A Review of Relevant Ontologies and Application of Reasoners

monarch INITIATIVE

Melissa Haendel, PhD

@ontowonka

OREGON HEALTH & SCIENCE UNIVERSITY

Outline

- Using ontologies and reasoners for classification
- Anatomy and Stage Ontologies
- Example of ontologies and reasoning at work: diagnosing diseases
- Environmental ontologies
- How to exchange data better

What is an Ontology?

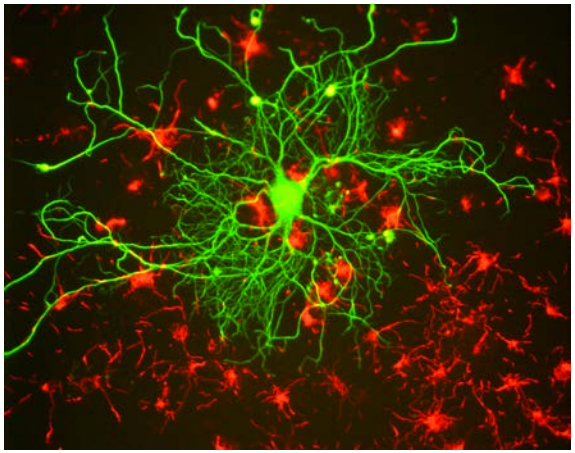
Definition:
A formal conceptualization of a specified domain

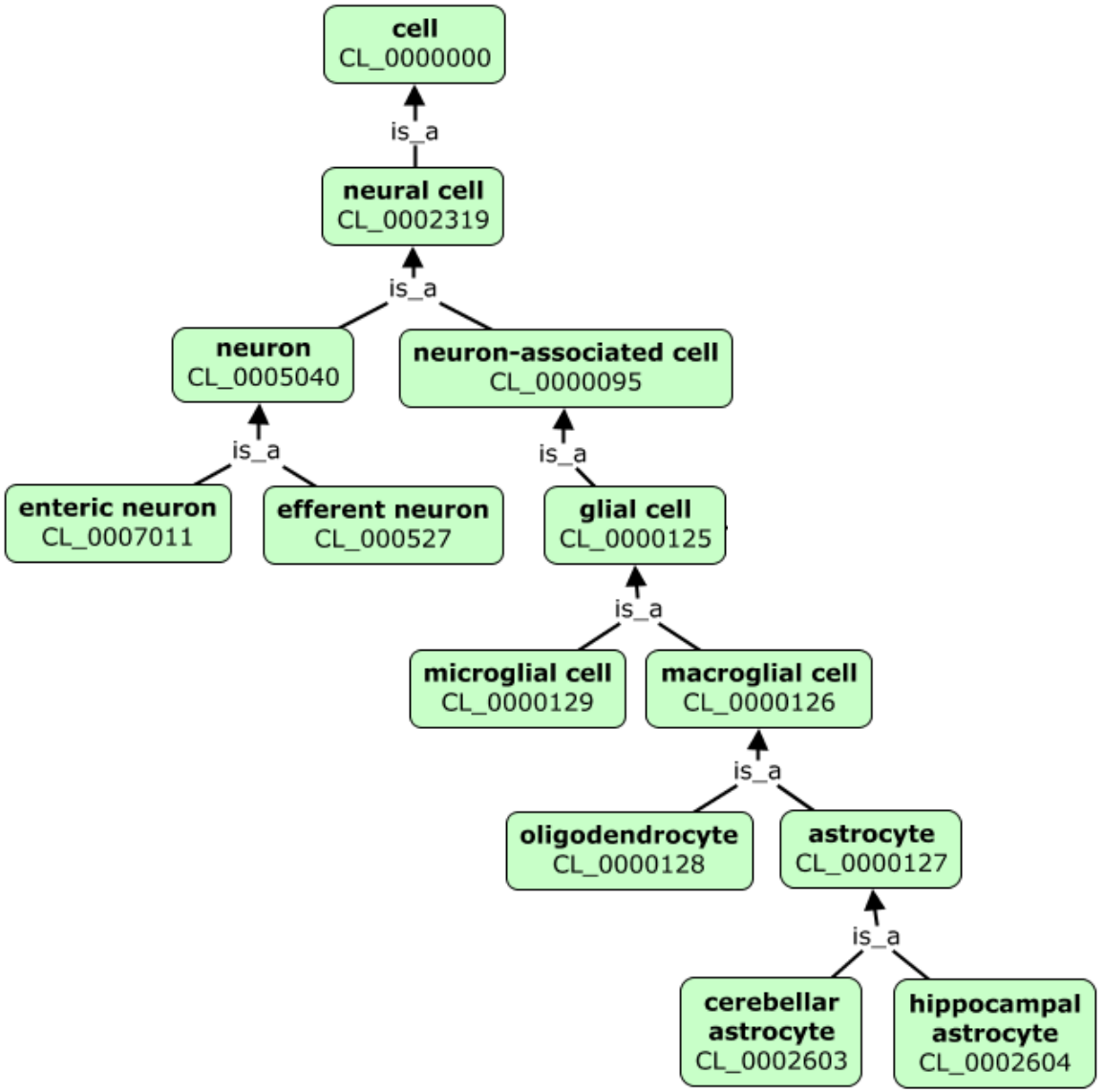
Key Features:

- Terms are defined
- 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

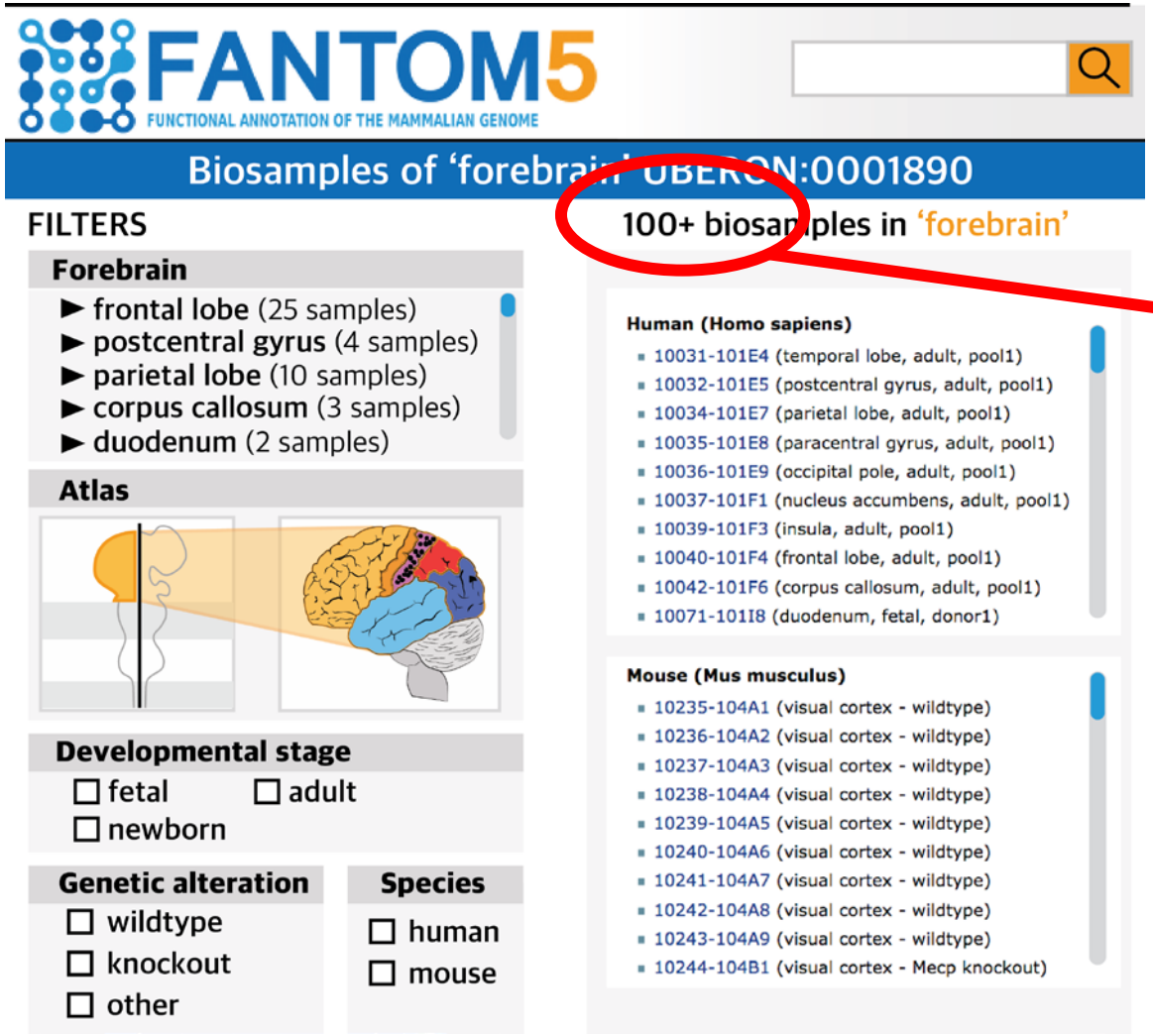
Examples:
SNOMED, Foundational Model of Anatomy, Gene Ontology, Linnean Taxonomy of species

Example taxonomy





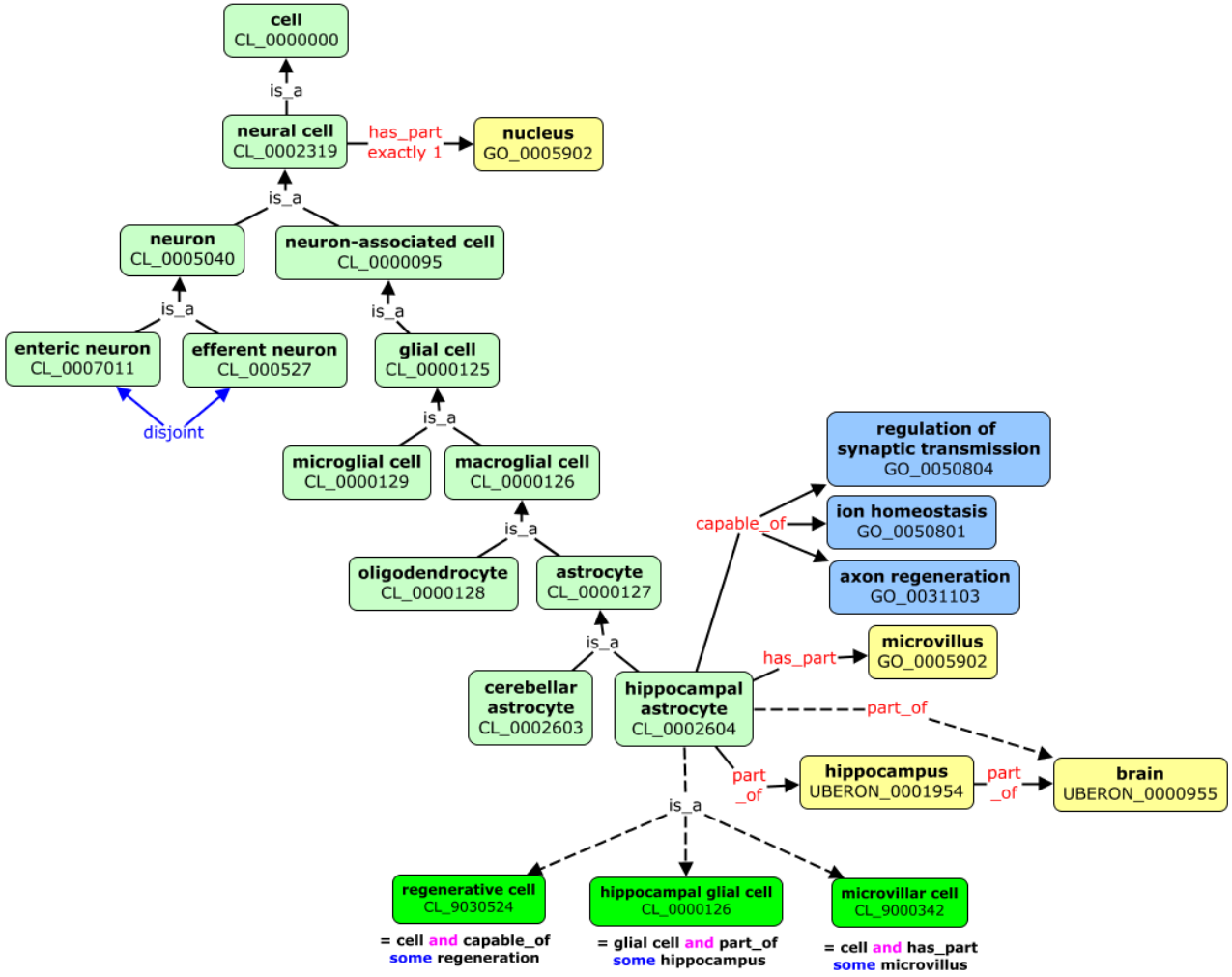
Ontologies enable queries to “just work” as you would hope.



Without ontological “subsumption” reasoning, synonym formalism, the user would either need to do 17 different queries, or get an incomplete set of results.

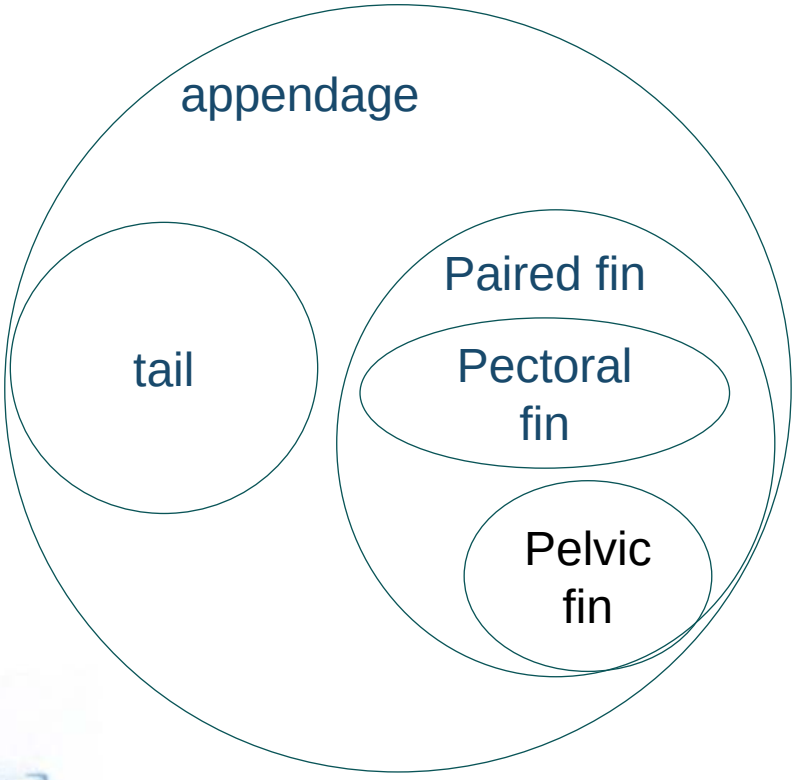
Ontologies

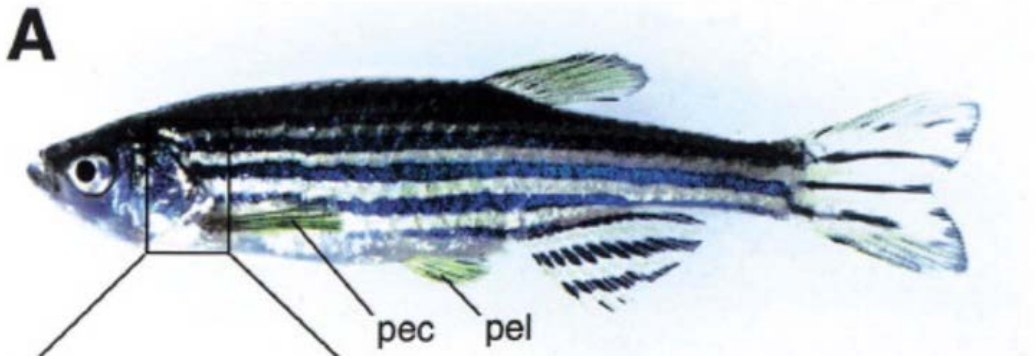
Ontologies support automated ***consistency checking***, ***inferred classification*** along different axes, and powerful ***graph-based applications***



Ontologies are formal classifications

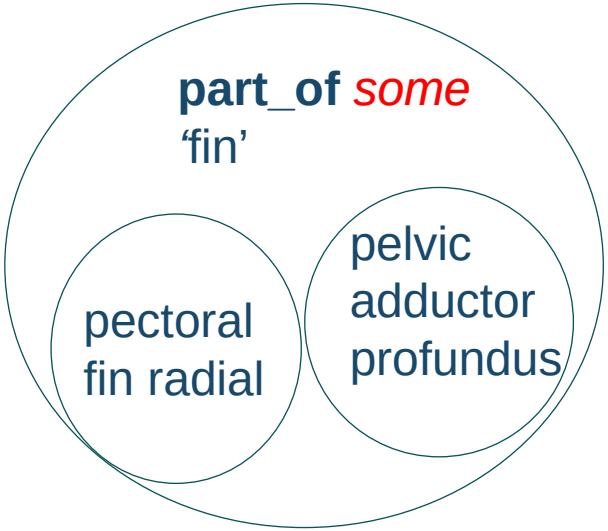
- Appendage
  - Tail
  - Median fin
  - Paired fin
    - Pectoral fin
    - Pelvic fin

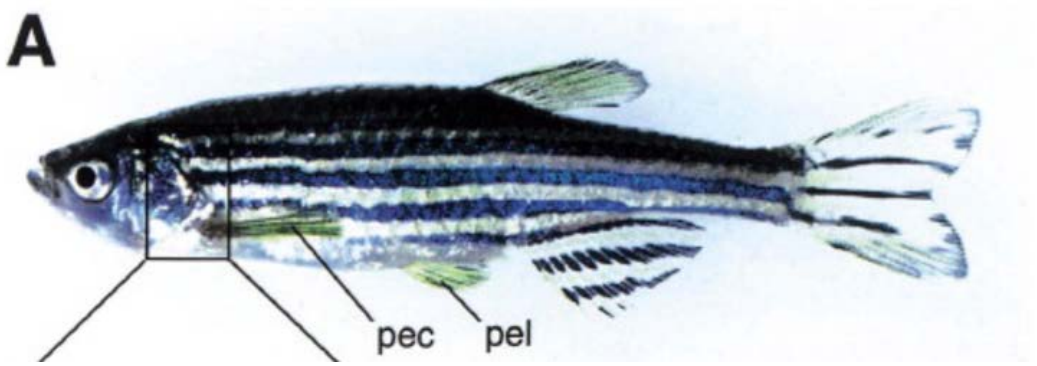




Relationships also support classification

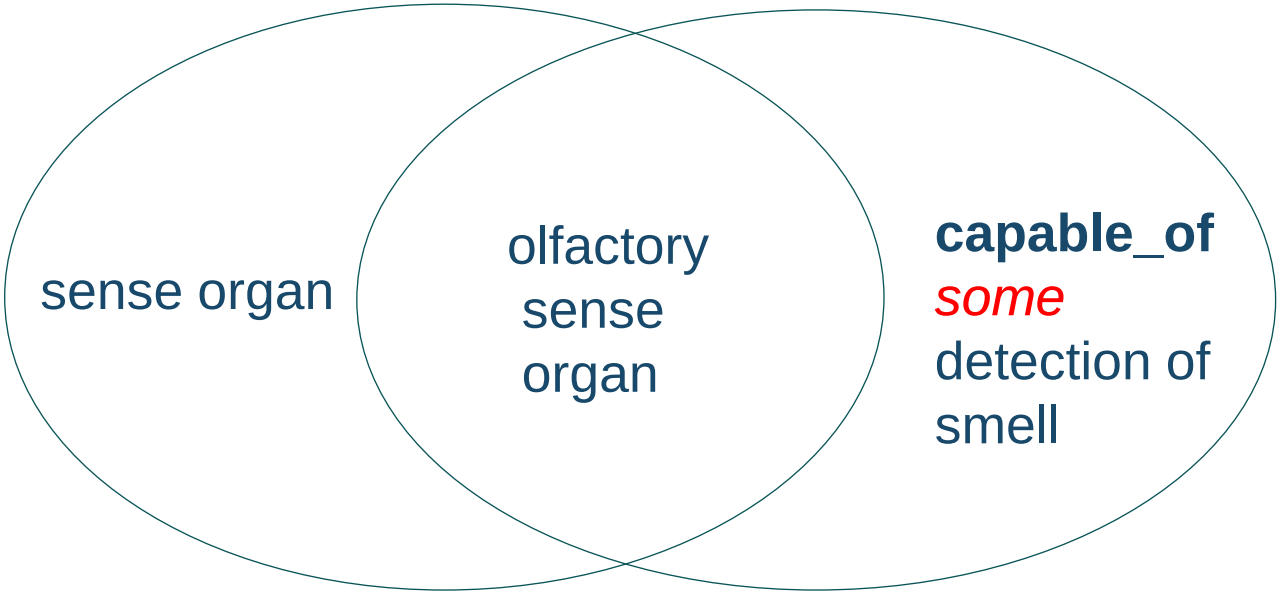
‘pectoral fin radial’ *SubClassOf* **part_of** *some* ‘fin’



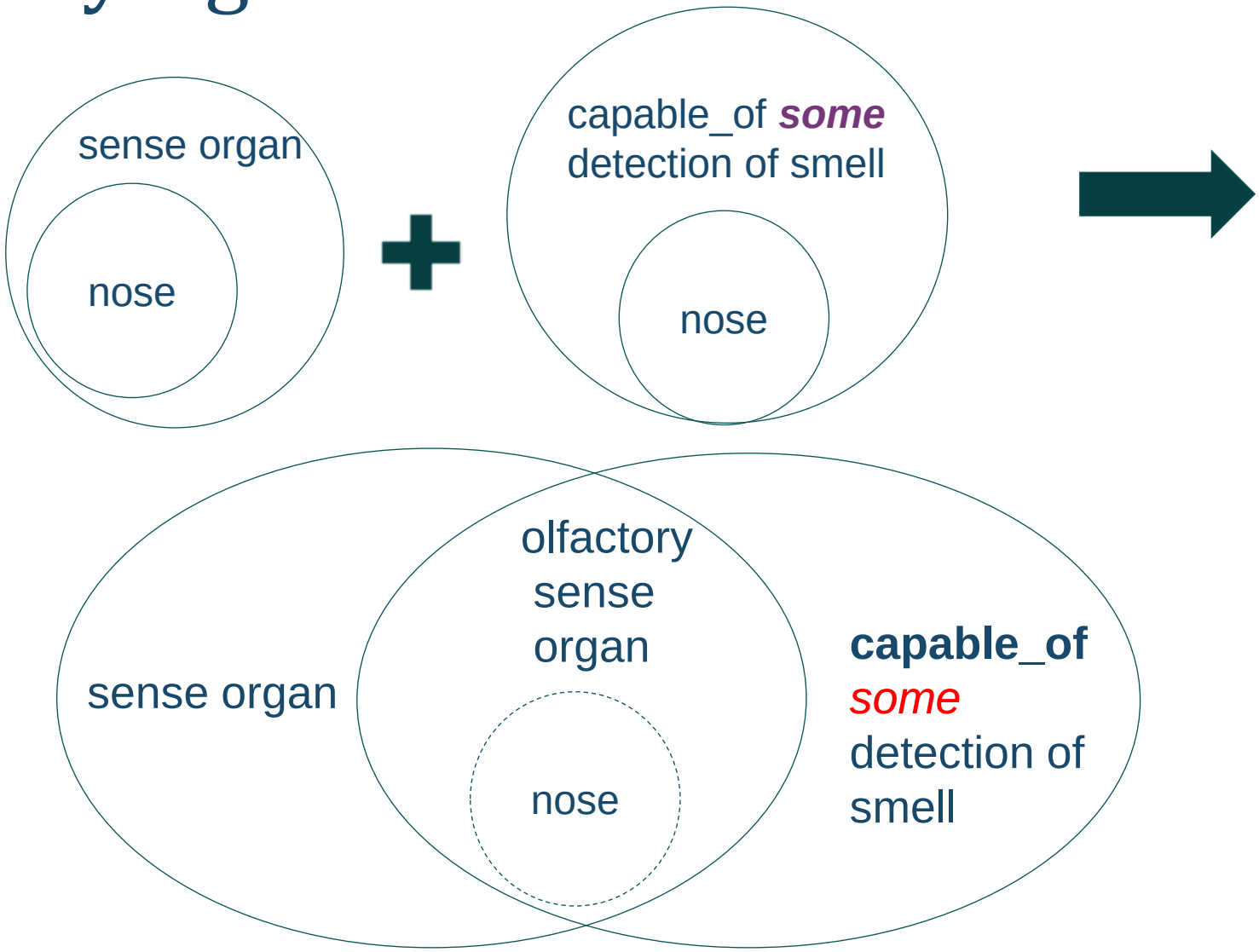


Necessary and sufficient conditions

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

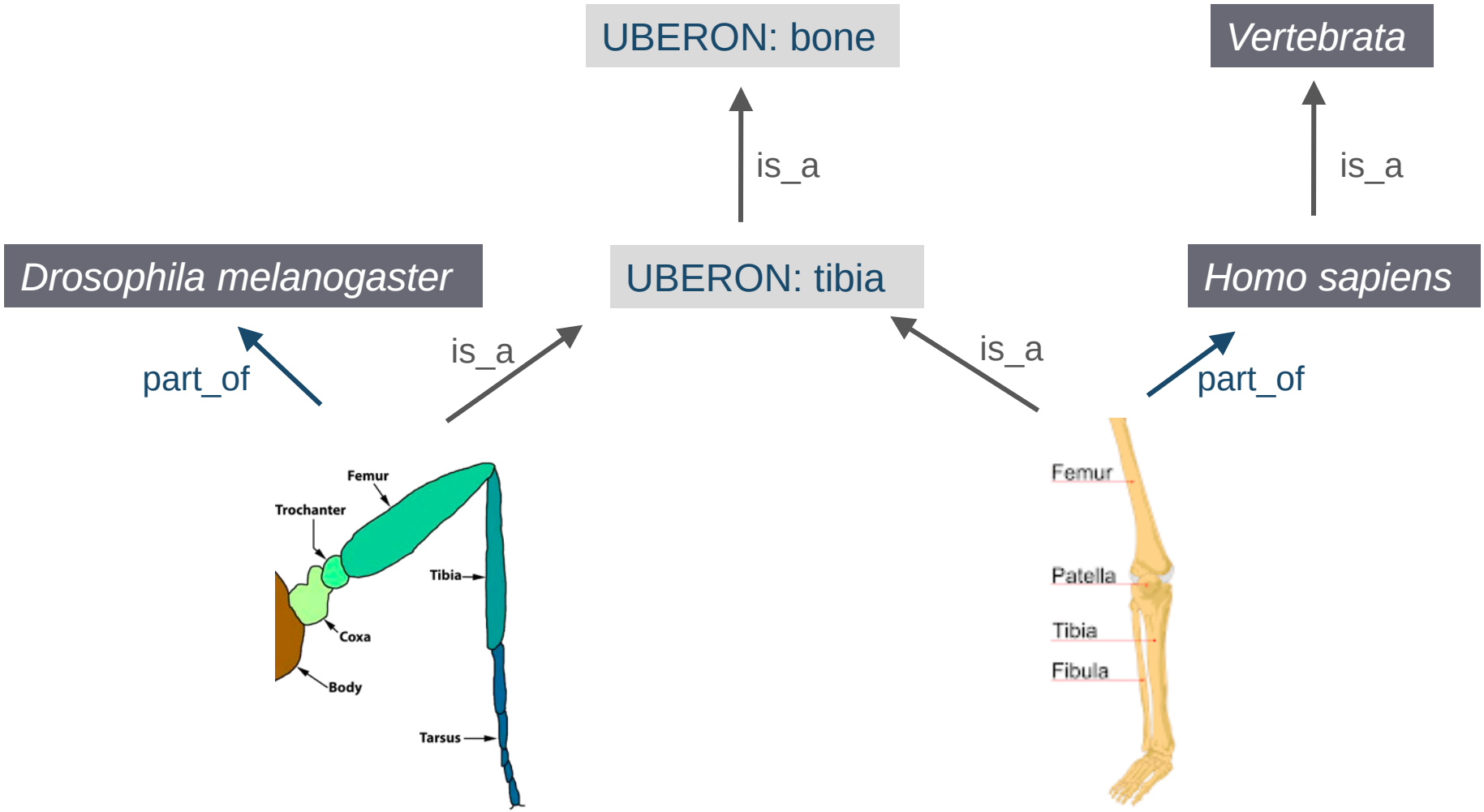


Classifying



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

Using reasoners to detect errors



Fruit fly FBbt ‘tibia’

Human FMA ‘tibia’

Using reasoners to detect errors

UBERON: bone

only_in_taxon

Vertebrata

is_a

is_a

Drosophila melanogaster

UBERON: tibia

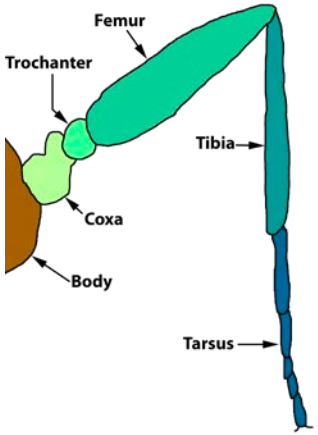
Homo sapiens

part_of

is_a

is_a

part_of



Fruit fly FBbt ‘tibia’

Human FMA ‘tibia’

Using reasoners to detect errors

UBERON: bone

only_in_taxon

Vertebrata

disjoint_with

is_a

is_a

Drosophila melanogaster

UBERON: tibia

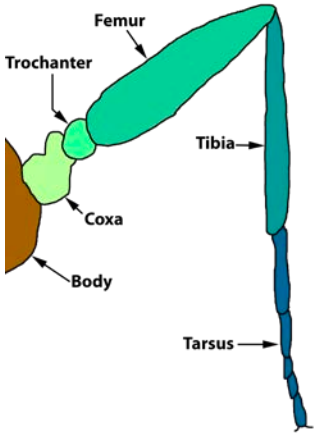
Homo sapiens

part_of

is_a

is_a

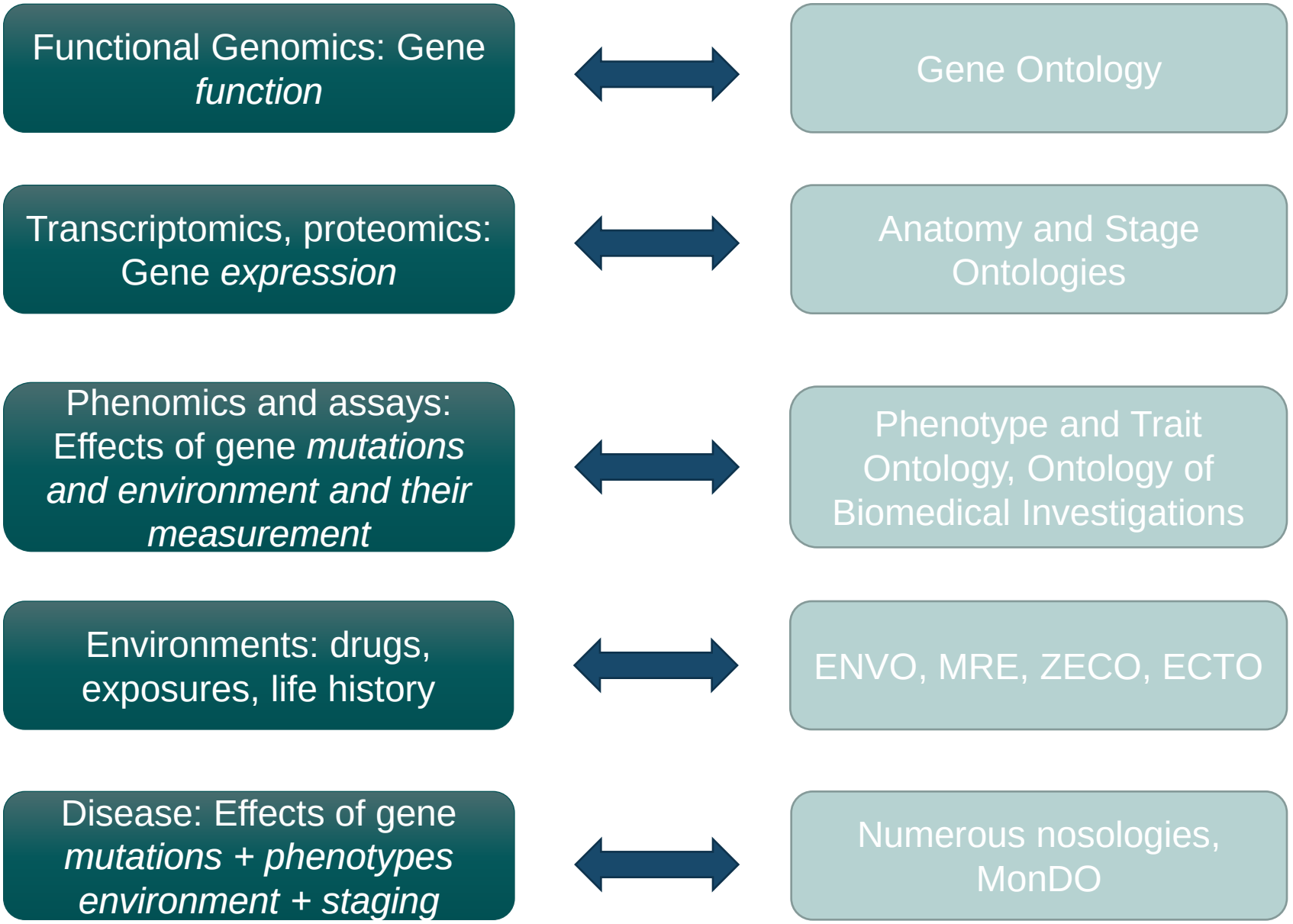
part_of



Fruit fly FBbt ‘tibia’

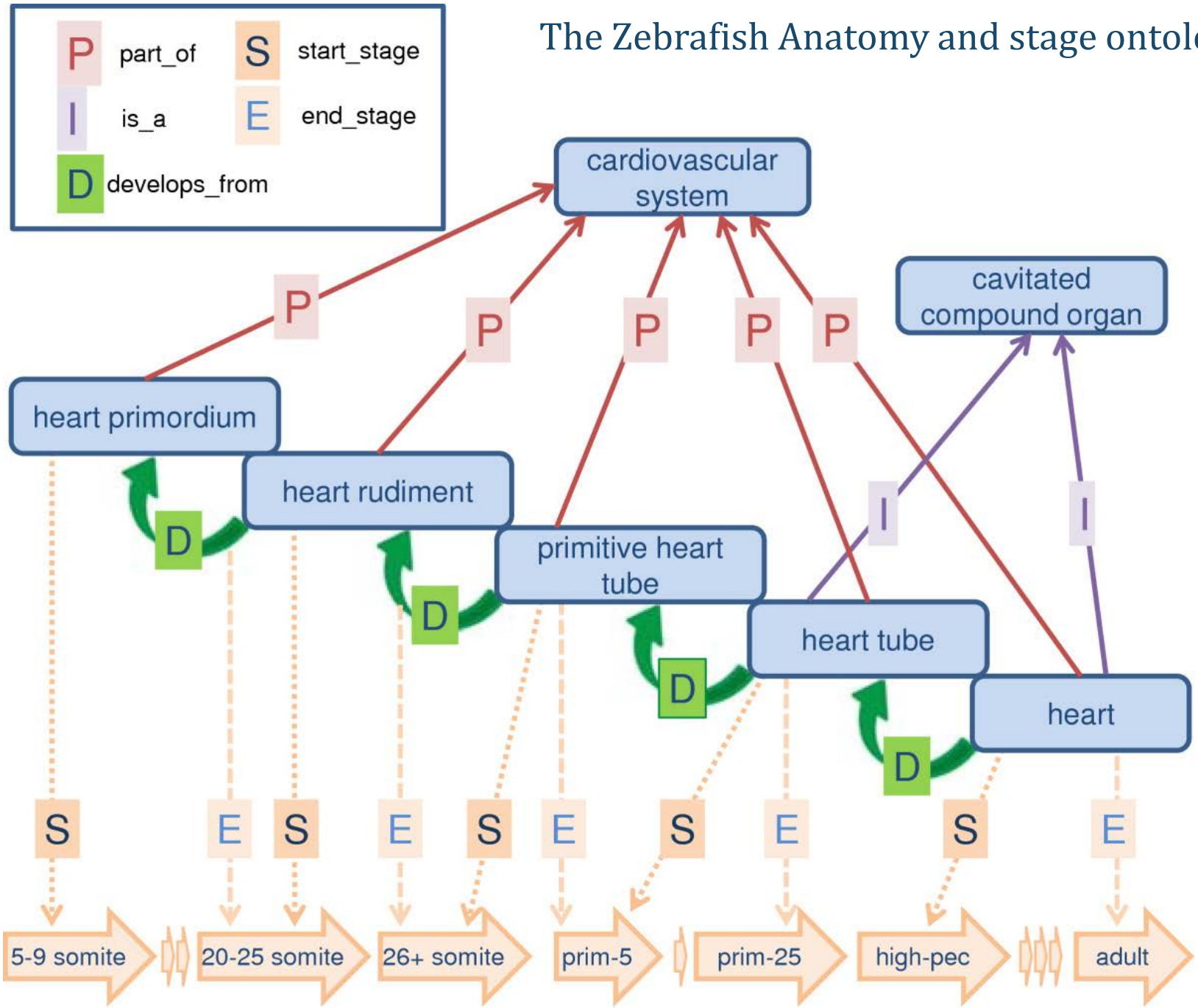
Human FMA ‘tibia’

A compendium of interoperable ontologies

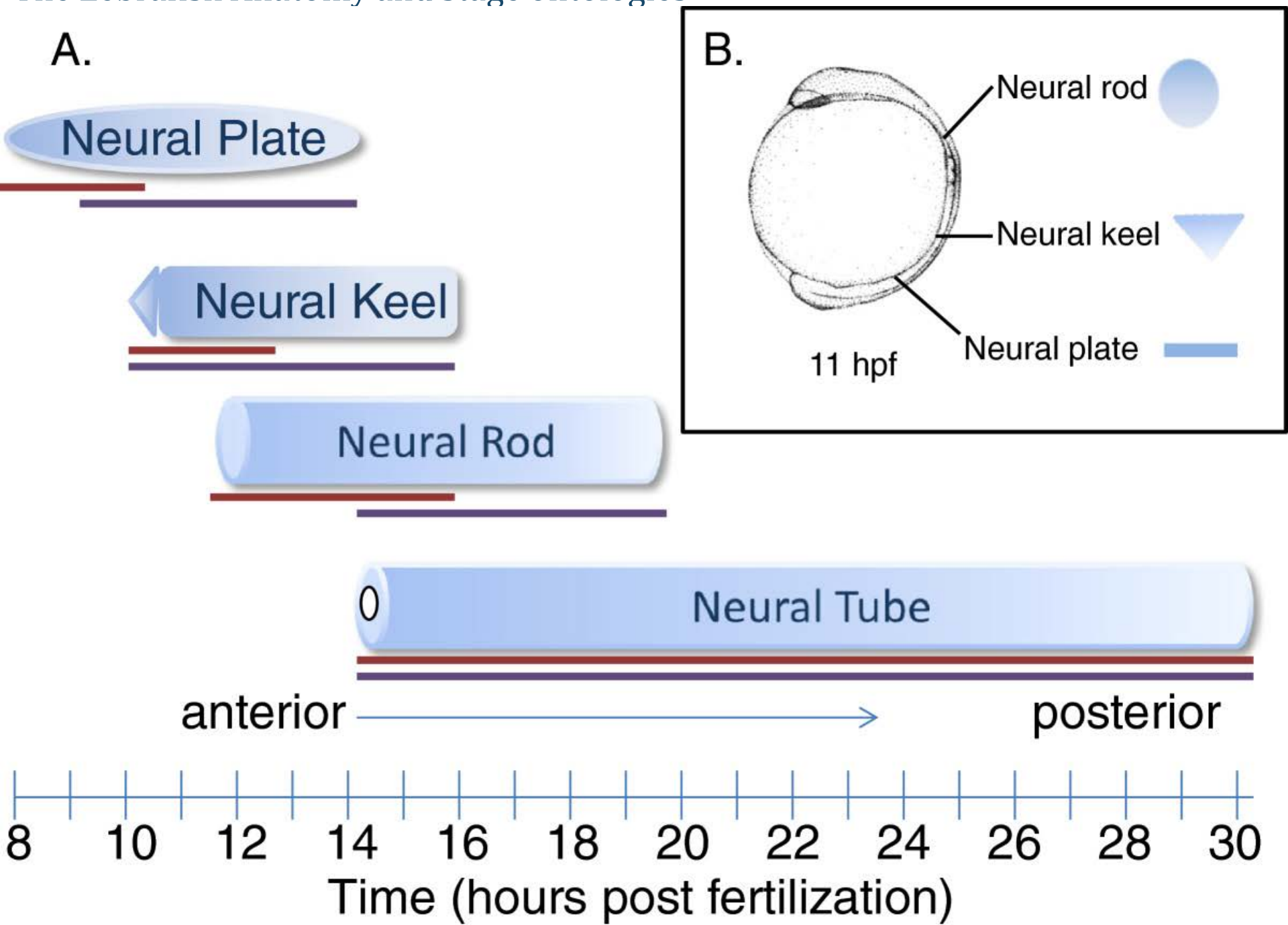


Anatomy and stage ontologies

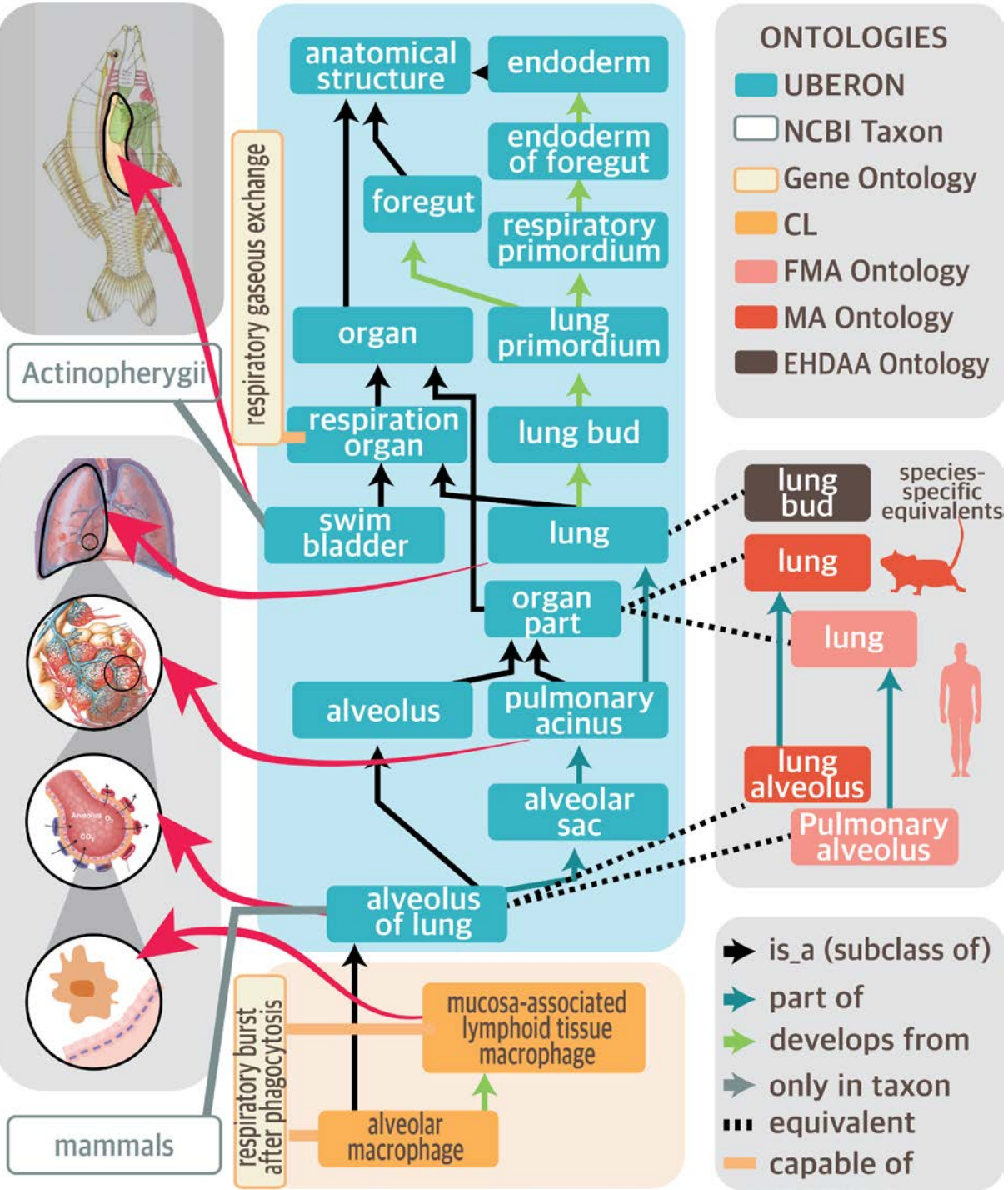
The Zebrafish Anatomy and stage ontologies



The Zebrafish Anatomy and stage ontologies



Uberon: bridging semantics for anatomy

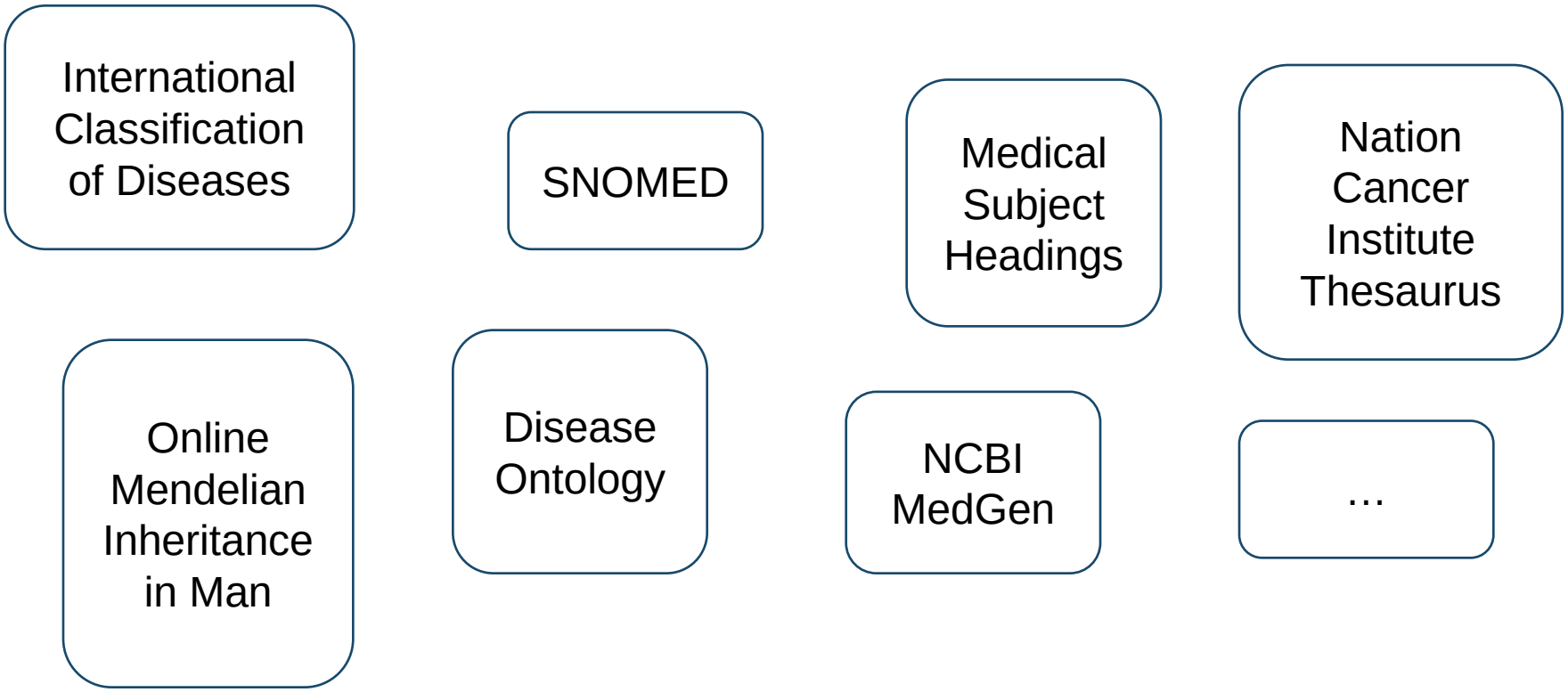


Mungall et al. (2012). **Genome Biology**, 13(1), R5. doi:10.1186/gb-2012-13-1-r5

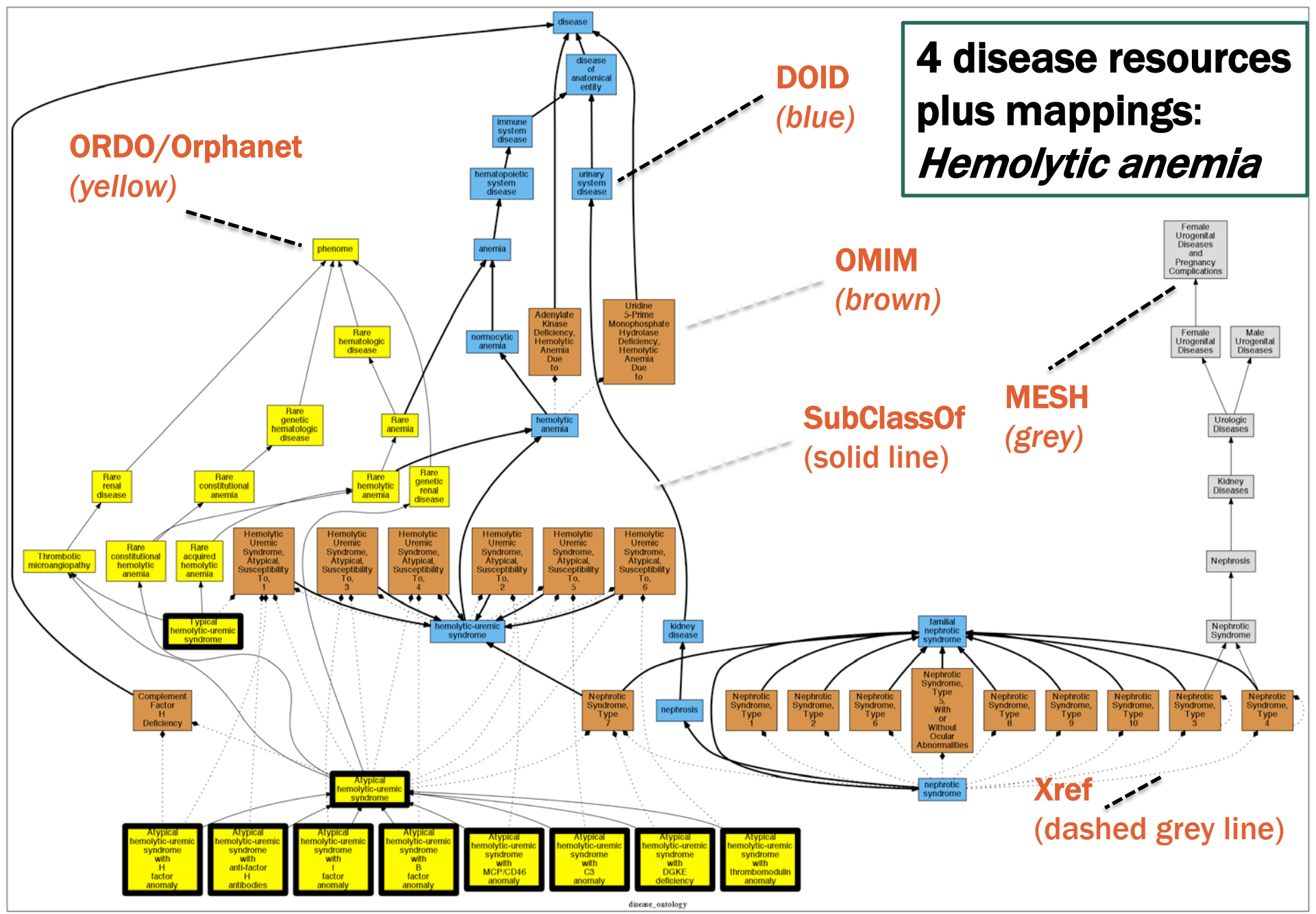
Köhler et al. (2014) **F1000Research** 2:30 Haendel et al. (2014) **JBMS** 5:21 doi:10.1186/2041-1480-5-21

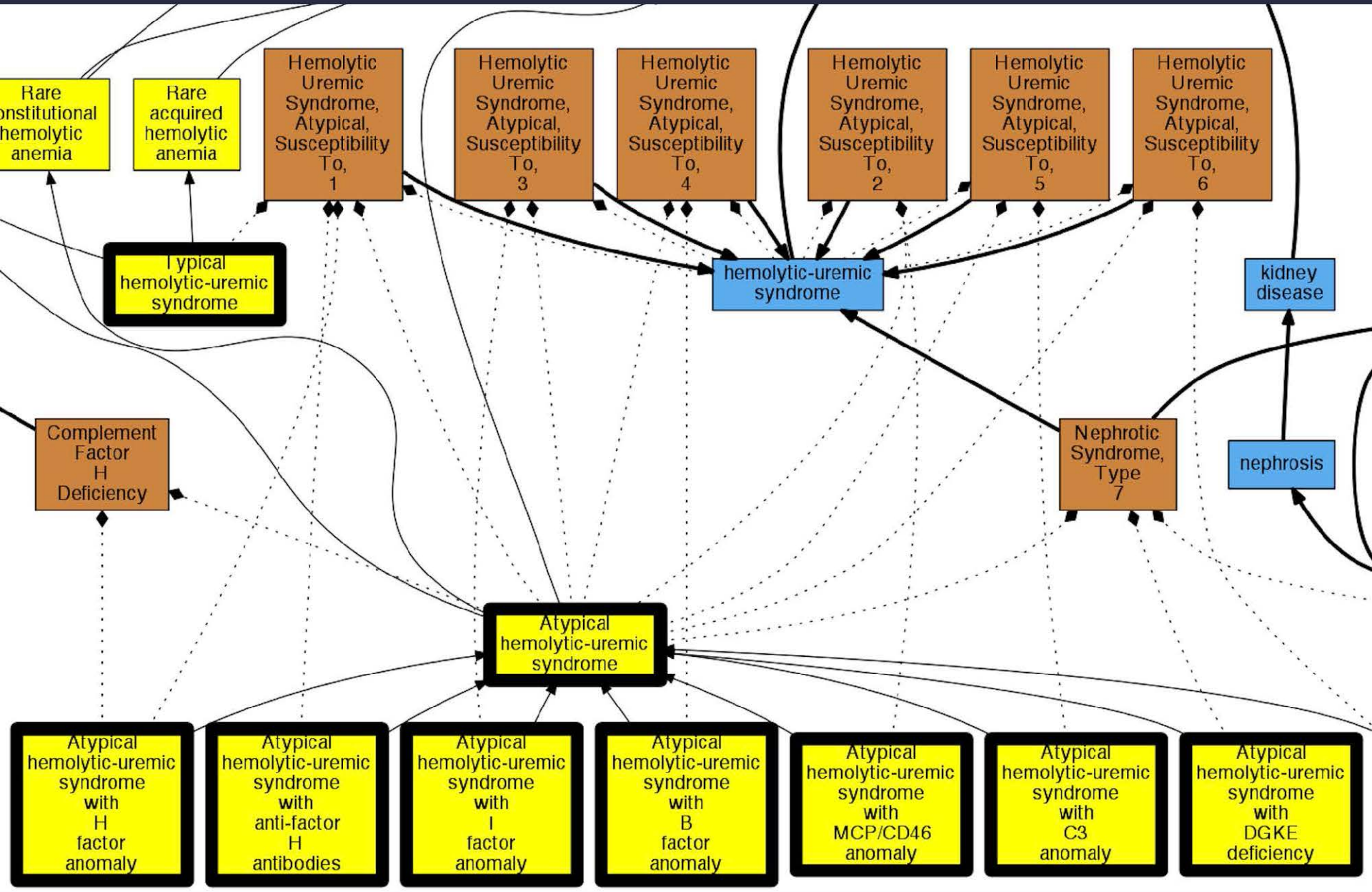
A merger of disease ontologies

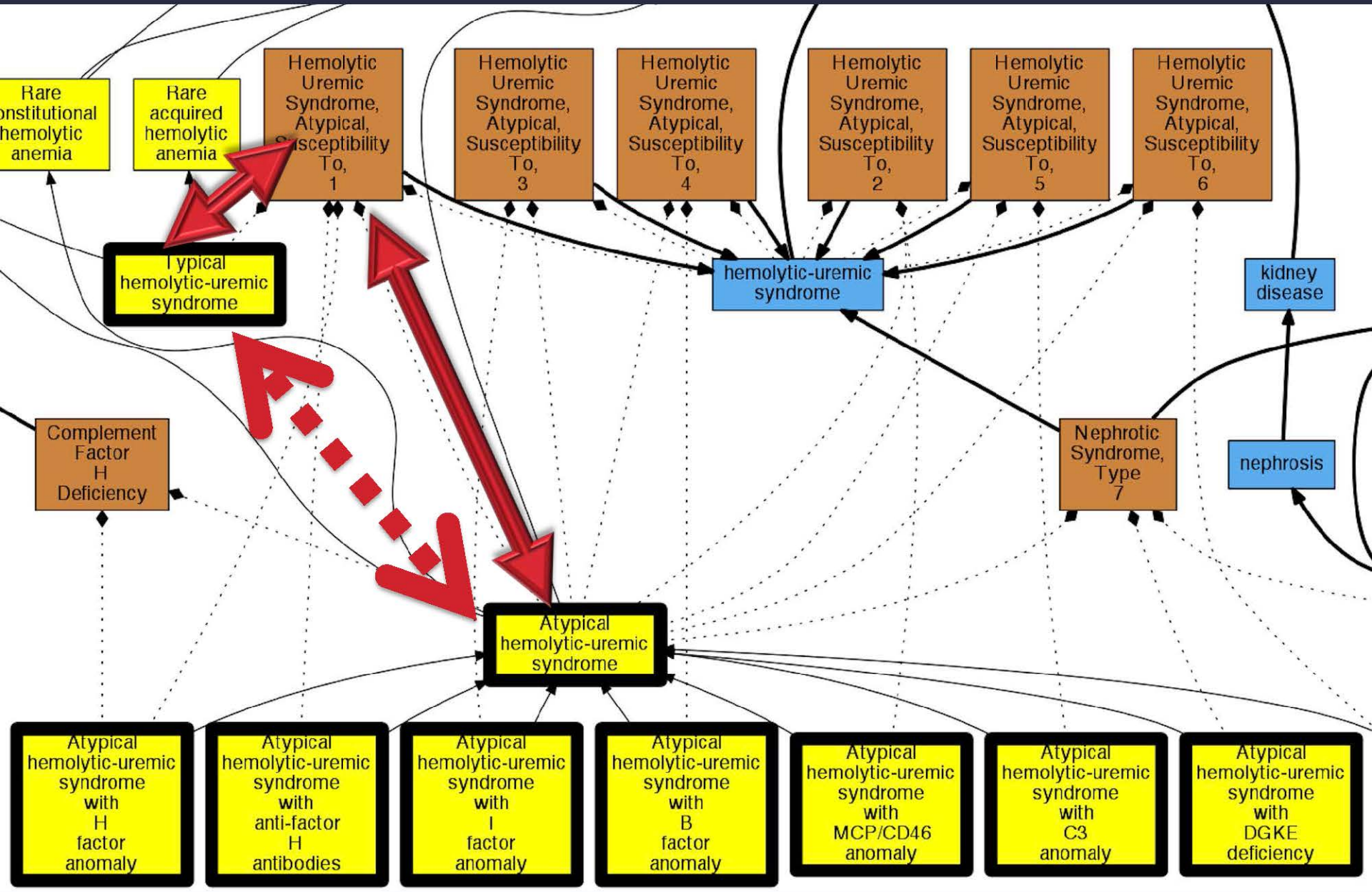
The challenge of multiple perspectives: how can we bridge these?

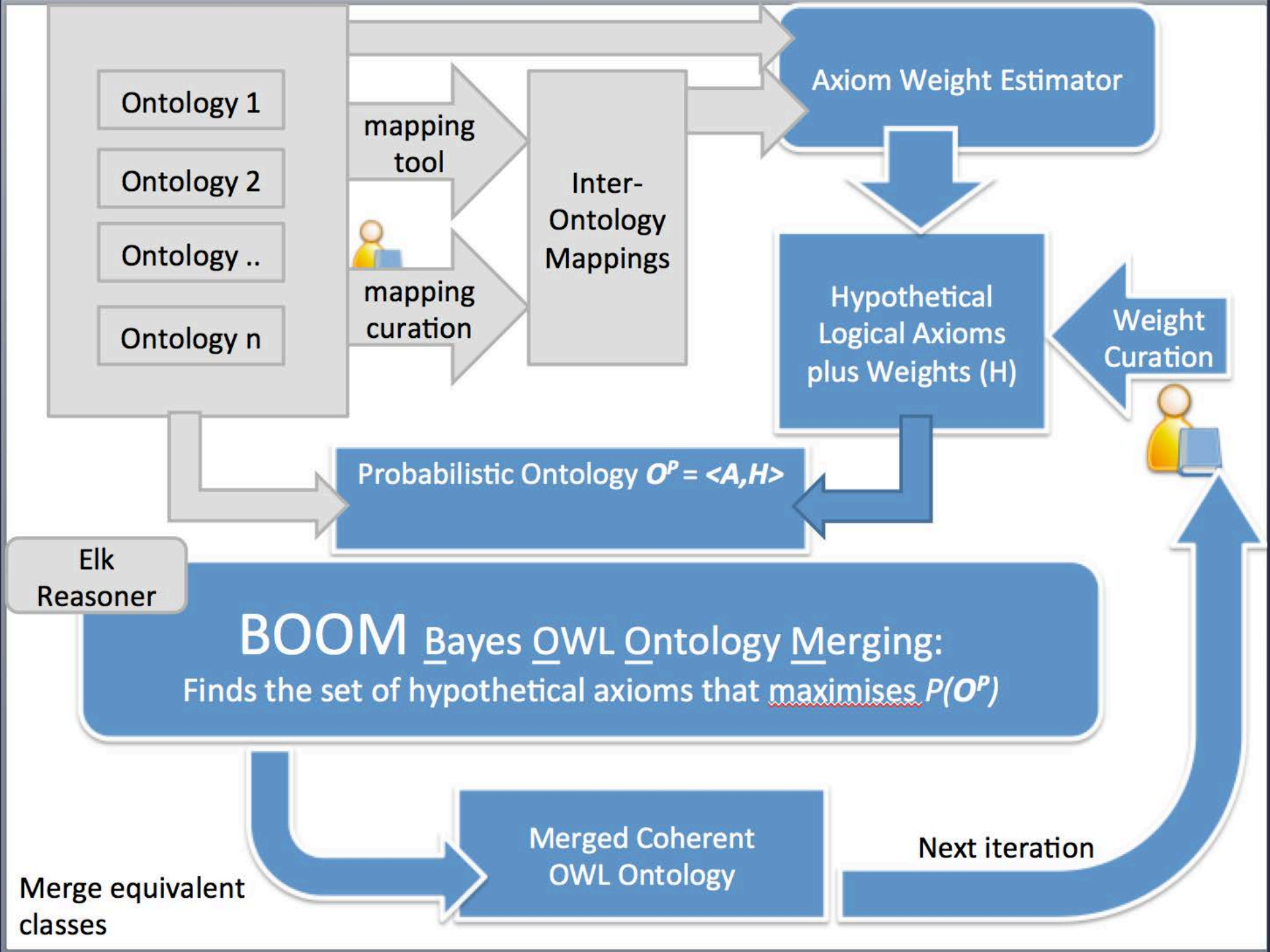


Disease classifications and lists…there are a lot of them


4 disease resources plus mappings: *Hemolytic anemia*
DOID (blue)
ORDO/Orphanet (yellow)
OMIM (brown)
MESH (grey)
SubClassOf (solid line)
Xref (dashed grey line)
disease
disease of anatomical entity
immune system disease
hematopoietic system disease
urinary system disease
phenome
anemia
normocytic anemia
Adenylate Kinase Deficiency, Hemolytic Anemia Due to
Uridine 5-Prime Monophosphate Hydrolase Deficiency, Hemolytic Anemia Due to
Rare hematologic disease
Rare genetic hematologic disease
Rare anemia
hemolytic anemia
Rare renal disease
Rare constitutional anemia
Rare hemolytic anemia
Rare genetic renal disease
Thrombotic microangiopathy
Rare constitutional hemolytic anemia
Rare acquired hemolytic anemia
Hemolytic Uremic Syndrome, Atypical, Susceptibility To, 1
Hemolytic Uremic Syndrome, Atypical, Susceptibility To, 3
Hemolytic Uremic Syndrome, Atypical, Susceptibility To, 4
Hemolytic Uremic Syndrome, Atypical, Susceptibility To, 2
Hemolytic Uremic Syndrome, Atypical, Susceptibility To, 5
Hemolytic Uremic Syndrome, Atypical, Susceptibility To, 6
Typical hemolytic-uremic syndrome
hemolytic-uremic syndrome
kidney disease
Complement Factor H Deficiency
Nephrotic Syndrome, Type 7
nephrosis
familial nephrotic syndrome
Nephrotic Syndrome, Type 1
Nephrotic Syndrome, Type 2
Nephrotic Syndrome, Type 6
Nephrotic Syndrome, Type 5, With or Without Ocular Abnormalities
Nephrotic Syndrome, Type 8
Nephrotic Syndrome, Type 9
Nephrotic Syndrome, Type 10
Nephrotic Syndrome, Type 3
Nephrotic Syndrome, Type 4
nephrotic syndrome
Atypical hemolytic-uremic syndrome
Atypical hemolytic-uremic syndrome with H factor anomaly
Atypical hemolytic-uremic syndrome with anti-factor H antibodies
Atypical hemolytic-uremic syndrome with I factor anomaly
Atypical hemolytic-uremic syndrome with B factor anomaly
Atypical hemolytic-uremic syndrome with MCP/CD46 anomaly
Atypical hemolytic-uremic syndrome with C3 anomaly
Atypical hemolytic-uremic syndrome with DGKE deficiency
Atypical hemolytic-uremic syndrome with thrombomodulin anomaly
Female Urogenital Diseases and Pregnancy Complications
Female Urogenital Diseases
Male Urogenital Diseases
Urologic Diseases
Kidney Diseases
Nephrosis
Nephrotic Syndrome
disease_ontology


Rare
onstitutional
hemolytic
anemia
Rare
acquired
hemolytic
anemia
Hemolytic
Uremic
Syndrome,
Atypical,
Susceptibility
To,
1
Hemolytic
Uremic
Syndrome,
Atypical,
Susceptibility
To,
3
Hemolytic
Uremic
Syndrome,
Atypical,
Susceptibility
To,
4
Hemolytic
Uremic
Syndrome,
Atypical,
Susceptibility
To,
2
Hemolytic
Uremic
Syndrome,
Atypical,
Susceptibility
To,
5
Hemolytic
Uremic
Syndrome,
Atypical,
Susceptibility
To,
6
Typical
hemolytic-uremic
syndrome
hemolytic-uremic
syndrome
kidney
disease
Complement
Factor
H
Deficiency
Nephrotic
Syndrome,
Type
7
nephrosis
Atypical
hemolytic-uremic
syndrome
Atypical
hemolytic-uremic
syndrome
with
H
factor
anomaly
Atypical
hemolytic-uremic
syndrome
with
anti-factor
H
antibodies
Atypical
hemolytic-uremic
syndrome
with
I
factor
anomaly
Atypical
hemolytic-uremic
syndrome
with
B
factor
anomaly
Atypical
hemolytic-uremic
syndrome
with
MCP/CD46
anomaly
Atypical
hemolytic-uremic
syndrome
with
C3
anomaly
Atypical
hemolytic-uremic
syndrome
with
DGKE
deficiency


Rare
onstitutional
hemolytic
anemia
Rare
acquired
hemolytic
anemia
Hemolytic
Uremic
Syndrome,
Atypical,
susceptibility
To,
1
Hemolytic
Uremic
Syndrome,
Atypical,
Susceptibility
To,
3
Hemolytic
Uremic
Syndrome,
Atypical,
Susceptibility
To,
4
Hemolytic
Uremic
Syndrome,
Atypical,
Susceptibility
To,
2
Hemolytic
Uremic
Syndrome,
Atypical,
Susceptibility
To,
5
Hemolytic
Uremic
Syndrome,
Atypical,
Susceptibility
To,
6
Typical
hemolytic-uremic
syndrome
hemolytic-uremic
syndrome
kidney
disease
Complement
Factor
H
Deficiency
Nephrotic
Syndrome,
Type
7
nephrosis
Atypical
hemolytic-uremic
syndrome
Atypical
hemolytic-uremic
syndrome
with
H
factor
anomaly
Atypical
hemolytic-uremic
syndrome
with
anti-factor
H
antibodies
Atypical
hemolytic-uremic
syndrome
with
I
factor
anomaly
Atypical
hemolytic-uremic
syndrome
with
B
factor
anomaly
Atypical
hemolytic-uremic
syndrome
with
MCP/CD46
anomaly
Atypical
hemolytic-uremic
syndrome
with
C3
anomaly
Atypical
hemolytic-uremic
syndrome
with
DGKE
deficiency


Ontology 1
Ontology 2
Ontology ..
Ontology n
mapping tool
mapping curation
Inter-Ontology Mappings
Axiom Weight Estimator
Hypothetical Logical Axioms plus Weights (H)
Weight Curation
Probabilistic Ontology O^P = <A,H>
Elk Reasoner
BOOM Bayes OWL Ontology Merging:
Finds the set of hypothetical axioms that maximises $P(O^P)$
Merged Coherent OWL Ontology
Next iteration
Merge equivalent classes

MonDO: Merged Ontology of Disease Entities

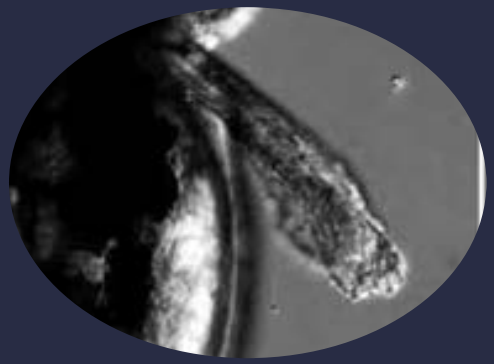
"Ontology"	Classes (before, after merge)	SubClass axioms	Mappings
Inputs:			
DOID	6878 → 6012	7082	36656
MESH (D)	11314 → 4152	19036	
OMIM (D)	7783 → 7783	0	31242
Orphanet (D)	8740 → 4683	15182	20326
OMIA	4833 → 4833	3120	355
DC	209 → 208	310	316
Medic	0	8630	3435
Output:			
MonDO	39757 → 27617	44837	

https://github.com/monarch-initiative/monarch-disease-ontology

Phenotype ontologies

Different communities use different languages

Palmoplantar hyperkeratosis

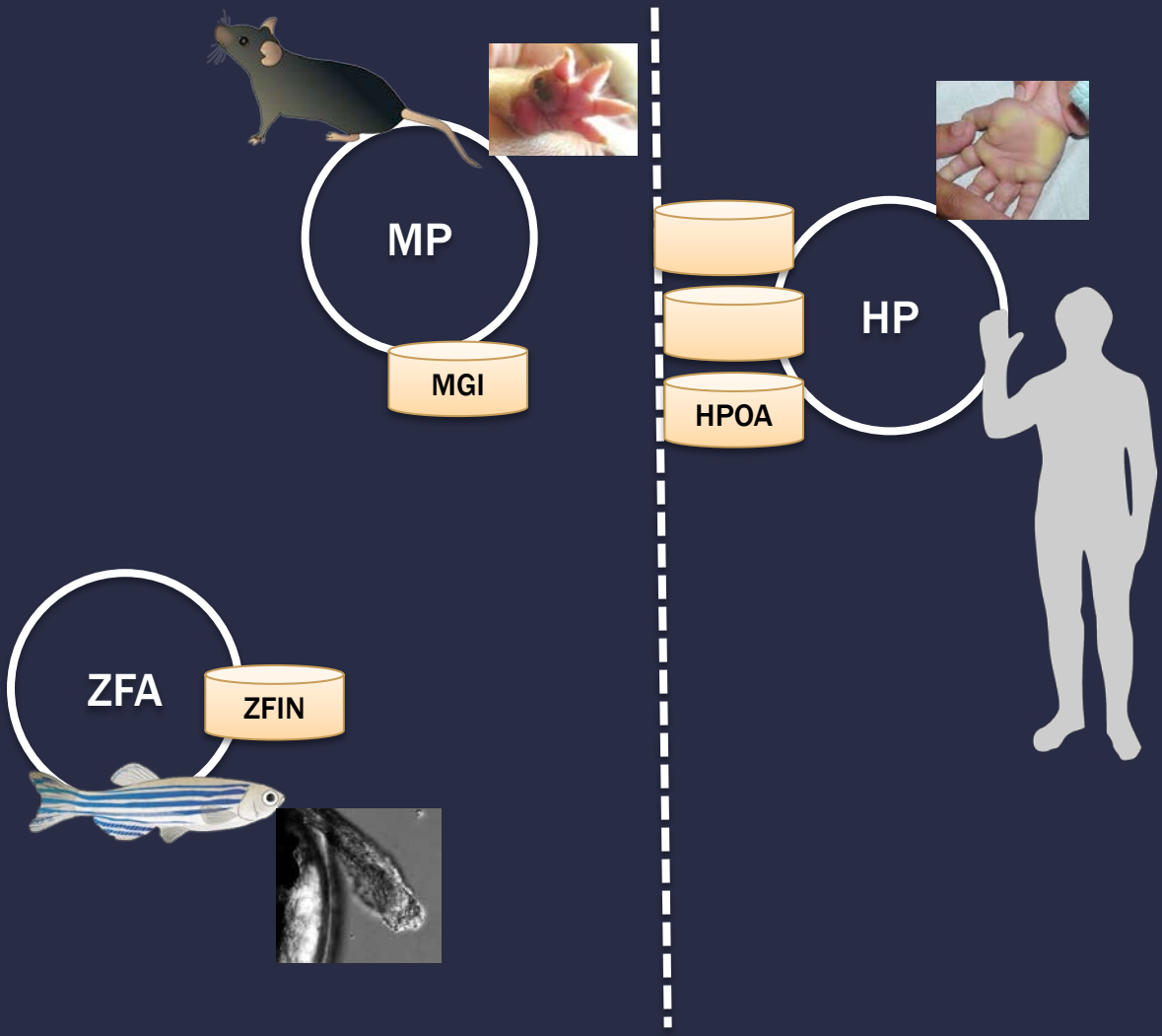


Degenerate fin epithelium

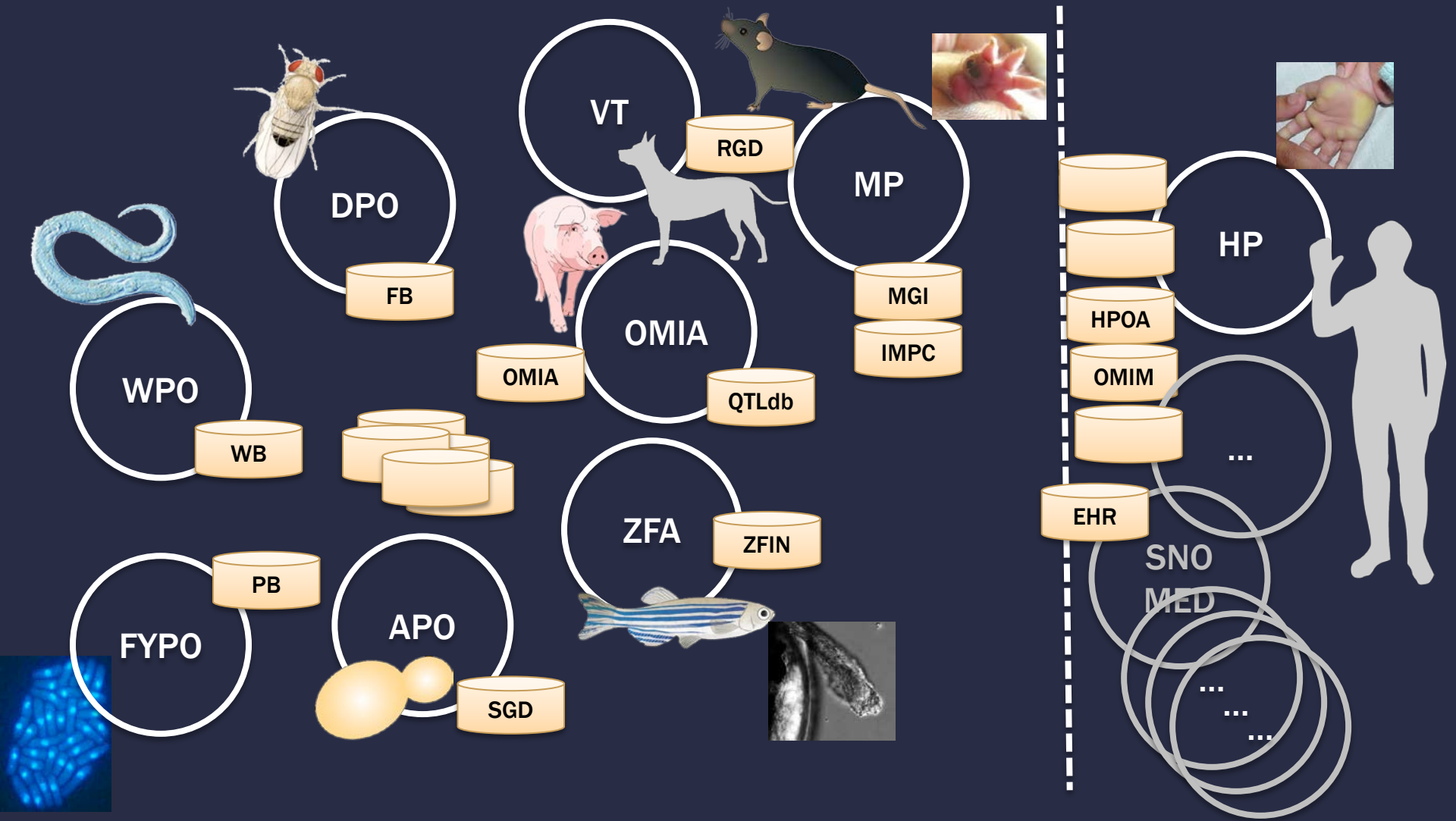
Thick hand skin

Ulcerated paws

Challenge: Each data source uses their own vocabulary/ontology



Challenge: Each data source uses their own phenotype vocabulary/ontology



Decomposition of complex concepts allows interoperability

Human phenotype

"Palmoplantar hyperkeratosis"

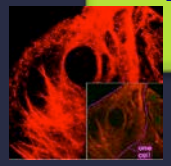
=

PATO

increased

GO

keratinization



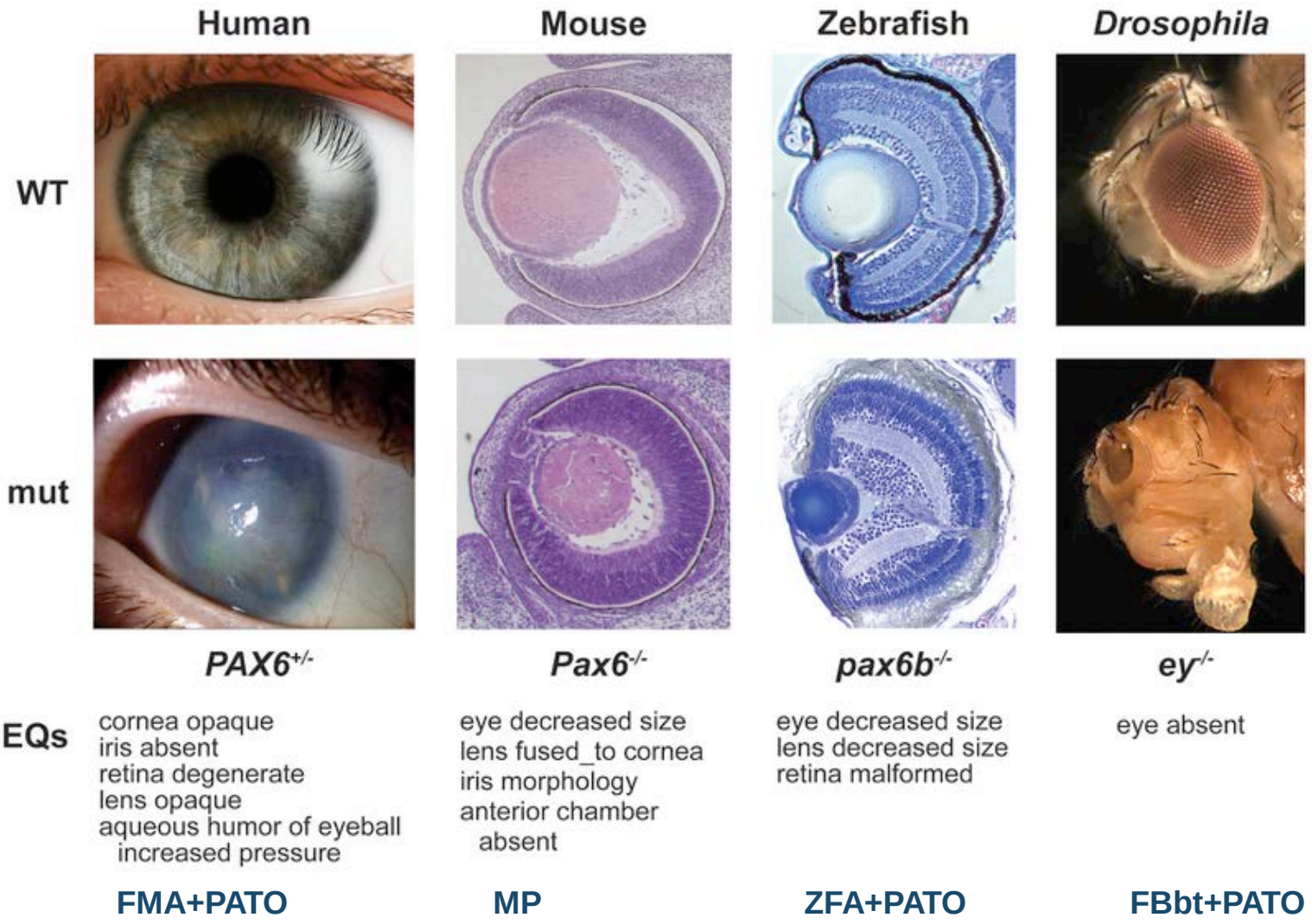
Uberon

Stratum corneum layer of skin

Autopod

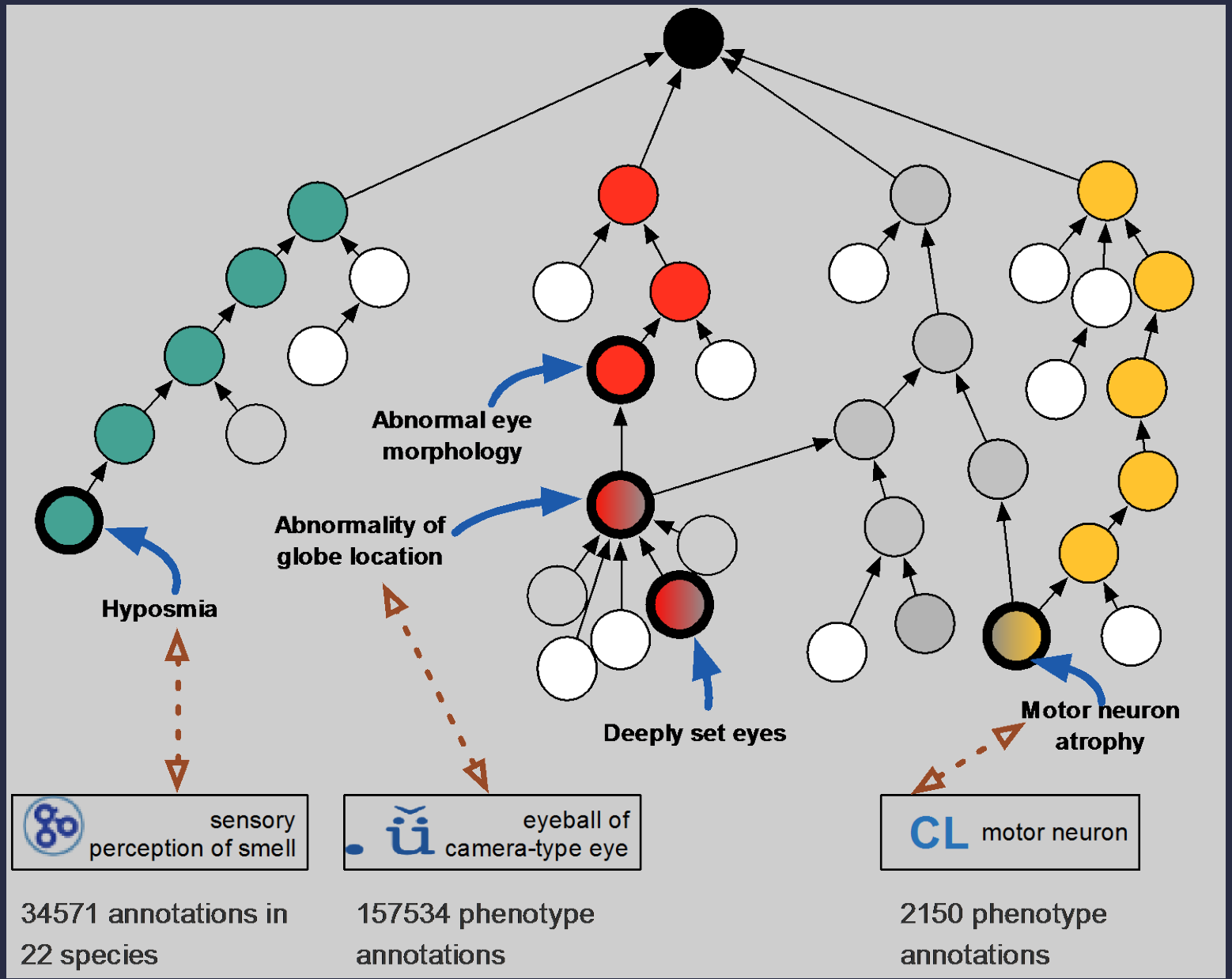
Species neutral ontologies, homologous concepts

Semantic similarity of phenotypes for disease discovery

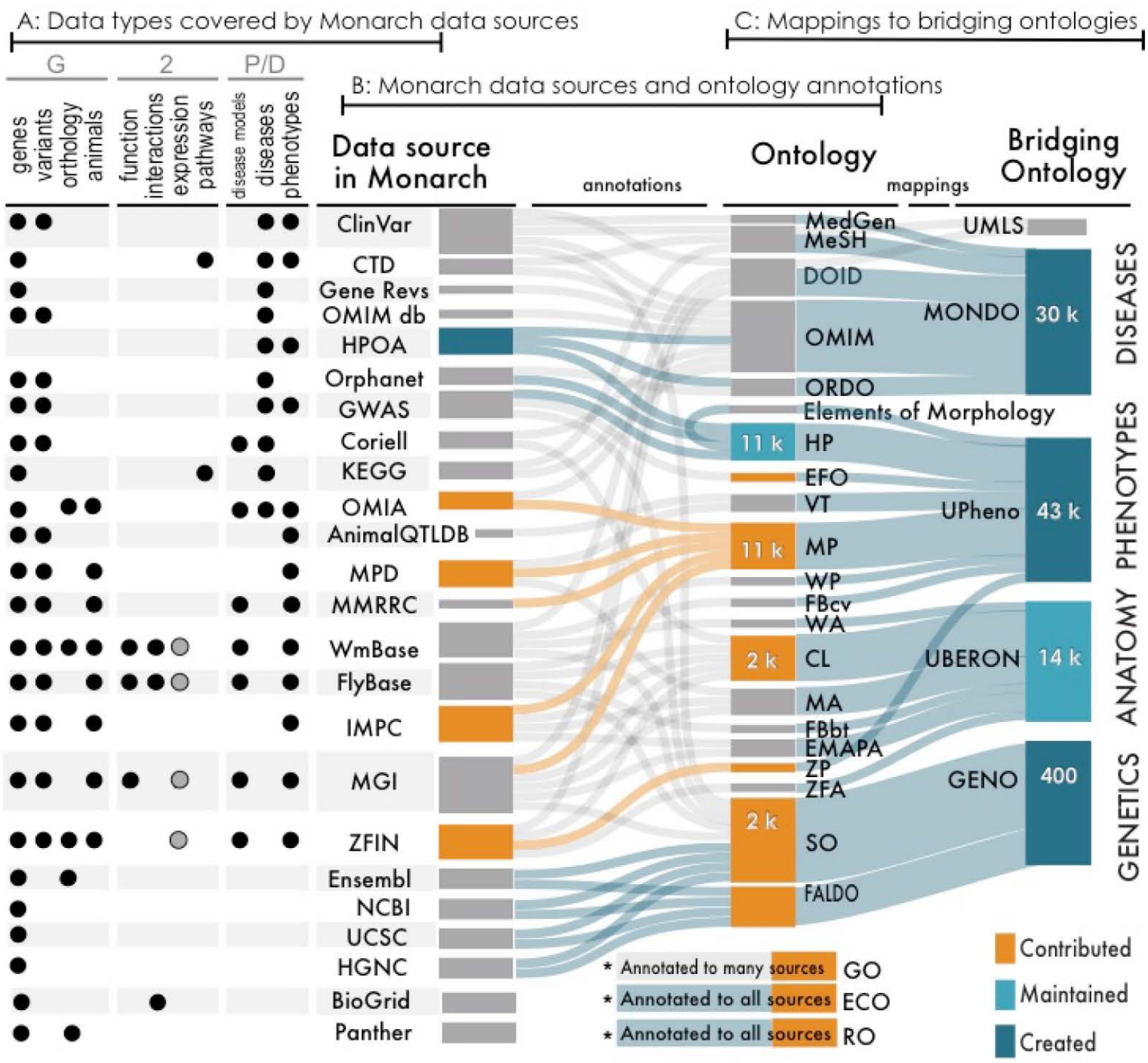


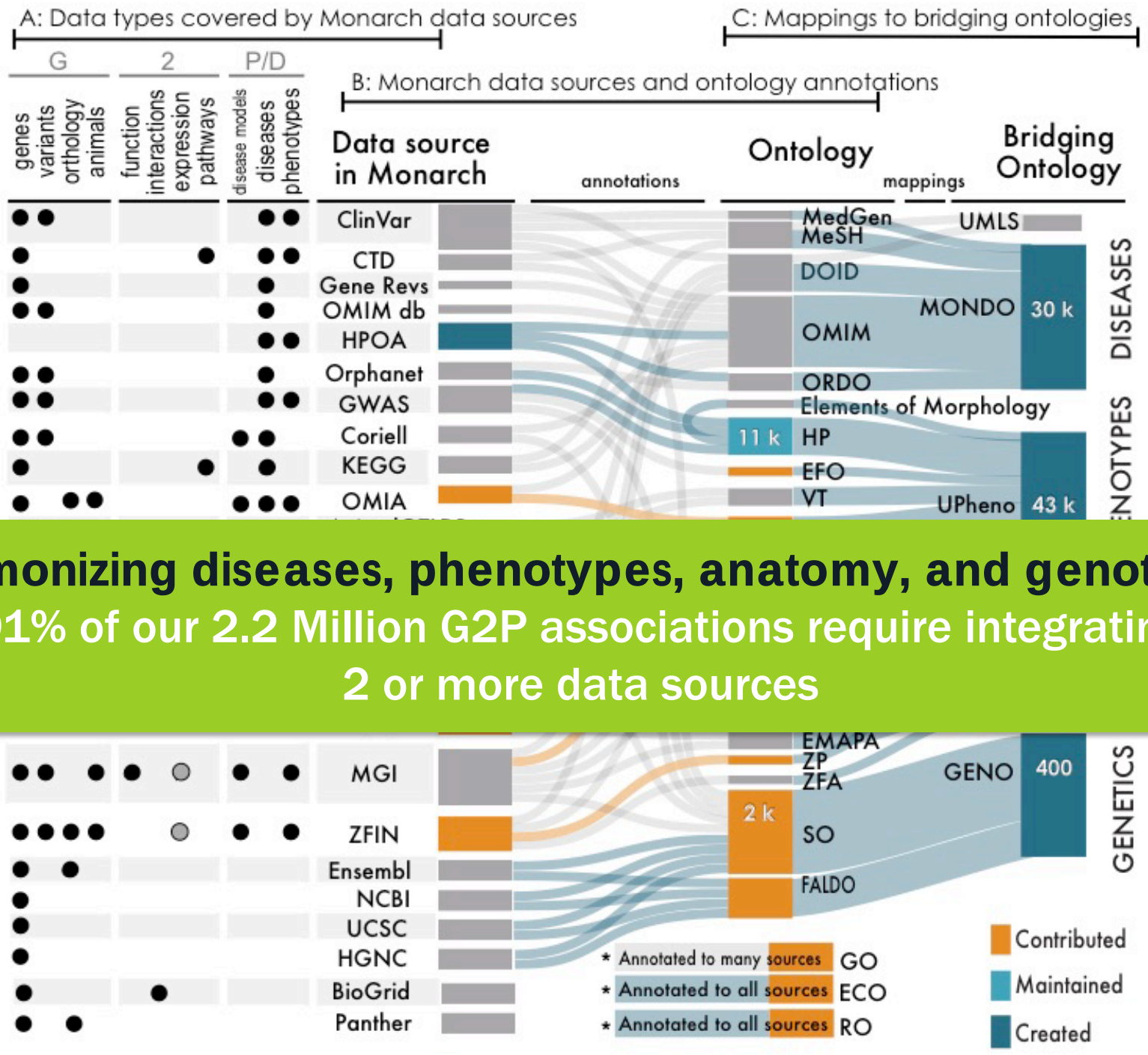
"Linking Human Diseases to Animal Models Using Ontology-Based Phenotype Annotation." PLoS Biol 7(11): e1000247. doi:10.1371/journal.pbio.1000247 Washington NL, Haendel MA, Mungall CJ, Ashburner M, Westerfield M, Lewis SE

The Human Phenotype Ontology for deep phenotyping



Ontologies at work: Data integration and disease diagnosis

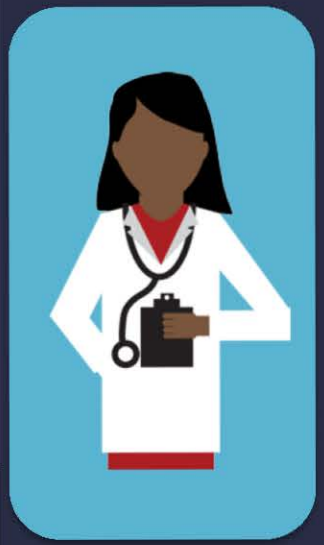

A: Data types covered by Monarch data sources
G
2
P/D
genes
variants
orthology
animals
function
interactions
expression
pathways
disease models
diseases
phenotypes
B: Monarch data sources and ontology annotations
Data source in Monarch
annotations
Ontology
mappings
Bridging Ontology
C: Mappings to bridging ontologies
ClinVar
CTD
Gene Revs
OMIM db
HPOA
Orphanet
GWAS
Coriell
KEGG
OMIA
AnimalQTLDB
MPD
MMRRC
WmBase
FlyBase
IMPC
MGI
ZFIN
Ensembl
NCBI
UCSC
HGNC
BioGrid
Panther
MedGen
MeSH
DOID
OMIM
ORDO
Elements of Morphology
11 k
HP
EFO
VT
11 k
MP
WP
FBcv
WA
2 k
CL
MA
FBbt
EMAPA
ZP
ZFA
2 k
SO
FALDO
UMLS
MONDO
30 k
UPheno
43 k
UBERON
14 k
GENO
400
DISEASES
PHENOTYPES
ANATOMY
GENETICS
* Annotated to many sources GO
* Annotated to all sources ECO
* Annotated to all sources RO
Contributed
Maintained
Created

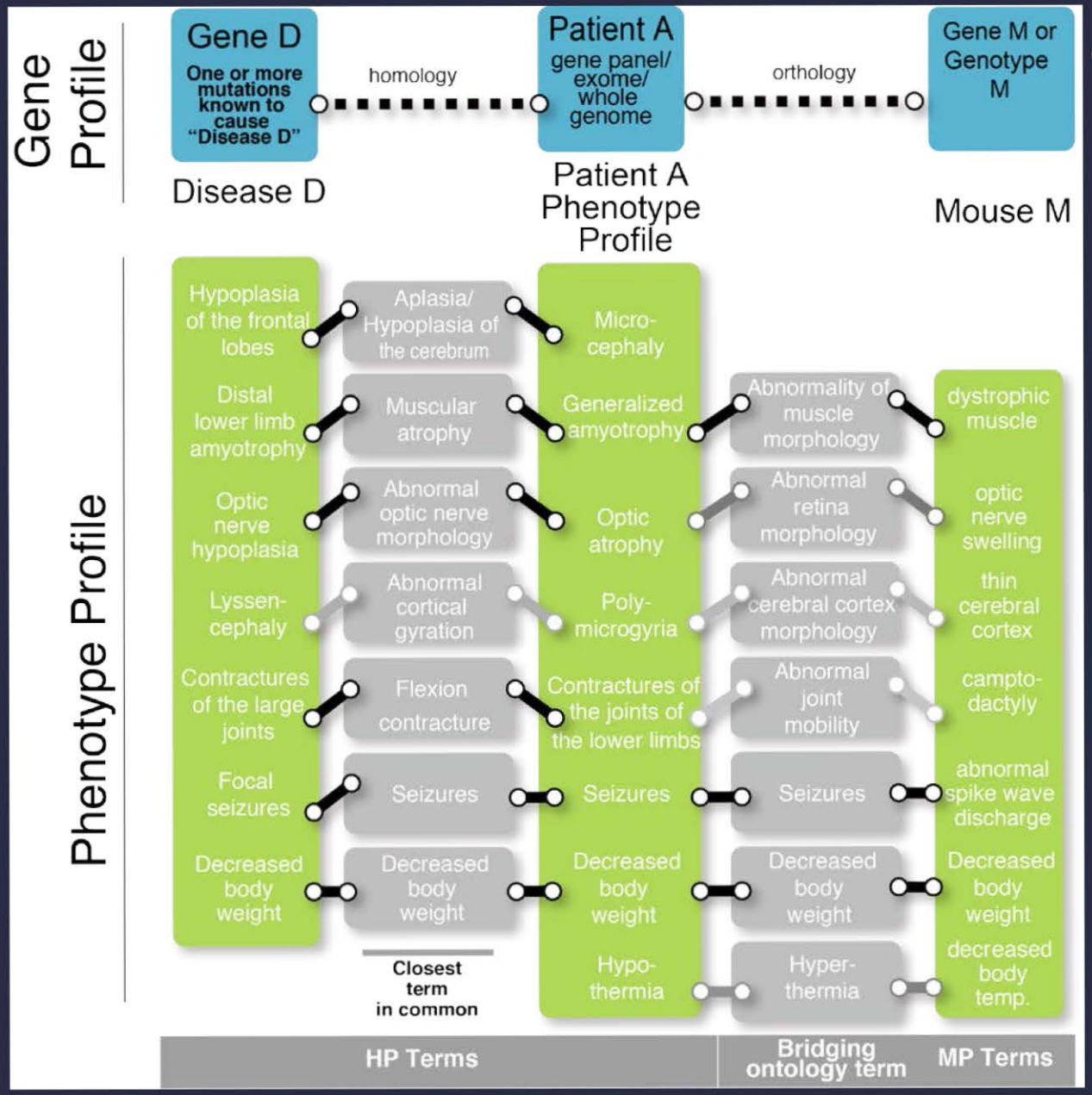


Harmonizing diseases, phenotypes, anatomy, and genotypes

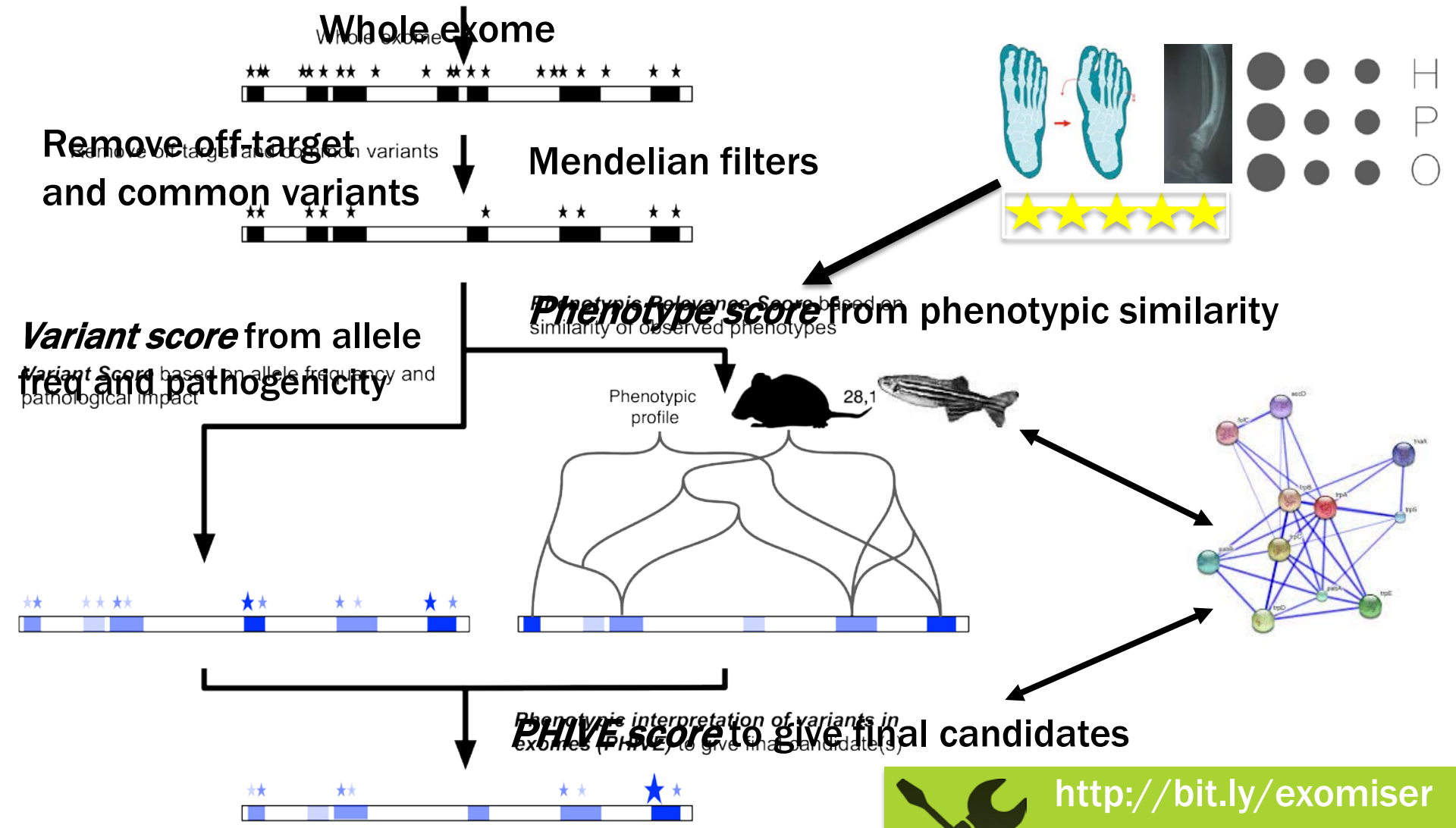
91% of our 2.2 Million G2P associations require integrating 2 or more data sources

Phenotypic matchmaking for disease diagnostics





Combining genotype and phenotype data for variant prioritization



Putting all that data to use to diagnose a rare platelet syndrome

http://bit.ly/stim1paper

EXOMISER

http://bit.ly/exomiser

Ranked STIM-1 variant maximally pathogenic based on cross-species G2P data, in the absence of traditional data sources

monarch INITIATIVE

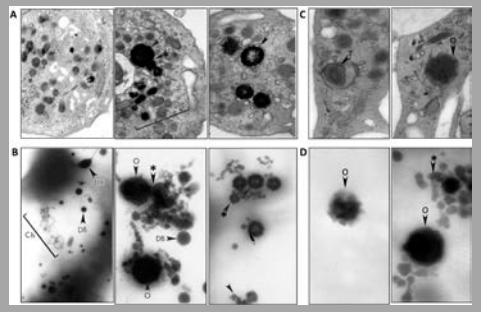
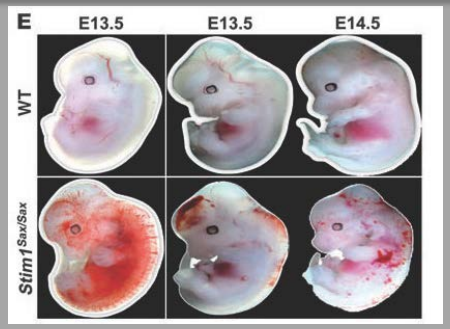
NIH Undiagnosed Diseases Program

MGI

Stim1$^{Sax/Sax}$

OMIM

CAGGC TGAGG

Phenotypic profile				N/A
Genes	Heterozygous, missense mutation STIM-1	Heterozygous, missense mutation STIM-1	N/A	

What about environment and exposure ontologies?

“the environment is everything that isn’t me”

—Albert Einstein

Can we sensibly make an ontology of everything that isn’t me?

1.	**Occupational diseases caused by exposure to agents arising from work activities**
1.1.	**Diseases caused by chemical agents**
1.1.1.	Diseases caused by beryllium or its compounds
1.1.2.	Diseases caused by cadmium or its compounds
1.1.3.	Diseases caused by phosphorus or its compounds
1.1.4.	Diseases caused by chromium or its compounds
1.1.5.	Diseases caused by manganese or its compounds
1.1.6.	Diseases caused by arsenic or its compounds
1.1.7.	Diseases caused by mercury or its compounds
1.1.8.	Diseases caused by lead or its compounds
1.1.9.	Diseases caused by fluorine or its compounds
1.1.10.	Diseases caused by carbon disulfide
1.1.11.	Diseases caused by halogen derivatives of aliphatic or aromatic hydrocarbons
1.1.12.	Diseases caused by benzene or its homologues
1.1.13.	Diseases caused by nitro- and amino-derivatives of benzene or its homologues
1.1.14.	Diseases caused by nitroglycerine or other nitric acid esters
1.1.15.	Diseases caused by alcohols, glycols or ketones
1.1.16.	Diseases caused by asphyxiants like carbon monoxide, hydrogen sulfide, hydrogen cyanide or its derivatives
1.1.17.	Diseases caused by acrylonitrile
1.1.18.	Diseases caused by oxides of nitrogen
1.1.19.	Diseases caused by vanadium or its compounds
1.1.20.	Diseases caused by antimony or its compounds

3. Occupational cancer

3.1. Cancer caused by the following agents

3.1.1. Asbestos

3.1.2. Benzid

3.1.3. Bis-chl

3.1.4. Chromi

3.1.5. Coal ta

3.1.6. Beta-na

3.1.7. Vinyl ch

3.1.8. Benzen

3.1.9. Toxic n

3.1.10. Ionizing

3.1.11. Tar, pit … residues of these s

3.1.12. Coke o

3.1.13. Nickel

3.1.14. Wood d

3.1.15. Arsenic

3.1.16. Berylliu

3.1.17. Cadmiu

3.1.18. Erionite

3.1.19. Ethylene oxide

3.1.20. Hepatitis B virus (HBV) and hepatitis C virus (HCV)

3.1.21. Cancers caused by other agents at work not mentioned in the preceding items where a

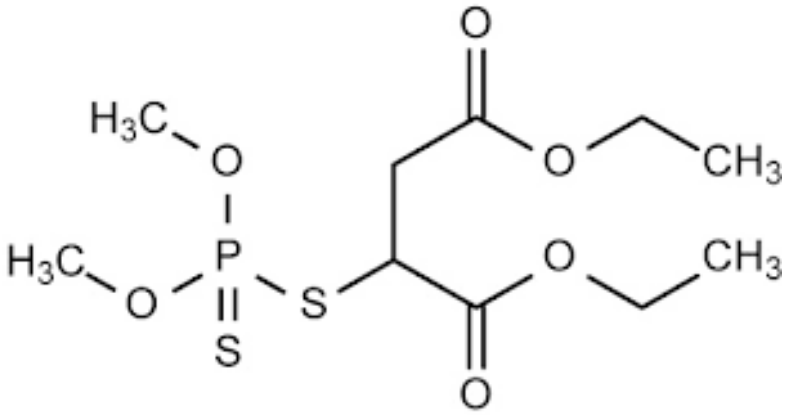
Can we make these lists computable?

Translate them into a form a machine can understand and reason over?

International Labour Organization

We have a precise machine-readable language for describing some environmental exposures

= CCOC(=O)CC(SP(=S)(OC)OC)C(=O)OCC

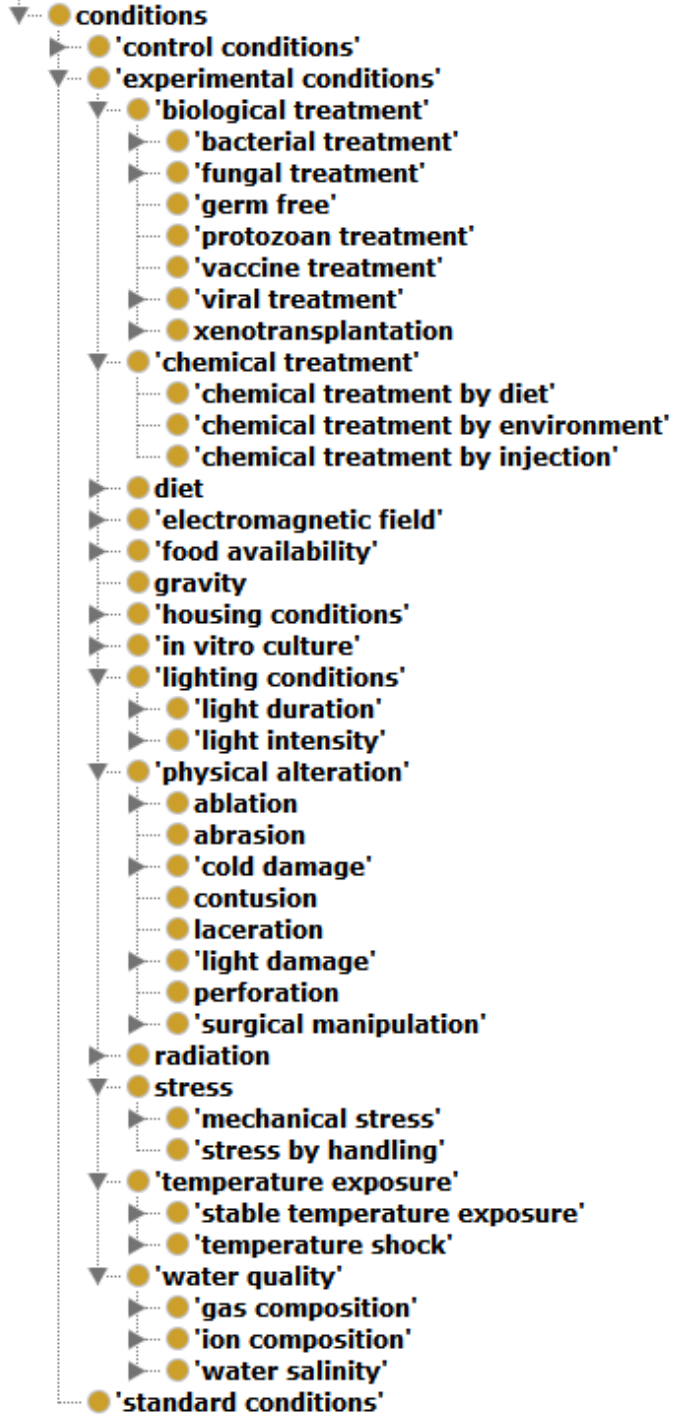


CHEBI:6651

CheBI is a chemical ontology

But others are harder to define

Image: Zol87 CC by/nc



The Zebrafish Environmental Conditions Ontology

https://github.com/ybradford/zebrafish-experimental-conditions-ontology

The Environment Ontology

- Originally created for metagenome samples
  - Characterize microbial environments
- Extended for ecological science
  - The “Earth Phenotype Ontology”
- Being adapted for human exposures

Material
- Water
- Soil
- Air

Features
- Natural
- Anthropogenic

Biome
- Terrestrial
- Aquatic
- Polar

Process
- Erosion
- Pollution
- Biological
  - Algal bloom…

Buttigieg, P. L., Morrison, N., Smith, B., Mungall, C. J., & Lewis, S. E. (2013). The environment ontology: contextualising biological and biomedical entities. Journal of Biomedical Semantics, 4(1), 43. doi:10.1186/2041-1480-4-43

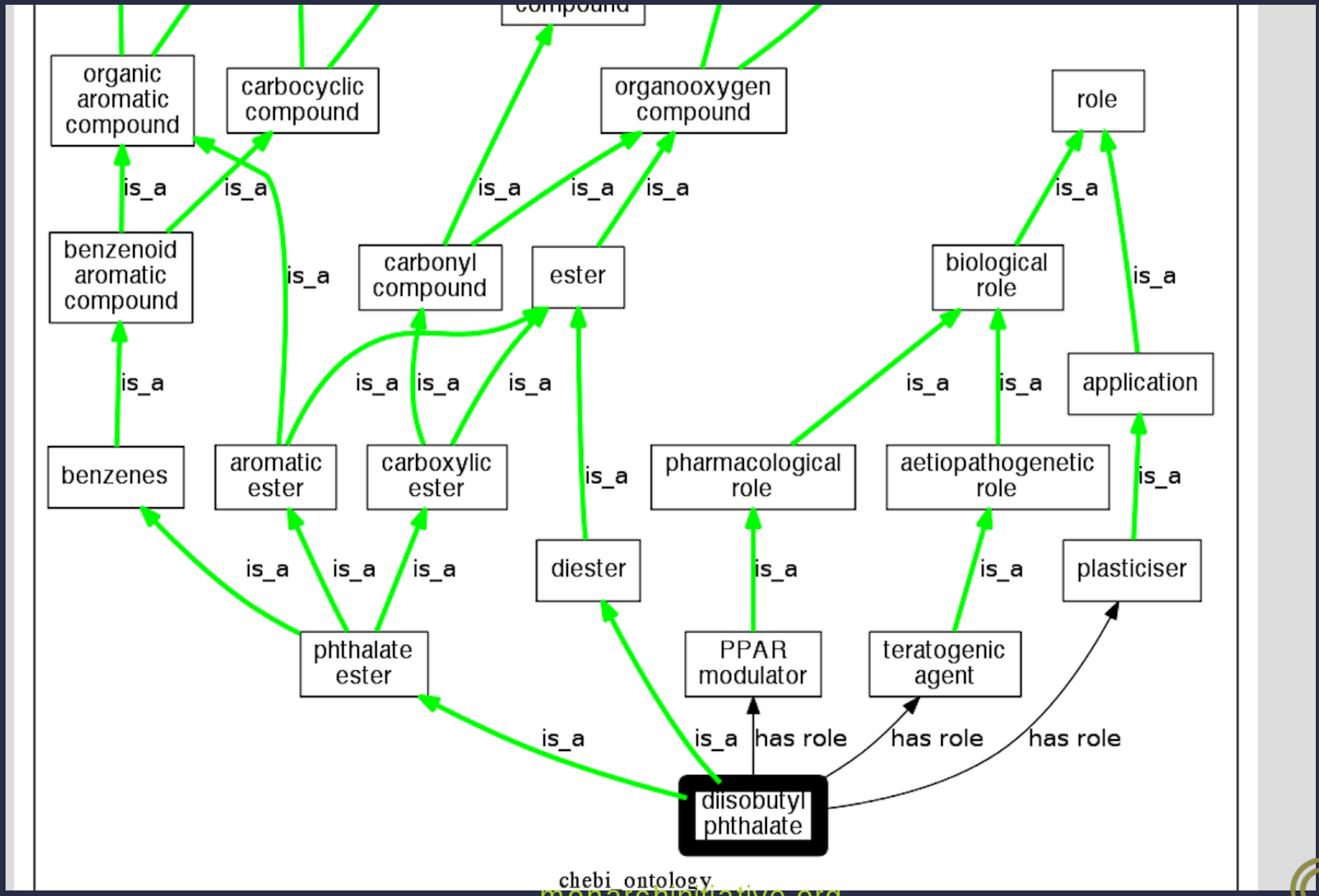
Biome: Food desert

Feature: Store (alcohol, sugar-rich food)

Material: Air, high particulate matter

Process: decreased investment in infrastructure

CHEBI: chemical classification



Environmental conditions, treatments and exposures ontology (ECTO)

PECO
Pombe experimental conditions ontology

ExO
Exposure ontology

NCI Thesaurus
clinical

ENVO

SNOMED
clinical

ZECO:
Zebrafish environmental conditions

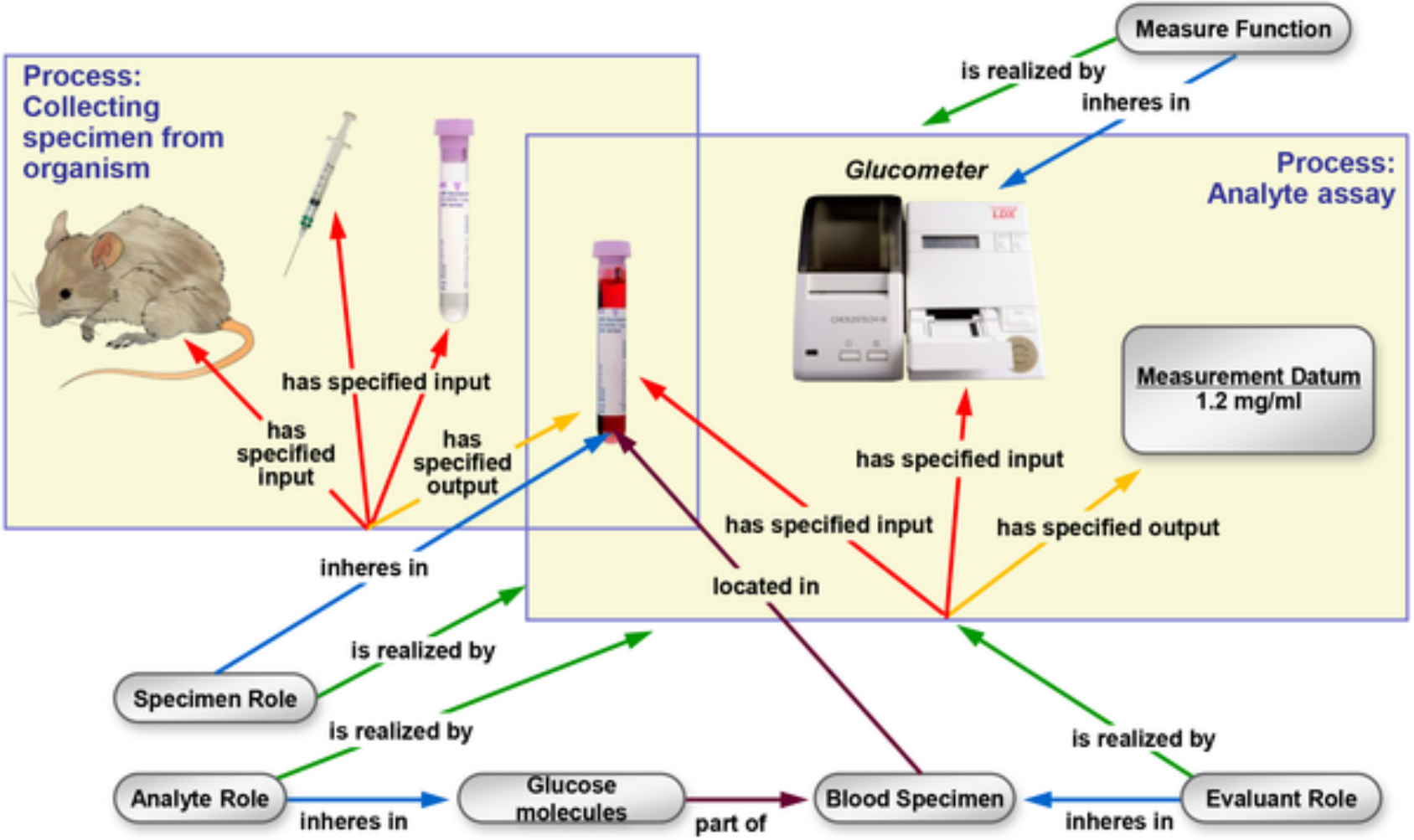
XCO
Experimental conditions ontology

MRE
Medically relevant exposures

https://github.com/cmungall/environmental-conditions

monarchinitiative.org

The Ontology of Biomedical Investigations



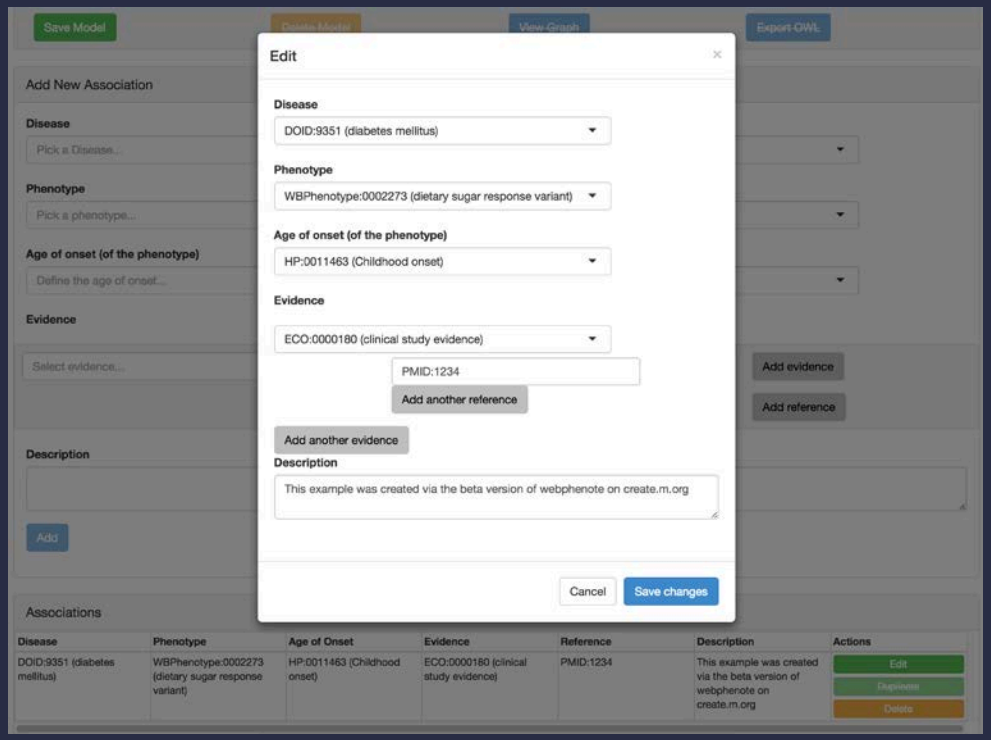
(2016) The Ontology for Biomedical Investigations. PLOS ONE 11(4): e0154556. doi:10.1371/journal.pone.0154556
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0154556

Recording and exchanging phenotype and environmental data…better

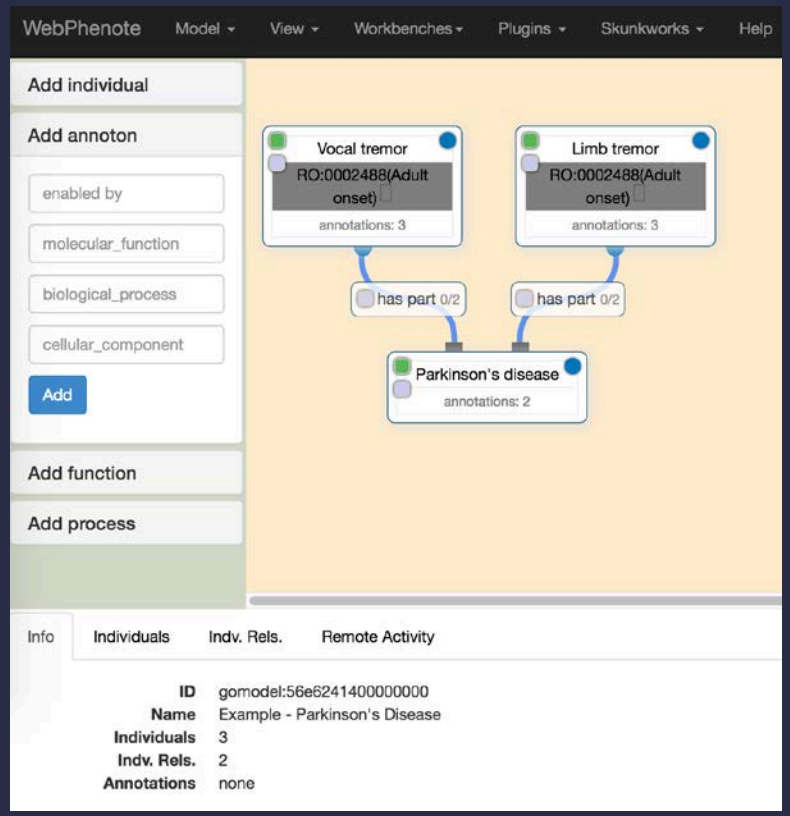
WebPhenote and Noctua

A causal/spatiotemporal network curation environment

Form-based



Graph-based



http://create.monarchinitiative.org/

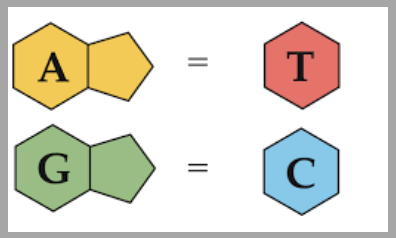
noctua.berkeleybop.org

GENEONTOLOGY
Unifying Biology

Computable encodings are essential

Genes + Environment = Phenotypes

Base pairs
Variant notation (eg. HGVS)



HGVS
HUMAN GENOME VARIATION SOCIETY

Medical procedure coding

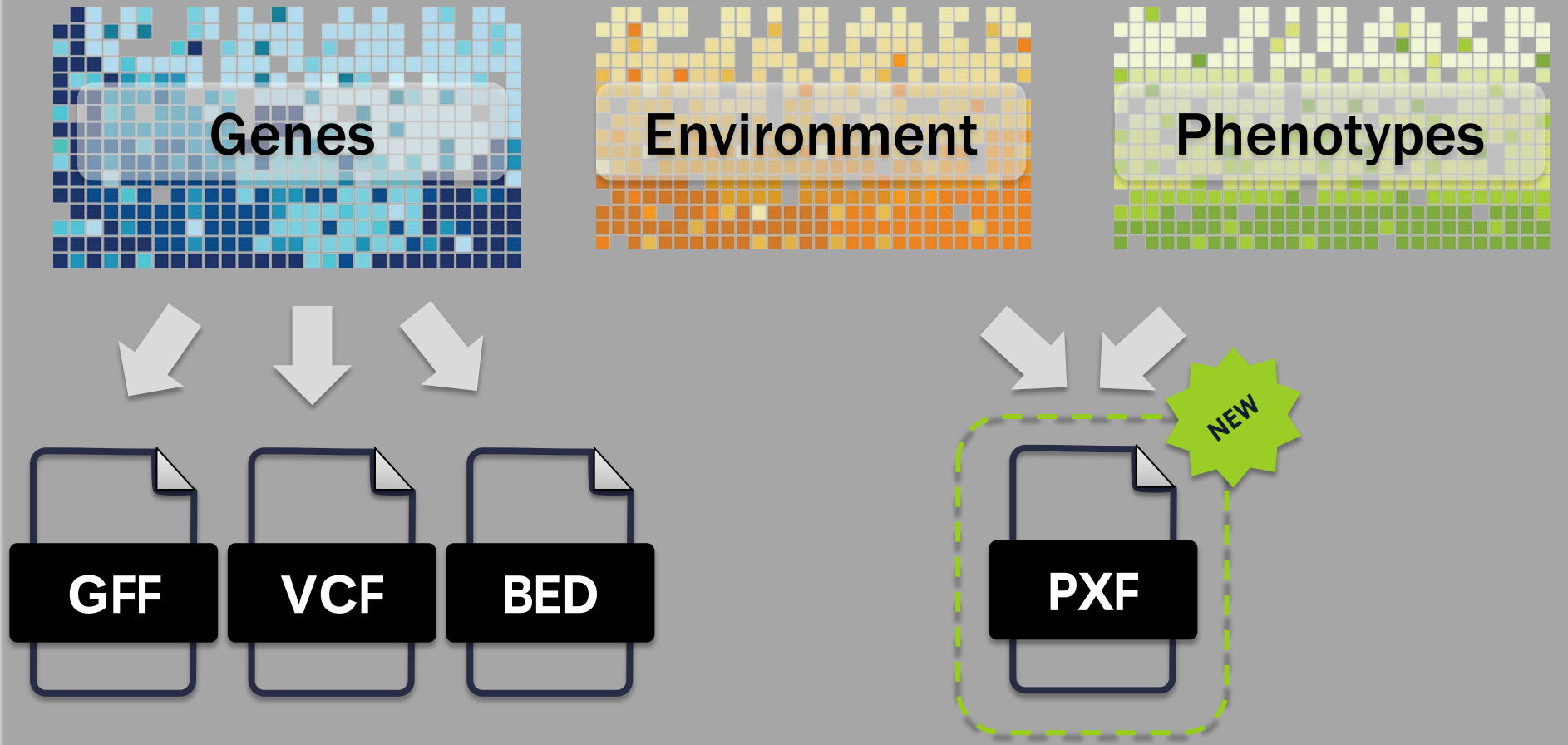
ICD-10

SNOMED INTERNATIONAL

Human Phenotype Ontology

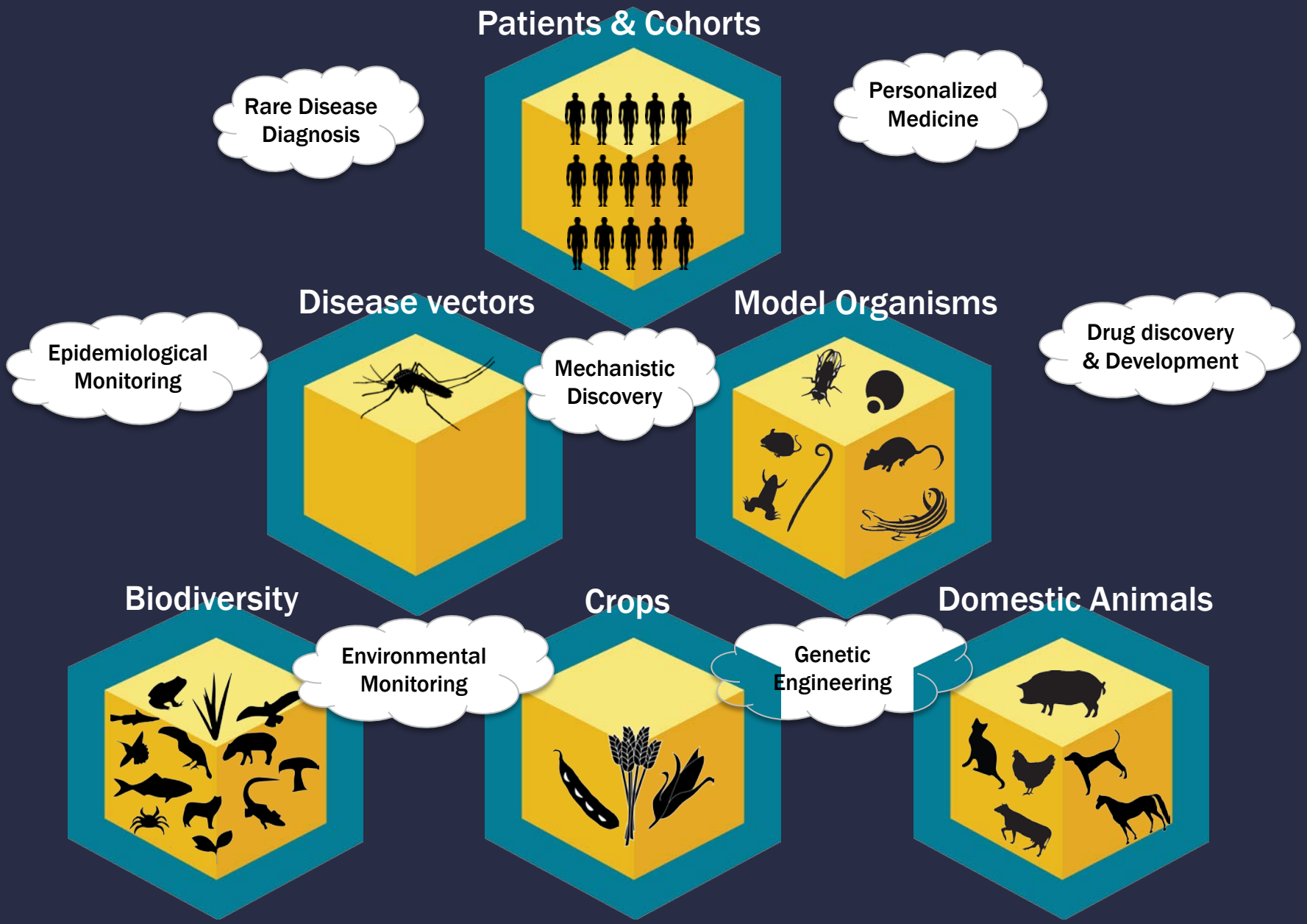
PheKB

Standard exchange formats exist for genes … but for phenotypes? Environment?



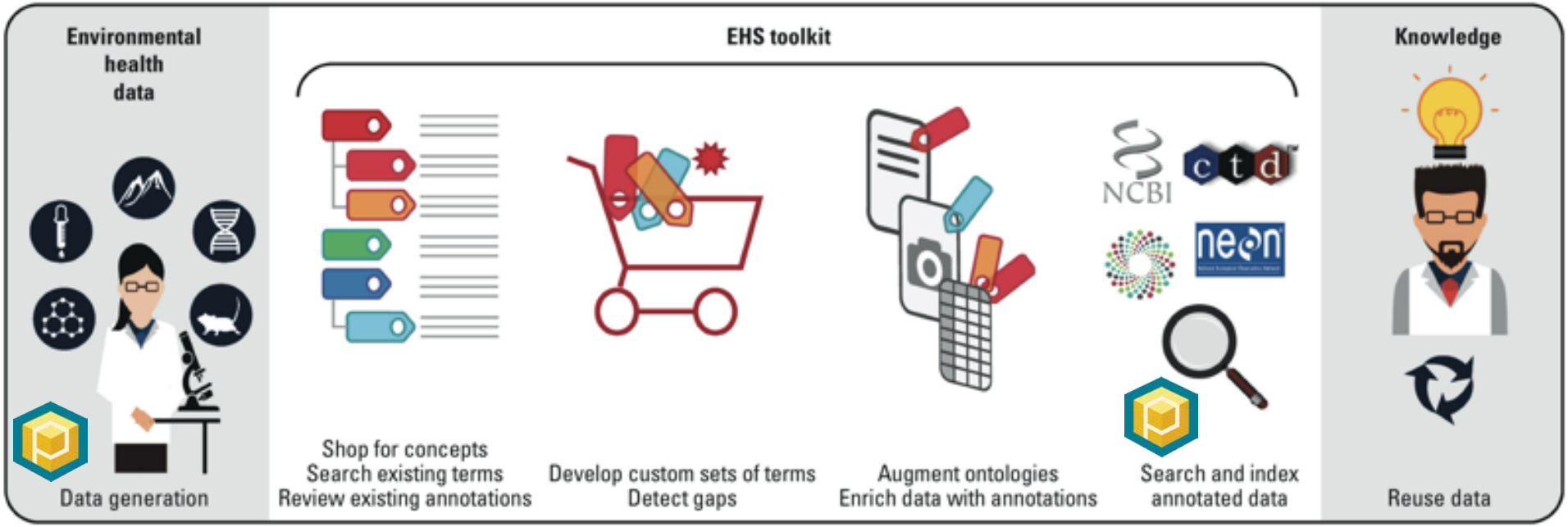
Global Alliance for Genomics & Health

If it is alive, it can be PhenoPackaged



Some biodiversity images adapted from http://i.vimeocdn.com/video/417366050_1280x720.jpg

A semantic vision for environmental health research



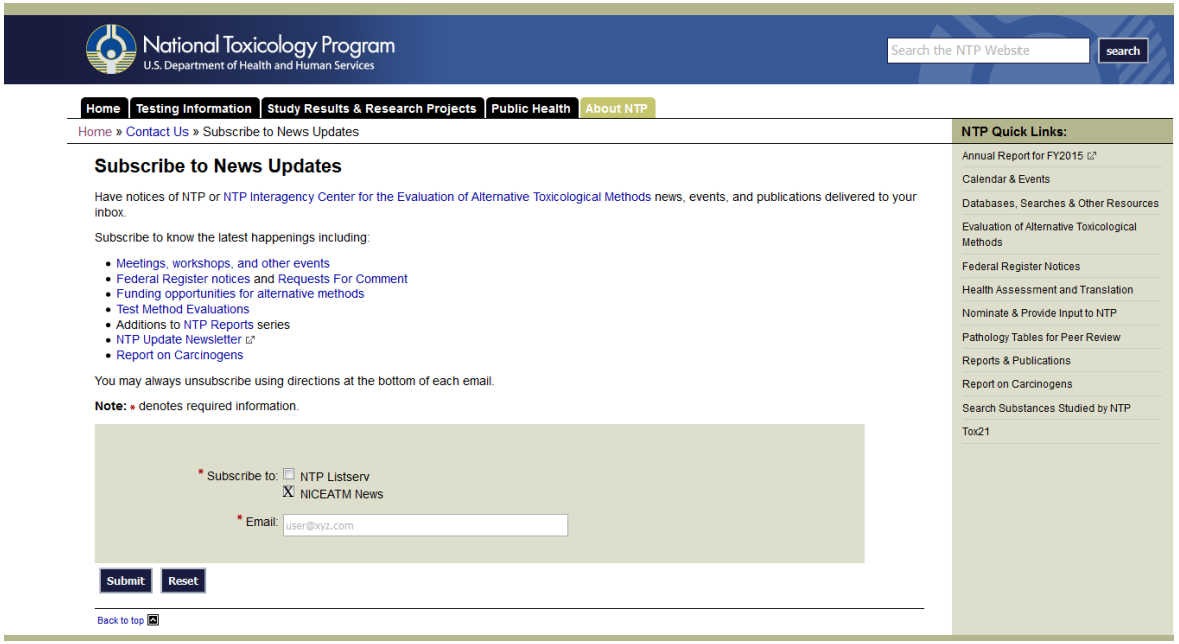
Laying a Community-Based Foundation for Data-Driven Semantic Standards in Environmental Health Sciences

https://ehp.niehs.nih.gov/15-10438/

NICEATM News

For updates on the SEAZIT project and other activities related to *in vitro* alternatives, subscribe to the NICEATM News email list.

- To subscribe to the NICEATM News email list, go to: https://tools.niehs.nih.gov/webforms/index.cfm/main/formViewer/form_id/361
- Check the NICEATM News box and click submit



Acknowledgements

OHSU
Matt Brush
Kent Shefchek
Julie McMurry
Tom Conlin
Nicole Vasilevsky
Dan Keith

Genomics England/Queen Mary
Damian Smedley
Jules Jacobson

Jackson Laboratory
Peter Robinson

ZFIN
Ceri VanSlyke
Yvonne Bradford

CTD
Carolyn Mattingly

Garvan
Tudor Groza

Max Plank
Pier Buttigieg

EBI
David Osumi-Sutherland

Lawrence Berkeley
Chris Mungall
Suzanna Lewis
Jeremy Nguyen
Seth Carbon
Nicole Washington

Charité
Sebastian Kohler

RTI
Jim Balhoff
Becky Boyles

Cyverse
Ramona Walls

FUNDING: NIH Office of Director: 2R24OD011883; NIH-UDP: HHSN268201300036C, HHSN268201400093P, NSF-DEB-0956049, NCINCI/Leidos #15X143, BD2K U54HG007990-S2 (Haussler) & BD2K PA-15-144-U01 (Kesselman)

With special thanks to Julie McMurry for excellent graphic design

The Monarch Initiative

www.monarchinitiative.org

PDs: Melissa Haendel, Chris Mungall, Peter Robinson

Funding:
NIH Office of Director: 1R24OD011883; NIH-UDP: HHSN268201300036C, HHSN268201400093P; NCINCI/Leidos #15X143, BD2K U54HG007990-S2 (Haussler) & BD2K PA-15-144-U01 (Kesselman)