



National Institute of Environmental Health Sciences  
*Your Environment. Your Health.*

---

# Best Practices: Ontology development and Curation

**Leigh Carmody, PhD**  
**Office of Data Science**  
**NIEHS**



National Institute of Environmental Health Sciences  
*Your Environment. Your Health.*



monarch  
INITIATIVE

### Advisory

- Nico Matentzoglou (Semanticity)
- Chris Mungall (LBL)
- Melissa Haendel (UNC)
- Peter Robinson (JAX/BIH)



hpo



mondo  
THE WORLD'S DISEASE CONCEPTS, UNIFIED



upheno  
ontology



maxo



ecto



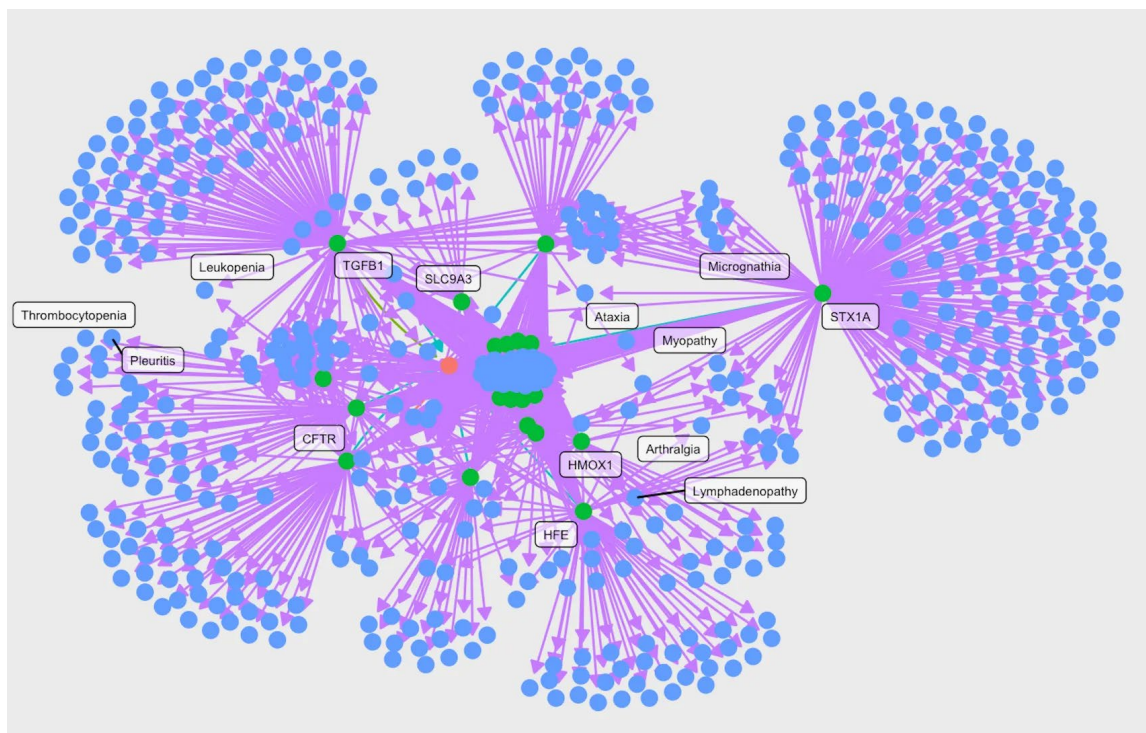
phenopackets



poet

# What do we want to do with (our) data?

- Extract meaningful insights
- Test hypotheses
- Identify patterns and trends
- Gain a deeper understanding of a phenomenon by analyzing the data contained within them



## FAIR Principles

- Findable
- Accessible
- Interoperable
- Reusable

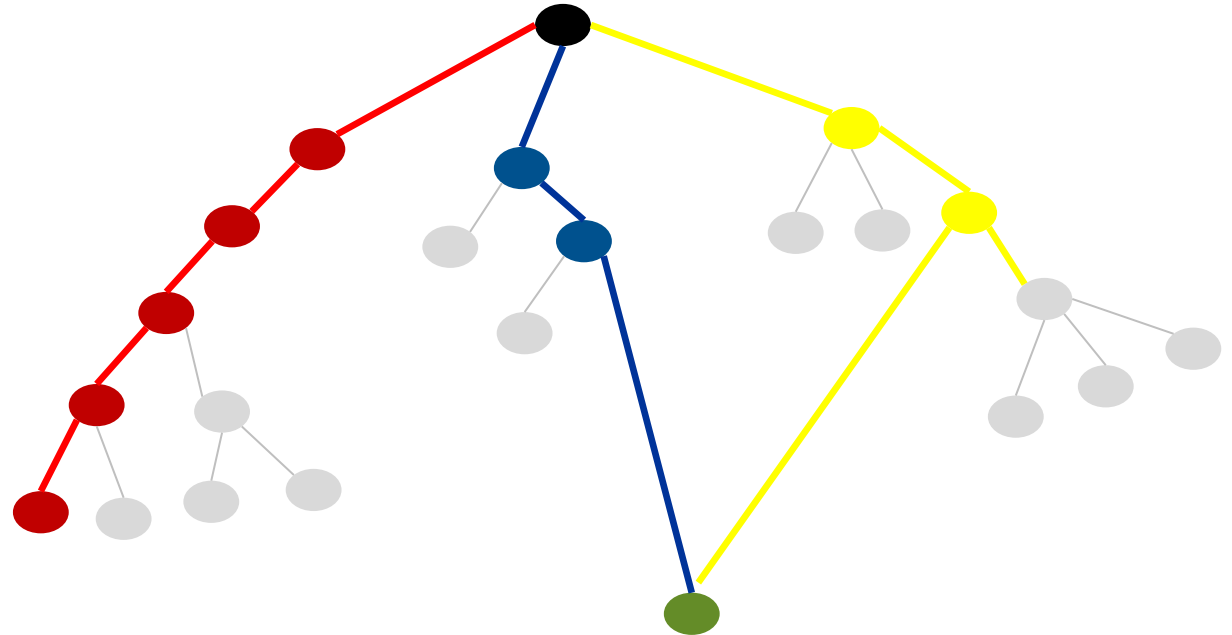
Data and vocabularies should be accessible to both humans and data machines.

## Definitions

- **Health and data standard terminologies:** is designed to provide a robust framework for representing and organizing knowledge within a specific domain, facilitating communication, and consistency across users and systems.
- **Common data element (CDE):** a standardized, precisely defined question that is paired with a set of specific allowable responses, that is then used systematically across different sites, studies, or clinical trials to ensure consistent data collection.
- **Ontology:** A systematic arrangement of important categories of concepts which exist in a field and showing the relations between them.

# Ontologies

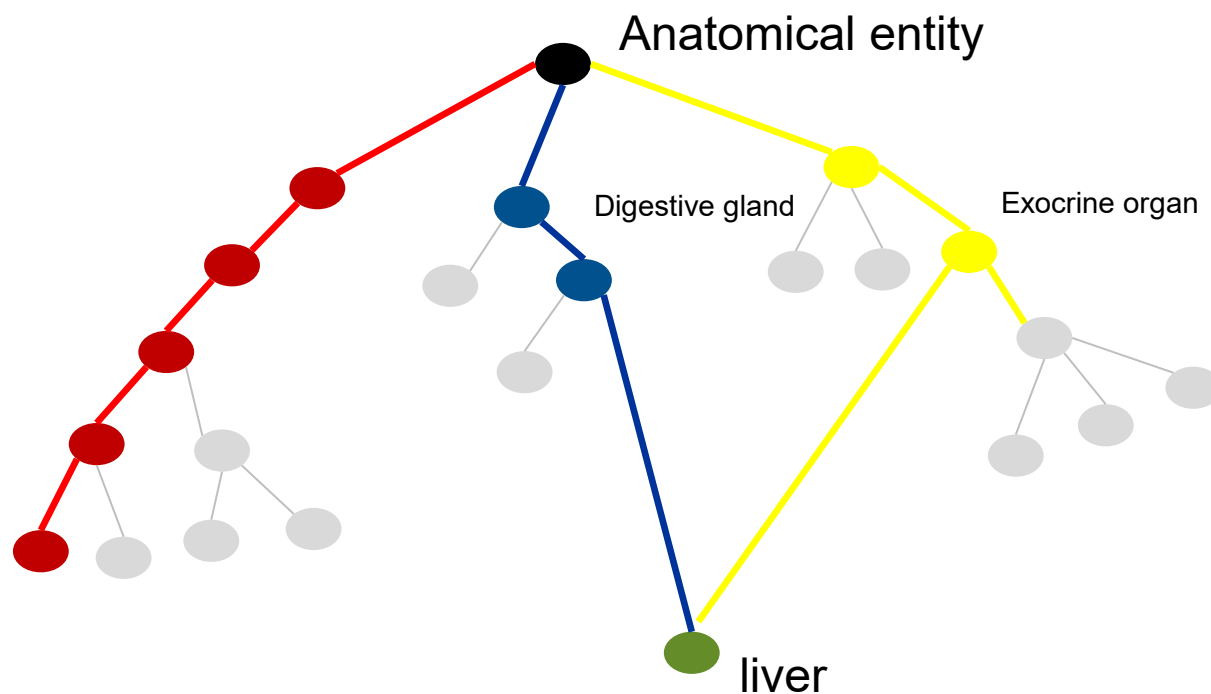
**Definition:** a set of concepts and categories in a subject area or domain that shows their properties and the relations between them.



# Ontologies

**Definition:** a set of concepts and categories in a subject area or domain that shows their properties and the relations between them.

**Uses:** Data management  
Ontologies can establish a common vocabulary and set of relationships to describe and understand data from different sources. This standardization allows systems to exchange data with a shared meaning, which is important for semantic interoperability.



# Best practices for ontology development

- Term requirements
- True-Path Rule
- Avoid Bundled Terms
- Pie Rule
- 5pm Rule



## Term requirements: Anatomy of an ontology term

\* **Label:** Deeply set eye    \* **ID:** HP:0000490

\* **Definition:** An eye that is more deeply recessed into the plane of the face than is typical.

\* **Parent Term:** Abnormality of globe location

**Synonyms:** Deep set eye, Deep-set eyes, Enophthalmos, Ocular depression, Sunken eye, Sunken eyes

**Reference:** PMID:19125427

**Comment:** This finding should be distinguished from a prominent supraorbital ridge or inferior orbital margin. In Deeply set eyes, the globe is recessed in comparison to the overall prominence of the face. There is no known objective measurement, and diagnosing this feature depends heavily on the experience of the observer.

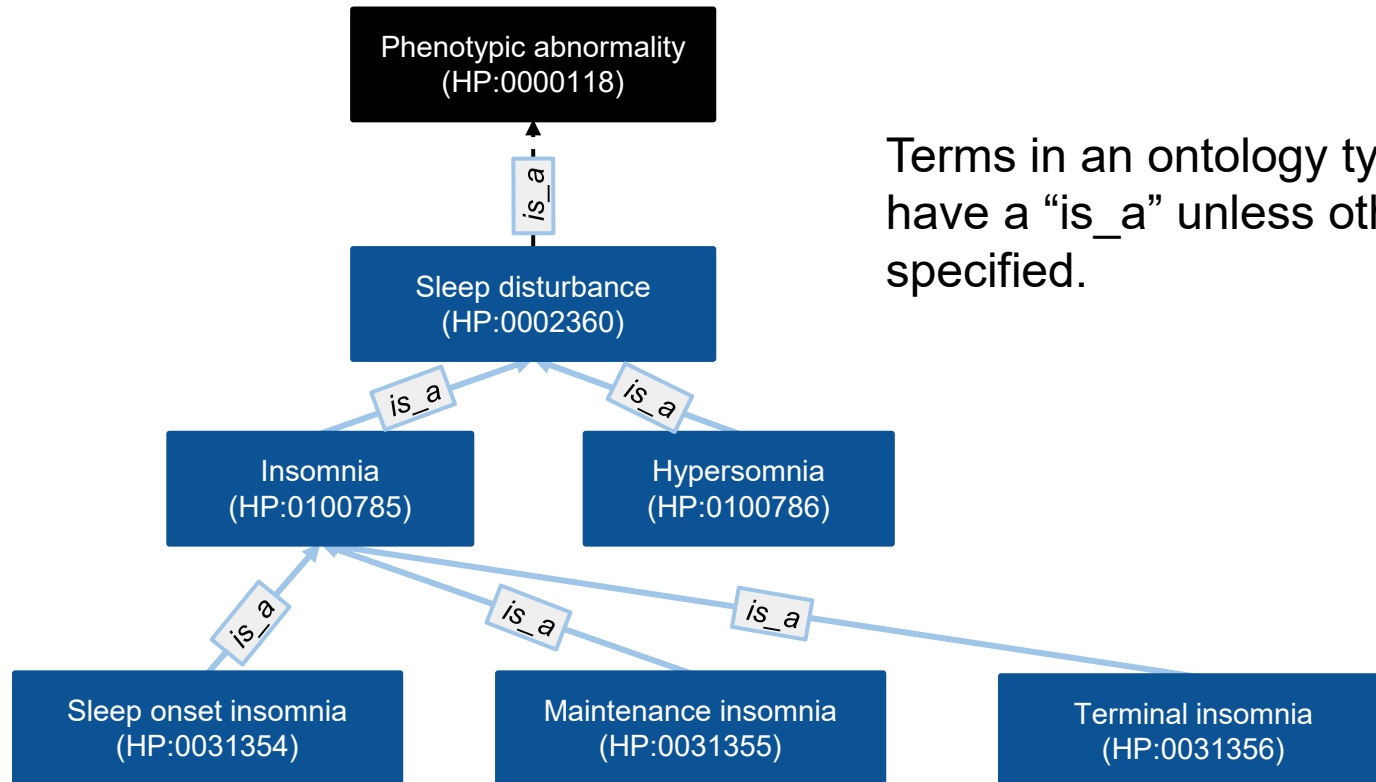
**Logical definition:**

'has part' some

(retracted and ('characteristic of' some 'eyeball of camera-type eye')

and ('has modifier' some abnormal))

# True-Path rule

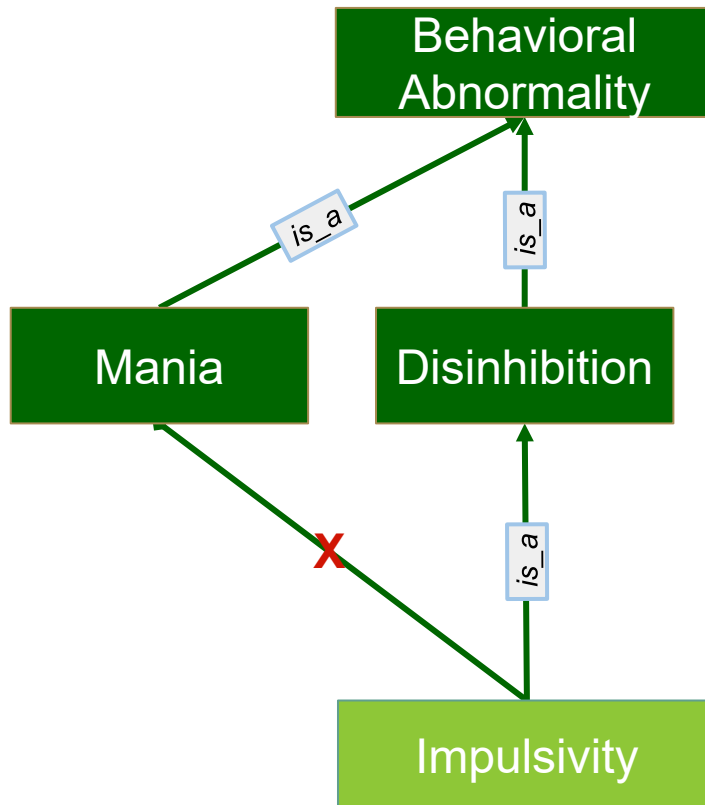


Terms in an ontology typically have a “is\_a” unless otherwise specified.



## True-Path rule

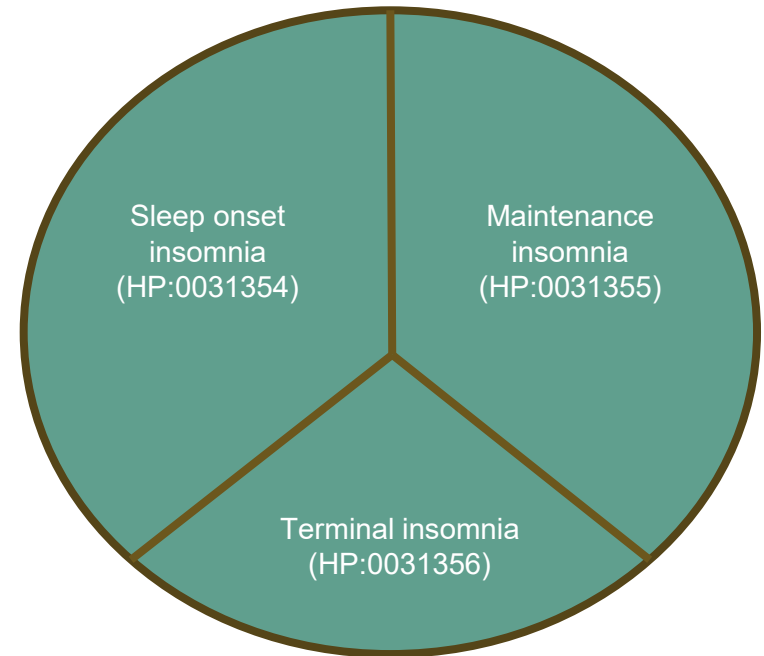
If a term can be used to describe a patient then all parents of that term can be used to describe that patient.



- Do patients with impulsive behavior always have mania?
- Do patients with impulsive behavior always have a behavioral abnormality?

## Pie Rule

The set of child terms of a given term should robustly cover the applications of the parent term, such that a parent term could always be replaced with a more specific child term.



## Avoid bundled terms

Avoid creating terms that refer to multiple terms that do not easily fit under a broader term. The only time to bundle terms is when they always occur together.

Examples from human phenotype issue tracker:

[‘Trunk and extremity distribution to skin lesions’](#)

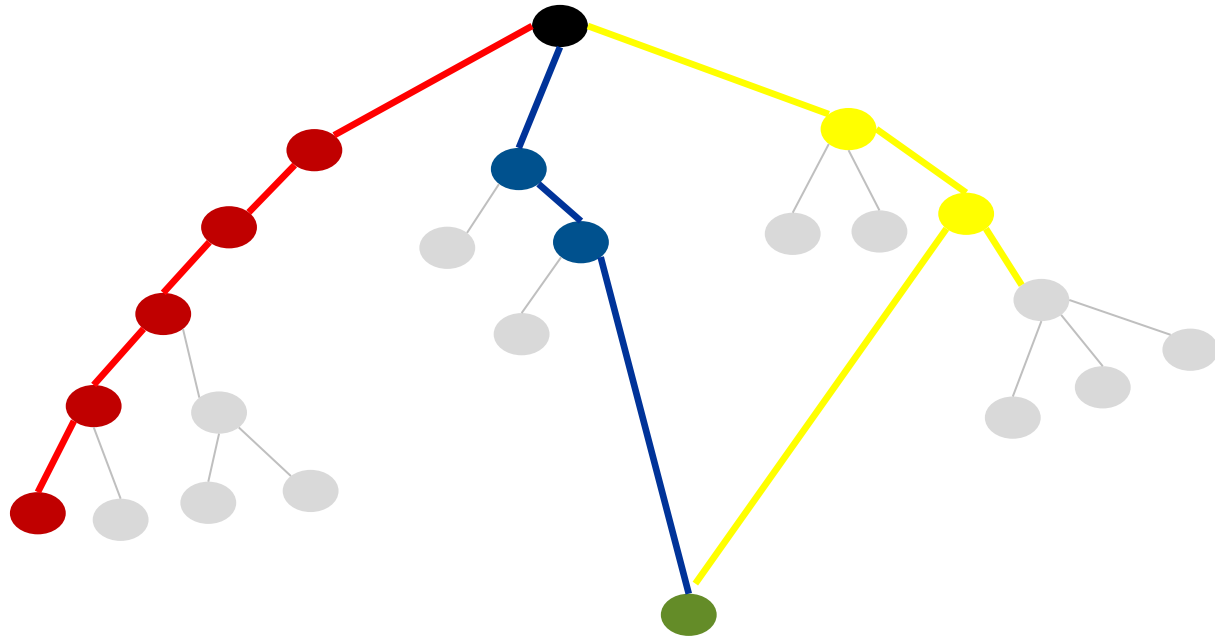
[‘Stocky habitus’](#): Body type defined by short stature and overweight, muscular appearance.

For curation: HP:0004322 Short stature, HP:0025502 overweight, Broad shouldered?.

# 5 O'clock rule

**(Findable).** At the end of a long day, the term should be easy and intuitive to find. That means clear labeling and thoughtful hierarchy.

There is another way to make sure there is a logical hierarchy and that is to leverage other ontologies.



\* **Label:** Deeply set eye      \* **ID:** HP:0000490

**Logical definition:**

'has part' some

(retracted and ('characteristic of' some 'eyeball of camera-type eye')  
and ('has modifier' some abnormal))



# Curation best practices

- Finding ontologies
- Selecting terms
- Adding/editing terms

# Finding an ontology: where?

## Different resources have different ontologies

- Ontology lookup service (OLS): <https://www.ebi.ac.uk/ols4>

- Ex. uPheno

```

renal/urinary system phenotype (2,534)
├─ Abnormality of the urinary system HP (1,178)
├─ renal system development phenotype (2)
│   └─ renal system development process quality, abnormal ZP (1)
├─ renal system morphology phenotype (1,574)
└─ renal/urinary system phenotype MP (692)
  
```

- Bioportal: <https://bioportal.bioontology.org/>

- Ex. Envo-  
• mapping

<a href="http://purl.obolibrary.org/obo/ENVO_01000600">http://purl.obolibrary.org/obo/ENVO_01000600</a>	NMDCO	LOOM
<a href="http://www.wikidata.org/entity/Q21555202">http://www.wikidata.org/entity/Q21555202</a>	HHEAR	LOOM
<a href="http://sweetontology.net/matrWater/Rainwater">http://sweetontology.net/matrWater/Rainwater</a>	SWEET	LOOM

- Ontobee: <https://ontobee.org/>

- SPARQL query templates

- OBO foundry: <https://obofoundry.org/>

- Tons of resources



## OBO foundry: <https://obofoundry.org/>

- Lots of available resources
- Collection of largely science/biology-based ontologies
- Ontologies have been vetted\*
  - Rating system\*

P1) Open

P2) Common Format - common formal language (BFO/COB compliant)

P3) URI/Identifier Space - ex. [http://purl.obolibrary.org/obo/UPHENO\\_3000004](http://purl.obolibrary.org/obo/UPHENO_3000004)

P4) Versioning

P5) Scope - extent of the domain or subject matter it intends to cover. P6) Textual Definitions - define terms

P7) Relations - Relations should be reused from the Relations Ontology (RO).

P8) Documentation

P9) Documented Plurality of Users \*\*

P10) Commitment To Collaboration

P11) Locus of Authority - There should be a person who is identifiably in charge

P12) Naming Conventions - The names in an ontology must be intelligible to scientists and amenable to natural language processing. Primary labels should be unique among OBO Library ontologies.




























P13) Notification of Changes

P16) Maintenance

P20) Responsiveness

## OBO foundry: <https://obofoundry.org/>

- Lots of available resources
- Collection of largely science/biology-based ontologies
- Ontologies have been vetted\*
  - Rating system\*: <https://oops.linkeddata.es/catalogue.jsp>

ID ^	Title ^	Description	Quick Access	OBO Principles ^	Social
<a href="#">cl</a>	Cell Ontology	The Cell Ontology is a structured controlled vocabulary for cell types in animals.	<a href="#">Home</a> <a href="#">Download</a> <a href="#">View</a> <a href="#">Search</a> <a href="#">Browse</a> <a href="#">Help</a>	 QC <b>WARN 2</b> 	 Stars 164
<a href="#">fbbt</a>	Drosophila gross anatomy	An ontology representing the gross anatomy of Drosophila melanogaster.	<a href="#">Home</a> <a href="#">Download</a> <a href="#">View</a> <a href="#">Search</a> <a href="#">Browse</a> <a href="#">Help</a>	 QC <b>WARN 2</b> 	 Stars 19
<a href="#">fbdv</a>	Drosophila development	A structured controlled vocabulary of the development of Drosophila melanogaster.	<a href="#">Home</a> <a href="#">Download</a> <a href="#">View</a> <a href="#">Search</a> <a href="#">Browse</a> <a href="#">Help</a>	 QC <b>INFO 1</b> 	 Stars 5
<a href="#">plana</a>	planaria-ontology	PLANA, the planarian anatomy ontology, encompasses the anatomy and life cycle stages for both __Schmidtea mediterranea__ biotypes.	<a href="#">Home</a> <a href="#">Download</a> <a href="#">View</a> <a href="#">Search</a> <a href="#">Browse</a> <a href="#">Help</a>	 QC <b>WARN 1</b> 	 Stars 5
<a href="#">po</a>	Plant Ontology	The Plant Ontology is a structured vocabulary and database resource that links plant anatomy, morphology and growth and development to plant...	<a href="#">Home</a> <a href="#">Download</a> <a href="#">View</a> <a href="#">Search</a> <a href="#">Browse</a> <a href="#">Help</a>	 QC <b>WARN 2</b> 	 Stars 60
<a href="#">uberon</a>	Uberon multi-species anatomy ontology	An integrated cross-species anatomy ontology covering animals and bridging multiple species-specific ontologies	<a href="#">Home</a> <a href="#">Download</a> <a href="#">View</a> <a href="#">Search</a> <a href="#">Browse</a> <a href="#">Help</a>	 QC <b>WARN 2</b> 	 Stars 137
<a href="#">wbls</a>	C. elegans development ontology	A structured controlled vocabulary of the development of <i>Caenorhabditis elegans</i> .	<a href="#">Home</a> <a href="#">Download</a> <a href="#">View</a> <a href="#">Search</a> <a href="#">Browse</a> <a href="#">Help</a>	 QC <b>INFO 1</b> 	 Stars 5
<a href="#">zfa</a>	Zebrafish anatomy and development ontology	A structured controlled vocabulary of the anatomy and development of the Zebrafish	<a href="#">Home</a> <a href="#">Download</a> <a href="#">View</a> <a href="#">Search</a> <a href="#">Browse</a> <a href="#">Help</a>	 QC <b>WARN 2</b> 	 Stars 6
<a href="#">aism</a>	Ontology for the Anatomy of the Insect SkeletoMuscular system	The AISM contains terms used in insect biodiversity research for describing structures of the exoskeleton and the skeletomuscular	<a href="#">Home</a> <a href="#">Download</a> <a href="#">View</a> <a href="#">Search</a> <a href="#">Browse</a> <a href="#">Help</a>	 QC <b>ERROR 1</b> 	 Stars 11

## OBO foundry: <https://obofoundry.org/>

- Lots of available resources
- Collection of largely science/biology-based ontologies
- Ontologies have been vetted\*
  - Rating system\*

P1) Open

P2) Common Format - common formal language (BFO/COB compliant)

P3) URI/Identifier Space - ex. [http://purl.obolibrary.org/obo/UPHENO\\_3000004](http://purl.obolibrary.org/obo/UPHENO_3000004)

P4) Versioning

**P5) Scope - extent of the domain or subject matter it intends to cover.**

**P6) Textual Definitions - define terms**

P7) Relations - Relations should be reused from the Relations Ontology (RO).

P8) Documentation

P9) Documented Plurality of Users \*\*

**P10) Commitment To Collaboration**

P11) Locus of Authority - There should be a person who is identifiably in charge

P12) Naming Conventions - The names in an ontology must be intelligible to scientists and amenable to natural language processing. Primary labels should be unique among OBO Library ontologies.

**P13) Notification of Changes**

**P16) Maintenance**

**P20) Responsiveness**

## Selecting terms

**Ex. Data:** Rat/Mouse/Guppy with liver adenoma, adenoma or carcinoma

tumor: MPATH:353 hepatocellular adenoma

Disease: MONDO:0018902/DOID:0050868 hepatocellular adenoma

Phenotype: MP:0003324 Increased liver adenoma incidence

ZP:0103622 hepatocellular adenoma liver increased amount, abnormal

# Selecting terms

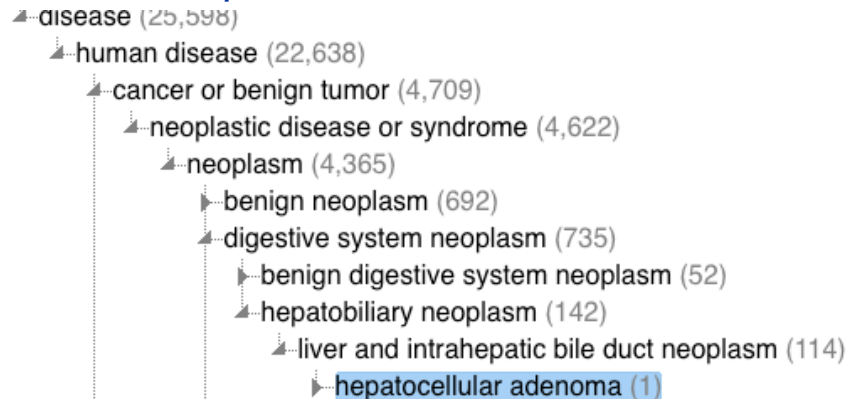
**Ex. Data:** Rat/Mouse/Guppy with liver adenoma, adenoma or carcinoma

tumor: MPATH:353 hepatocellular adenoma

- MPATH= [Mouse](#) pathology ontology
- Does not have an issue tracker.
- There is a contact information and there is will to make minimal changes to the ontology.

Disease: MONDO:0018902/DOID:0050868 hepatocellular adenoma

- [Is\\_a](#) relationship means this is a human disease.



Phenotype: MP:0003324 Increased liver adenoma incidence

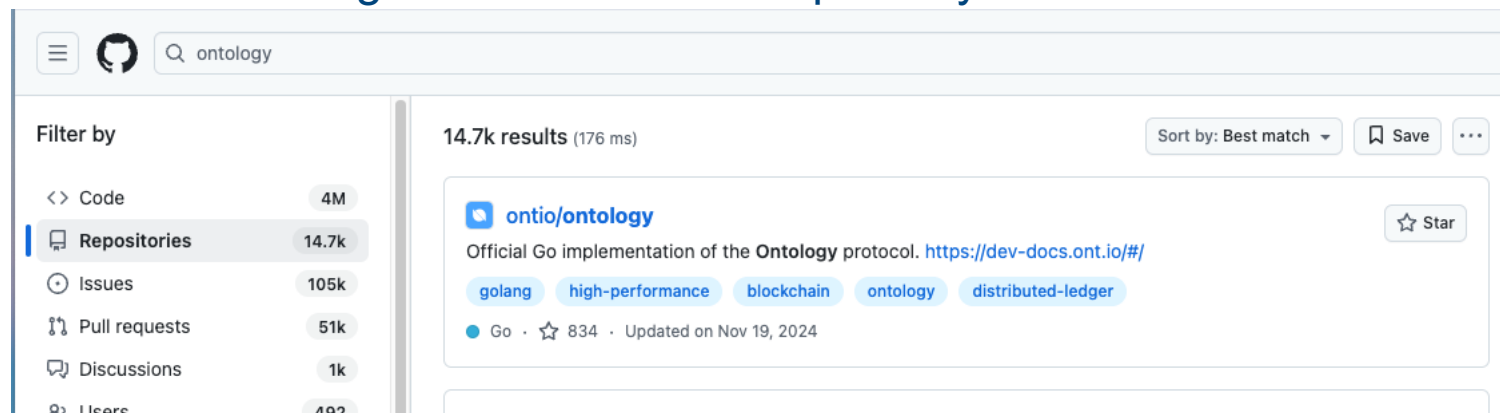
ZP:0103622 hepatocellular adenoma liver increased amount, abnormal

- Deal with separate phenotype ontologies for adding terms\*
- Different phrasing
- ZP is zebrafish ontology

# Adding terms/editing ontologies

## Github:

Most ontologies have a Github repository.



Can search in Github, but Google search of ontology name +github is the quickest route.



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

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

<https://github.com/mgijax/mammalian-phenotype-ontology>

<https://github.com/obophenotype/upheno>

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

<https://github.com/EnvironmentOntology/envo>

<https://github.com/ebi-chebi/ChEBI>

<https://github.com/obophenotype/uberon>

<https://github.com/geneontology/go-ontology>

<https://github.com/obophenotype/cell-ontology>

<https://github.com/PROconsortium/PRoteinOntology>

# What makes a good github ticket?

<https://oboacademy.github.io/oobook/explanation/writing-good-issues/>

- Search existing issues before creating a new one

- Short, descriptive title

- Short explanation of the problem and include an example

- Mention related issues (use # for same repo, and url for different repo)

- Tag users who should be aware of the issue

- Give specifics about how to complete/close ticket.

- Be polite!



# Good\* term requests

Issues 258 Pull requests 2 Actions Projects Wiki Security Insights

abnormal jejunum goblet cell morphology - amend typo in term label #4080

New issue



Open



anna-anagnostop opened yesterday

...

Correct typo in label for abnormal jejunum goblet cell morphology MP:0013798; change jejunum to jejunal.



Assignees

anna-anagnostop

Labels

## Tracheobronchomegaly #11038



anna-anagnostop

Open



MickeySegal opened 2 days ago

...

**Preferred term label:**

Tracheobronchomegaly

**Synonyms**

**Definition (free text, please give PubMed ID)**

dilatation of the intrathoracic trachea and of the major bronchi

**Parent term (use hpo.jax.org/app)**

**Diseases characterized by this term ? (e.g. Orphanet or OMIM number)**

Mounier-Kuhn syndrome

**Your nano-attribution (ORCID)**

Create sub-issue



## Additional Resources

- OBO academy <https://oboacademy.github.io/oobook/>

- Lessons:

[Contributing to OBO ontologies](#)

[Ontology Pipelines with ROBOT](#)

[Developing an OBO Ontology](#)

[Ontology Design](#)

[Disease and Phenotype Ontologies](#)

[Leveraging ChatGPT for ontology curation](#)

[Using Ontologies in Practice--LinkML can help](#)

- Tutorials:

Protégé

Ontology Pipelines - ODK, ROBOT, etc,

Templates

Git, GitHub and Collaborative Workflows

OntoGPT

Command line

Exomiser

Information Extraction

## Additional Resources

- OBO academy <https://oboacademy.github.io/oobook/>
  - Youtube account to watch tutorials <https://www.youtube.com/@obo-academy>
- OBO foundry <https://obofoundry.org/>
  - Principles and best practices
- Ontology editor: <https://protege.stanford.edu/>
  - [https://protege.stanford.edu/publications/ontology\\_development/ontology101.pdf](https://protege.stanford.edu/publications/ontology_development/ontology101.pdf)
- Tutorial: Introduction to RDF and OWL
  - <https://github.com/CSIRO-enviro-informatics>
- EMBL-EBI Training <https://www.ebi.ac.uk/training/on-demand>
  - Ontology trainings and tutorials

## Conclusion

- Data standards are critical to integrating data
- Ontologies are especially helpful by establishing relationships.
- Following ontology development best practices aids all users of ontologies
- There is no one way to annotate data, but using ontology terms aids in linking data
- There is no perfect ontology. There are mistakes, and areas that are not as well covered. As a user, it is critical to help an ontology by requesting terms, fixing errors

# Thank You.

- Office of Data Science
  - Charles Schmitt, PhD
  - Darius Bost, PhD
  - Mike Conway, MSc
  - Jennifer Foster, PhD
  - Rupali Gupta, PhD
  - Ying Liu, PhD, MBA
  - Sue Nolte, MSc
  - Deep Patel, MSc
  - Maria Shatz, PhD, MSc
  - Amlan Talukder, PhD