by core competency addressed

The N-lighten Search results connect you back to the listing for those resources in eagle-i and provides access to the resources through the enclosed link. For open source educational materials, access to the resources is obtainable simply by clicking on the link for the desired resource. For resources requiring a learning management system, the resource is viewable through the N-lighten LMS sandbox after clicking on the link, creating a N-lighten password and then accessing the resource. All shared educational resources are governed by Creative Commons licensing for attribution and use.

Under development

Driving a user interface for quality data capture

SWEET Semantic Web Entry & Editing Tool

Workbench > Casey Eye Institute

Casey Eye Institute switch organizations

All Resource Types

- Biological Specimen add new
- Human Study add new
- Instrument add new
- Organism or Virus add new
- Protocol add new
- Reagent add new
- Research Opportunity add new
- Service add new
- Software add new

Status Legend

- Draft
- In Curation
- Published
- Withdrawn

Form Actions:

Save Save and unlock Cancel

Organism or Virus Name*

Organism or Virus Type* Request a term.

Organism or Virus Description

Organism or Virus Additional Name

Location Casey Eye Institute <Inst>

Contact <none> See choices from all organizations. + ×

Developed by <none> See choices from all organizations. + ×

Genetic alteration <none> See choices from all organizations. + ×

Related Technique Technique Request a term. + ×

Inventory Number

Related Publication or Documentation <none> See choices from all organizations. + ×

Transgenic Construct <none> See choices from all organizations. + ×

Medical Questionnaire Ontology (MQO)

Health questionnaire

1. Do you drink alcohol?
2. If yes, at what age did you begin drinking alcohol?
3. In the past year, which of the following alcoholic beverages have you used?
 - Beer, regular
 - Beer, light
 - Red wine
 - White wine
 - Liquor
 - None
4.

Alcohol consumption

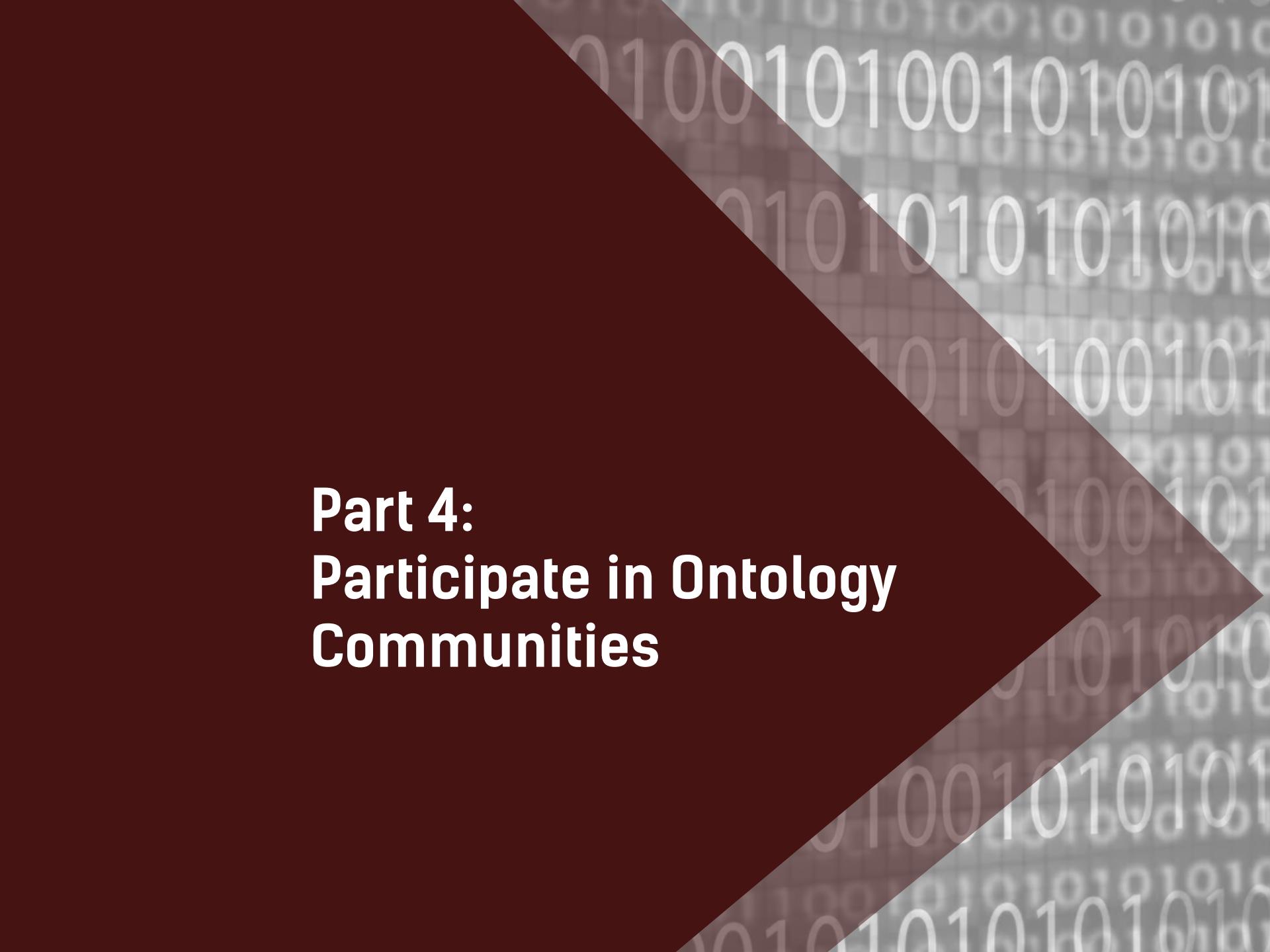
- ◆ Currently consumes alcohol
- ◆ Does not currently consume alcohol
- ◆ Never consumed alcohol

Age of initial alcohol consumption

Type of alcohol in past year

- ◆ Beer, regular
- ◆ Beer, light
- ◆ Red wine
- ◆ White wine
- ◆ Liquor
- ◆ None

Goal: The MQO can be used for automatic generation of health questionnaires in a computable format.



Part 4: **Participate in Ontology Communities**

Expert contributions to ontologies

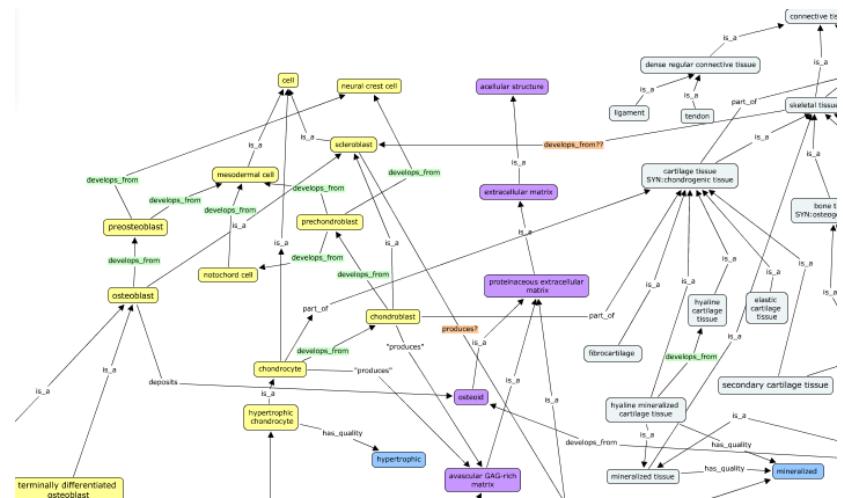


<https://bbsrc.ukri.org/news/people-skills-training/2015/151019-n-new-pool-of-experts-and-committee-members/>

How can I collaborate or stay informed?

- The OBO Foundry has a number of listservs to coordinate collaborative efforts
 - Many ontologies are linked from the OBO Foundry site, if you would like to contribute to specific ontologies
- Note that most of the listservs are fairly low traffic, and these are archived

Simple tools for modeling a domain



Spreadsheets:

- Google docs
 - Excel

Visualization software

- Cmap
 - Vue
 - Caggle
 - GraphViz

Ontology issue trackers

- A tracker is a place to put a formal ontology request
- Trackers have long been used in the software community for keeping track of bugs, feature requests, etc
- In the ontology community, they are quite valuable because they provide an open, documented, structured requests for changes or additions
- Tracker IDs can be referenced in ontology metadata-such as in an editor note or definition annotation

Github: A common ontology tracker

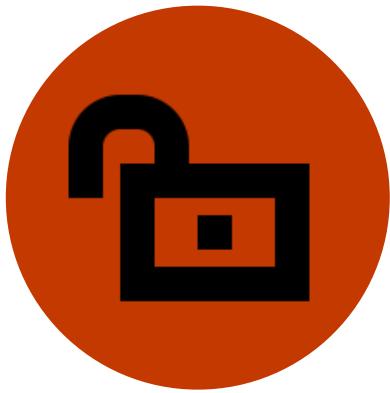


A screenshot of a GitHub repository page for "obophenotype / human-phenotype-ontology". The page shows the repository's main statistics: 321 issues (321 open, 3,814 closed), 2 pull requests, 6 projects, and a wiki. The "Issues" tab is selected. A search bar filters for "is:issue is:open". A green "New issue" button is visible. The list of open issues includes:

- ① NTR: splenopancreatic fusion new term request #4156 opened an hour ago by nicolevasilevsky
- ① NTR: Lower cranial nerve dysfunction new term request #4155 opened an hour ago by nicolevasilevsky
- ① add syn to 'Subvalvular aortic stenosis' synonyms #4154 opened an hour ago by nicolevasilevsky
- ① Synonym request: Abnormality of stapedial reflex #4153 opened 3 hours ago by LCCarmody
- ① add syn to Hyperpigmentation of the skin? synonyms #4152 opened 4 hours ago by nicolevasilevsky
- ① add syn to Preauricular skin tag synonyms #4151 opened 4 hours ago by nicolevasilevsky

<https://github.com/obophenotype/human-phenotype-ontology/issues>

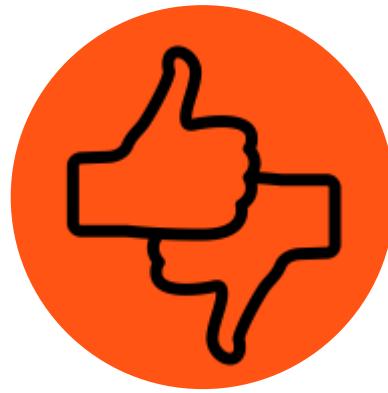
Ontology trackers



Open



Time



Discussions

How do you write a tracker request?

- 1) Provide as much information as possible, in order to facilitate the change you are requesting and future reference
- 2) For new terms, or term rearrangements, provide the intended hierarchy – both SubClass as well as any other relations required (such as partonomy)
- 3) Provide text definitions, that make sense in the Genus Differentia context, for all new or edited terms
- 4) Provide attribution for the definitions
- 5) Provide (nano)attribution for yourself

Example tracker request

NTR: splenopancreatic fusion #4156

Edit

New issue

Open

nicolevasilevsky opened this issue an hour ago · 0 comments



nicolevasilevsky commented an hour ago

Member



...

For new term requests, please provide us with the following information:

1. Preferred term label

Splenopancreatic fusion

2. Synonyms

Spleno-pancreatic fusion

Splenic pancreatic fusion

3. Textual definition (should be understandable even for non-specialists, please include a PubMed ID for relevant articles providing additional information about the suggestion)

Fusion of the pancreatic tail and spleen.

PMID:18398855

4. Parent term (use HPO app, HPO Browser or OLS)

HP_0025408 'Abnormal spleen morphology'

HP_0012090 'Abnormality of pancreas morphology'

Assignees

pnrobinson

Labels

new term request

Projects

None yet

Milestone

No milestone

Notifications

Unsubscribe

You're receiving notifications because you authored the thread.

2 participants



Example tracker request

5. Which diseases are characterized by this term ? (e.g. Orphanet- or OMIM-id)

Schinzel-Giedion syndrome (PMID:18398855)
trisomy 13 (PMID:28843273)

 Lock conversation

6. Your nano-attribution (ORCID-id or label, e.g. *HPO:probinson* (organization:name))

<https://orcid.org/0000-0001-5208-3432>

This is for the Kid's First Data Portal - the description is 'Splenopancreatic fusion (microscopic)'

Practice making a term request

- Go to: https://github.com/nicolevasilevsky/CSH_IntroToOntologies/issues
(Short URL): www.bit.ly/cshissue
 - Create a new term request for a term in any ontology
-

Example:

New term request for HPO: Presence of anti-pollen antibody in blood

- Look up recommended parent class
- Provide a text definition
- Provide a source (like a PubMed ID)
- Provide nano-attribution

HPO Workbench

- Users can view HPO and HPO annotations via HPO workbench
- Can make suggestions via Workbench:
 - New child terms
 - New annotations
 - Corrections
 - Mistaken annotations

The screenshot shows the HPO Workbench application window. The title bar reads "Human Phenotype Ontology Workbench". The main content area displays "SICKLE CELL ANEMIA" with its "Disease ID: OMIM:603903" and "Name: SICKLE CELL ANEMIA". A sidebar on the left lists navigation options: "Browse HPO Terms or Diseases" (selected), "HPO Terms", "Diseases" (selected), and "New Annotation". Below this is a search bar with "autocomplete HPO term..." and a "Go" button. A list of clinical features is shown: "Clinical course", "Clinical modifier", "Frequency", "Mode of inheritance", and "Phenotypic abnormality". The main right panel is titled "HPO Terms (n=17)" and lists 17 terms with their IDs, labels, and definitions. The first few rows include: "HP:0001744 Splenomegaly" (Abnormal increased size of the spleen), "HP:0001878 Hemolytic anemia" (A type of anemia caused by premature destruction of red blood cells (hemolysis)), and "HP:0200023 Priapism" (A painful and harmful medical condition in which the erect penis doesn't return to its flaccid state, despite the absence of both physical and psychological stimulation, within four hours). Other terms listed are Hematuria, Leukocytosis, Autosomal recessive inheritance, Cholelithiasis, and others. At the bottom of the main panel are buttons for "Export hierarchical Summary", "Suggest correction", "Export ontology as Excel file", "Suggest new child term", "Suggest new annotation", and "Report mistaken annotation". To the right, there's a "Disease databases:" section with checkboxes for "All" (selected), "Orphanet", "OMIM", and "DECIPHER". The footer of the window says "HPO Workbench, v. 0.1.8, © Monarch Initiative 2018".

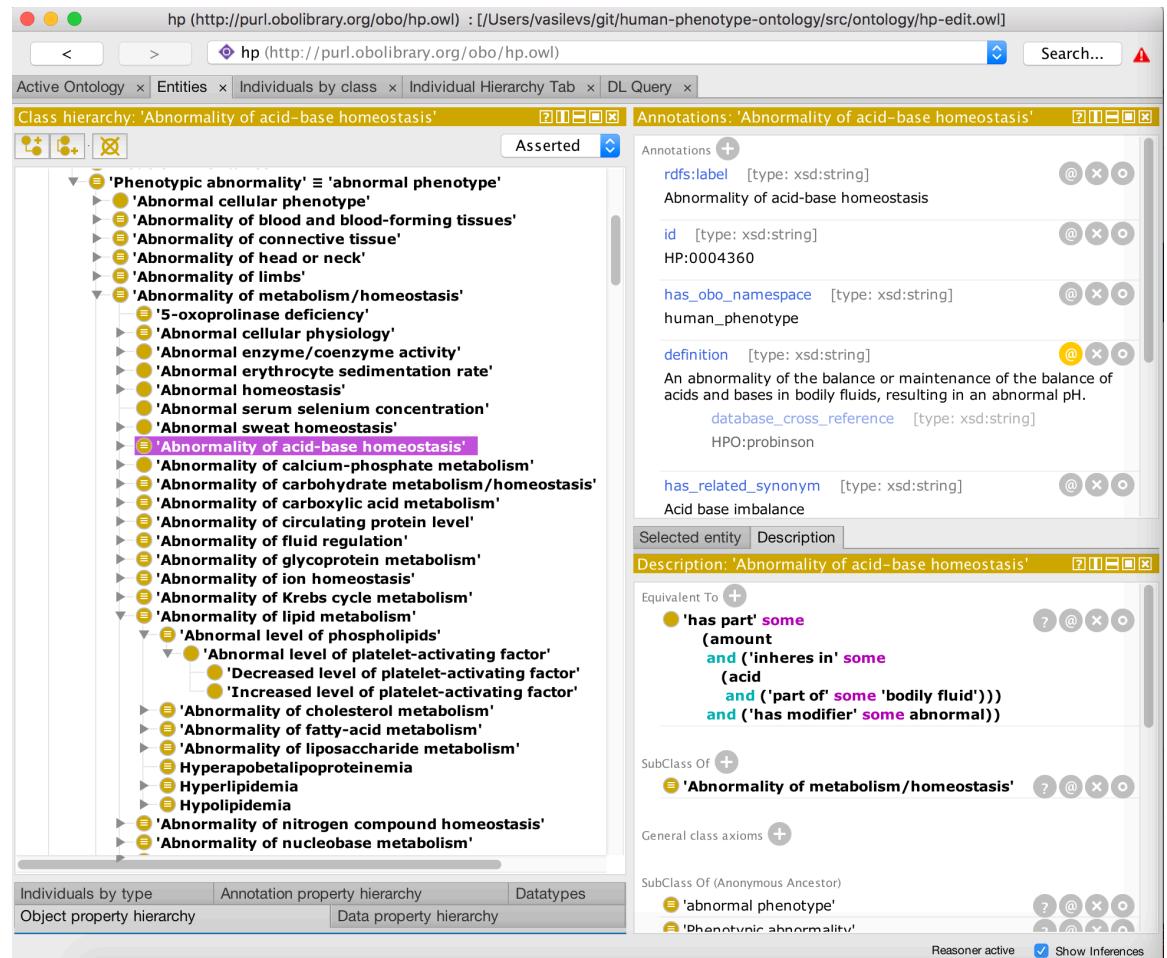
ID	Label	Definition
HP:0001744	Splenomegaly	Abnormal increased size of the spleen.
HP:0001878	Hemolytic anemia	A type of anemia caused by premature destruction of red blood cells (hemolysis).
HP:0200023	Priapism	A painful and harmful medical condition in which the erect penis doesn't return to its flaccid state, despite the absence of both physical and psychological stimulation, within four hours.
HP:0000790	Hematuria	The presence of blood in the urine. Hematuria may be gross hematuria (visible to the naked eye) or microscopic hematuria (detected by dipstick or microscopic examination of the urine).
HP:0001974	Leukocytosis	An abnormal increase in the number of leukocytes in the blood.
HP:0000007	Autosomal recessive inheritance	A mode of inheritance that is observed for traits related to a gene encoded on one of the autosomes (i.e., the human chromosomes 1-22) in which a trait manifests in homozygotes. In the context of medical genetics, autosomal recessive disorders manifest in homozygotes (with two copies of the mutant allele) or compound heterozygotes (whereby each copy of a gene has a distinct mutant allele).
HP:0001081	Cholelithiasis	Hard, pebble-like deposits that form within the gallbladder.

Part 5: Tools for Building Ontologies

Protégé

Many ontologies
are built using the
software tool,
Protégé

Can also use
Protégé for
viewing ontologies



ROBOT

[view on github](#)
[getting started](#)
[common errors](#)
[chaining commands](#)
[global options](#)
[makefile](#)

ROBOT is an OBO Tool

ROBOT is a tool for working with [Open Biomedical Ontologies](#). It can be used as a command-line tool or as a library for any language on the Java Virtual Machine.

Click on the command names in the sidebar for documentation and examples, and visit our JavaDocs for [robot-core](#) and [robot-command](#) for technical details.

For a “how-to” covering the major commands and features of ROBOT, visit our tutorial located [here](#).

Want to learn more about ontologies?

Online Ontology Tutorial

What: Basics of ontology development using software tool Protégé

Website: <https://github.com/OHSUBD2K/BDK14-Ontologies-101>

What: ROBOT tutorial

Website: <https://github.com/rctauber/robot-tutorial>

Protégé Course

Name: Protégé Short Course

When: TBD (Last course was Oct 22-24, 2018)

Where: Stanford University, Stanford, CA

Website: <https://protege.stanford.edu/short-courses.php>



Part 6: Resources

Ontology Conferences

ICBO

International Conference on Biomedical Ontologies (ICBO) 2019

When: TBD

Where: University of Buffalo, Buffalo, NY

Website: <http://icbo.buffalo.edu/index.html>



ISMB

International Systems for Molecular Biology

When: July 21 - 25, 2019

Where: Basel, Switzerland

Website: <https://www.iscb.org/ismbeccb2019>





Biocuration 2019

7th - 10th April 2019

The West Road Concert Hall, Cambridge, UK

<https://www.biocuration2019.org/>

Resources

Ontology browsers:

- OBO Foundry: <http://www.obofoundry.org/>
- BioPortal: <http://bioportal.bioontology.org/>
- Ontology Lookup Service: <https://www.ebi.ac.uk/ols/index>
- Ontobee: <http://www.ontobee.org/>
- QuickGO: <https://www.ebi.ac.uk/QuickGO/>
- AmiGO: <http://amigo.geneontology.org/amigo>

Related Reading:

- What are ontologies: <https://ontotext.com/knowledgehub/fundamentals/what-are-ontologies/>
 - Short article describing the value of ontologies
- Recommended readings https://github.com/OHSUBD2K/BDK14-Ontologies-101/blob/master/BDK14_exercises/BDK14_RecommendedReadings.pdf
 - A more extensive list of recommended readings

Acknowledgements

Translational and Integrative Sciences Lab and Monarch Initiative

- Melissa Haendel
- Carlo Torniai
- Chris Mungall
- Peter Robinson
- Sebastian Köhler
- Jim Balhoff
- Matt Brush
- David Osumi-Sutherland
- Nico Matentzoglu
- Julie McMurry (graphic design)

Some content was adapted from:

BDK14: Introduction to OWL2 and data reasoning. Nicole Vasilevsky, Melissa Haendel, Chris Mungall, David Osumi-Sutherland, Matt Yoder, Carlo Torniai, and Simon Jupp.
<https://github.com/OHSUBD2K/BDK14-Ontologies-101>



Part 7: **Exercises**

Ontology Problem set

Go to:

[https://github.com/prog4biol/pfb2017/blob/master/
problemsets/ontology_problemset.md](https://github.com/prog4biol/pfb2017/blob/master/problemsets/ontology_problemset.md)

(Short URL): www.bit.ly/ontologyproblemset

Modeling a domain

- What domains are you working on?
- If you had to build an ontology, what are some classes you would include?
- What existing ontologies could you reuse?

Thank you!

Contact:

 vasilevs@ohsu.edu

 [@N_Vasilevsky](https://twitter.com/N_Vasilevsky)