



mondo

THE WORLD'S DISEASE CONCEPTS, UNIFIED

Nicole Vasilevsky, OHSU

Chris Mungall, LBNL

ClinGen Biocurator Working Group Meeting, October 22, 2020

These slides:
<https://bit.ly/clingen-mondo>

Overview

1

Overview
of Mondo

2

View
Mondo

3

Website

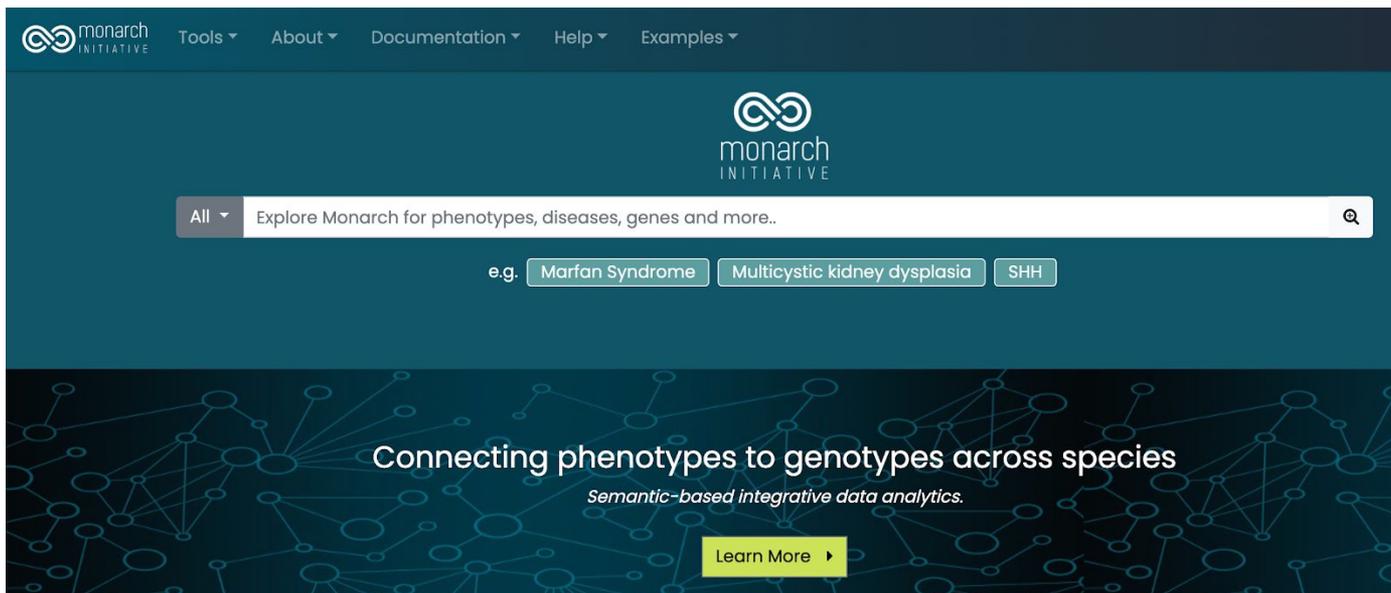
4

New term
requests

5

Gene-based
names

Monarch: connecting diseases, phenotypes, and genes



The screenshot shows the Monarch Initiative website. At the top left is the Monarch Initiative logo. To its right are navigation links: Tools, About, Documentation, Help, and Examples. In the center is the Monarch Initiative logo and a search bar with the placeholder text "Explore Monarch for phenotypes, diseases, genes and more..". Below the search bar are three example buttons: "Marfan Syndrome", "Multicystic kidney dysplasia", and "SHH". At the bottom of the page is a dark blue banner with a network diagram background. The banner contains the text "Connecting phenotypes to genotypes across species" and "Semantic-based integrative data analytics." Below this text is a yellow "Learn More" button with a right-pointing arrow.

What is an ontology?

on·tol·o·gy

/än'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

NCBI Taxonomy

Uberon

Gene Ontology

Aves (birds)

is_a

Cygnus (swans)

part_of

is_a

is_a

Cygnus columbianus (tundra swan)

Cygnus atratus (black swan)

is_a

Cygnus columbianus bewickii (Bewick's swan)

webbed pes (webbed foot)

is_a

Pes (foot)

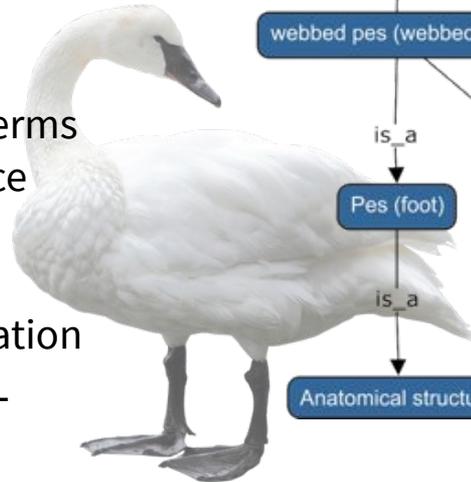
capable of

Swimming behavior

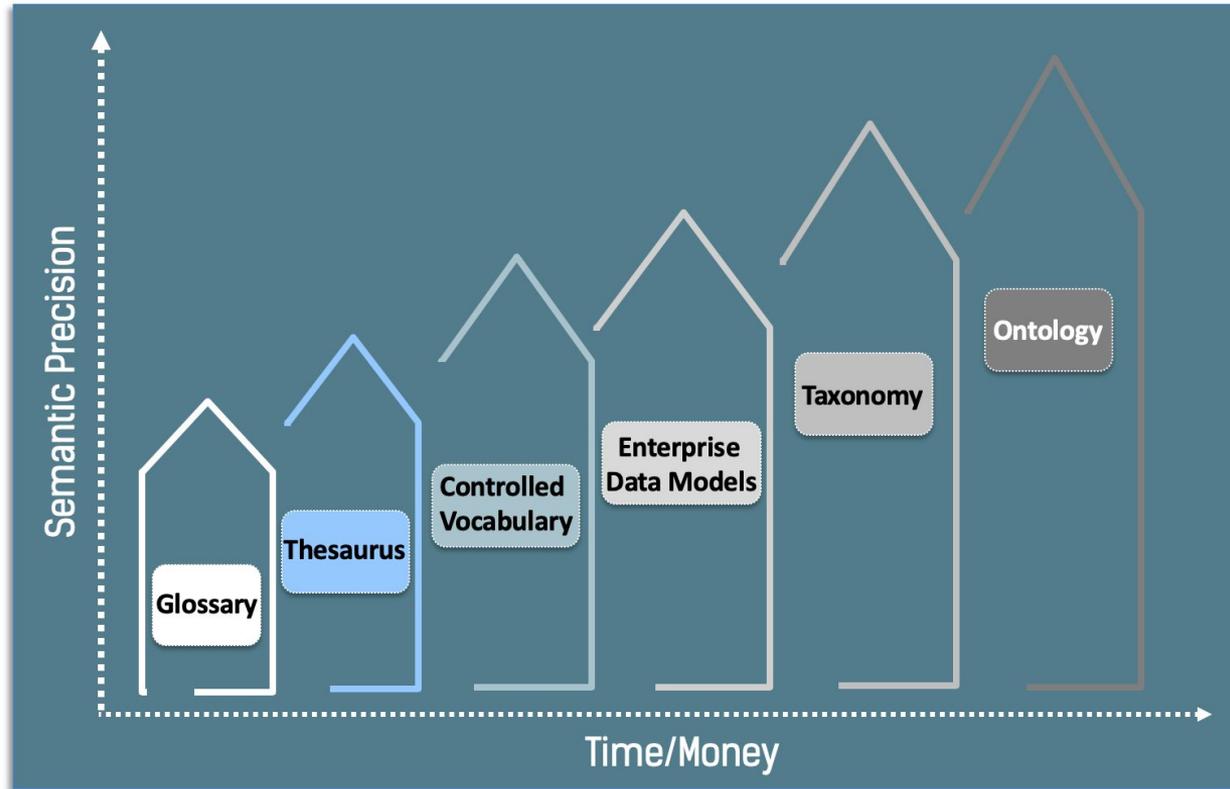
is_a

Anatomical structure

Biological Process



Complexity of Vocabulary Types



What is the most clinically useful way to define and group diseases?

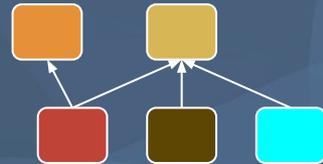
CANCER **COMPLEX**
INFECTIOUS **RARE**
MENDELIAN

We needed:

- disease concepts spanning multiple categories
- a systematic way of relating these concepts

Why not just use mappings?

- Many terminologies / ontologies / lists include mappings
 - These can be used to cross-walk
- Problems:
 - Often mutually inconsistent
 - N^2 sets of mappings!
 - Not 1:1 equivalents





A Census of Disease Ontologies

Annual Review of Biomedical Data Science

Vol. 1:305-331 (Volume publication date July 2018)

First published as a Review in Advance on May 9, 2018

<https://doi.org/10.1146/annurev-biodatasci-080917-013459>

Melissa A. Haendel,^{1,2} Julie A. McMurry,¹ Rose Relevo,¹ Christopher J. Mungall,³ Peter N. Robinson,⁴ and Christopher G. Chute⁵

¹Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, Oregon 97239, USA; email: haendel@ohsu.edu

²Linus Pauling Institute, Oregon State University, Corvallis, Oregon 97331, USA

³Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

⁴The Jackson Laboratory, Farmington, Connecticut 06032, USA

⁵School of Medicine, School of Public Health, and School of Nursing, Johns Hopkins University, Baltimore, Maryland 21205, USA



Download PDF



Supplemental Material



Article Metrics

Permissions | Reprints | Download Citation | Citation Alerts

No one system fit the bill

- There are a wealth of disease resources
 - *specialized*
 - (e.g OMIM: Mendelian)
 - *generalized*
 - (e.g. MESH, SNOMED)
- No single resource sufficient
 - *specialized*
 - did not include concepts we needed
 - *generalized*
 - lacked sufficient depth/precision in key domains

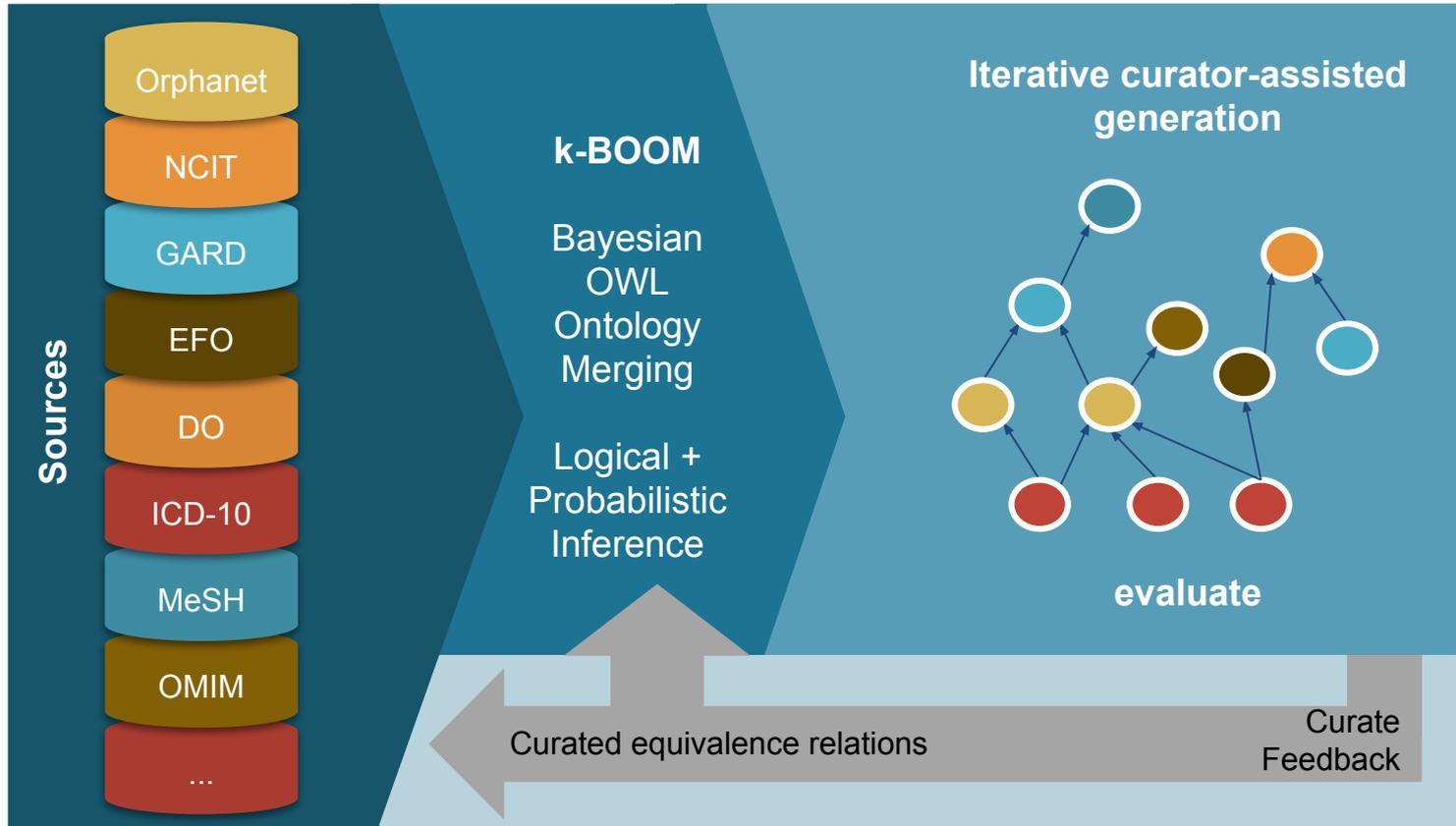
Assessment of synonyms, hierarchies, and mappings across ontology sources for example diseases EDS and Pancreatic cancer

Source ontology	General information			Ehlers–Danlos syndrome				Pancreatic cancer			
	Has definition ^b	Hierarchy type	Language translation	Xrefs	Synonyms	Parents	Descendants	Xrefs	Synonyms	Parents	Desce
DO	Yes	Single	No	6/11	0/2	1/1	6/6	4/11	0/9	2/2	37/37
ICD-10	~Yes ^c	Single	No	–	–	1/1	–	–	–	1/1	–
ICD-11	Yes	Multi	Planned	–	1/2	2/2	26/26	–	2/7	3/3	23/23
MedDRA	No	Multi	Yes	–	–	2/2	–	–	–	1/2	–
MeSH	~Yes ^d	Single	Yes	–	22/23 ^e	3/4	16/21	–	12/25 ^e	3/3	14/14
MonDO	Yes	Multi	Planned	–	0/2	7/7	36/36	–	8/8	2/2	44/44
NCIt	Yes	Multi	No	–	–	2/2	7/7	–	3/3	2/2	68/68
OMIM ^f	Yes	Flat	No	–	7/7	–	–	2/2	–	–	–
Orphanet	No	Multi	No	2/5	–	6/6	24/24	5/5	0/2	7/7	10/10
SNOMED	N	Multi	Yes	–	2/11	8/8	21/21	–	5/5	3/3	47/47
UMLS	No	Multi	Yes	–	27/46	2/2	7/8	–	50/62	–	6/6

Wide heterogeneity in:

- placement of diseases hierarchically, and therefore meaning,
- mapping to other diseases,
- the number and typing of synonyms

Evidence-based merging of equivalent classes



Relationships to other resources

- In obo format version and in OLS these are database cross references ('xrefs')
- In owl version we use explicit *logical axioms*, e.g:
 - **equivalentTo**
 - **relatedTo**

OLS / Mondo Disease Ontology **MONDO** / **MONDO:0001586**  Copy

mucopolysaccharidosis type 1

 http://purl.obolibrary.org/obo/MONDO_0001586  Copy

property value

exactMatch

<http://linkedlifedata.com/resource/umls/id/C2713321>,

exactMatch NCIT:C85053, exactMatch

Orphanet:579, exactMatch DOI:12802,

closeMatch

<http://identifiers.org/snomedct/267453008>,

exactMatch

<http://linkedlifedata.com/resource/umls/id/C0023786>,

exactMatch <http://identifiers.org/meddra/10056886>,

closeMatch

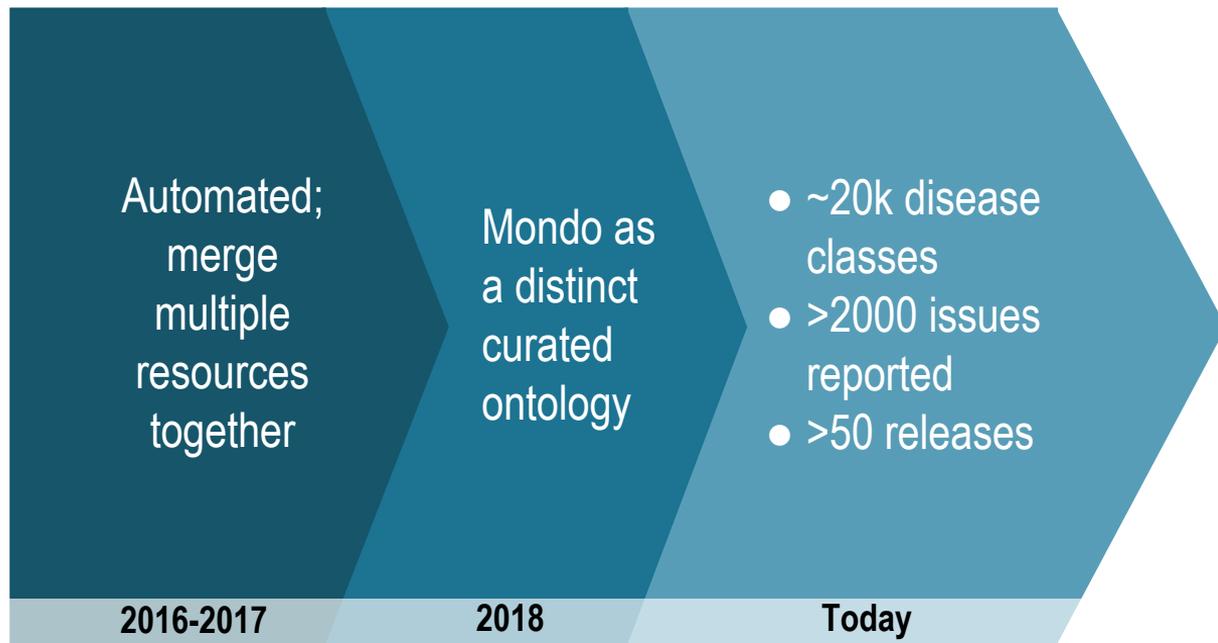
<http://identifiers.org/snomedct/190938004>,

exactMatch

<http://identifiers.org/snomedct/75610003>,

exactMatch <http://identifiers.org/mesh/D008059>

Current Status



MONDO IDs assigned and tracked for each concept

Use of standard ontology engineering practices

Periodically aligned and synced with existing resources

Released monthly

Mondo community

Where to view Mondo



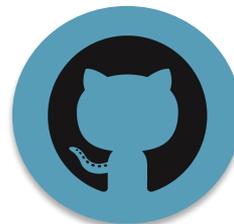
OBO Foundry

obofoundry.org/ontology/mondo



Ontology Lookup Service

ebi.ac.uk/ols/ontologies/mondo



GitHub

github.com/monarch-initiative/mondo

Community developed

Weekly Calls

Fridays, 9am PT/12pm ET
Zoom

Join our mailing list:

<https://groups.google.com/forum/#!forum/mondo-users>

Mondo website



[Home](#) [Sources](#) [Users](#) [Contributors](#) [Editors](#) [Resources](#) [FAQ](#)

Mondo Disease Ontology

The Mondo Disease Ontology (Mondo) aims to harmonize disease definitions across the world.

About

Numerous sources for disease definitions and data models currently exist, which include [HPO](#), [OMIM](#), [SNOMED CT](#), [ICD](#), [PhenoDB](#), [MedDRA](#), [MedGen](#), [ORDO](#), [DO](#), [GARD](#), etc; however, these sources partially overlap and sometimes conflict, making it difficult to know definitively how they relate to each other. This has resulted in a proliferation of mappings between disease entries in different resources; however mappings are problematic: collectively, they are expensive to create and maintain. Most importantly, the mappings lack completeness, accuracy, and precision; as a result, mapping calls are often inconsistent between resources. The UMLS provides intermediate concepts through which other resources can be mapped, but these mappings suffer from the same challenges: they are not guaranteed to be one-to-one, especially in areas with evolving disease concepts such as rare disease.

In order to address the lack of a unified disease terminology that provides precise equivalences between disease concepts, we created Mondo, which provides a logic-based structure for unifying multiple disease resources.

Mondo's development is coordinated with the [Human Phenotype Ontology \(HPO\)](#), which describes the individual phenotypic features that constitute a disease. Like the HPO, Mondo provides a hierarchical structure which can be used for classification or "rolling up" diseases to higher level groupings. It provides mappings to other disease resources, but in contrast to other mappings between ontologies, we precisely annotate each mapping using strict semantics, so that we know when two disease names or identifiers are equivalent or one-to-one, in contrast to simply being closely related.

For more details, please see these [slides](#).

Mondo website

[Home](#)[Sources](#)[Users](#)[Contributors](#)[Editors](#)[Resources](#)[FAQ](#)

FAQs

What is an ontology?

An ontology is a formal, computational representation of knowledge in a particular domain or area of knowledge, such as diseases or anatomy. Terms are arranged in a hierarchy, and the terms and relationships between them are defined using both human readable and machine readable definitions, allowing logical inference and sophisticated queries. They are expressed in a knowledge representation language like RDF or OWL.

I want to request a new term or request a change to Mondo, how do I do so?

All requests should go on our [GitHub issue tracker](#). You will need to create a free [GitHub](#) account if you do not already have one.

What kind of information should I include when I create a ticket on the GitHub tracker?

We have [guidelines](#) on how to make a good term request. Also note, when you click [New Issue](#) on our issue tracker, it will give you options to choose a pre-formatted template that will suggest the type of information to include on a ticket. If none of the templates fit your issue, you can [open a blank issue](#).

How often is Mondo released?

Mondo is released around the first of each month. See the latest release [here](#).

Mondo website



Home Sources Users **Contributors** Editors Resources FAQ

Mondo Disease Ontology

The Mondo Disease Ontology (Mondo) aims to harmonize disease definitions across the world.

About

Numerous sources for disease definitions and data models currently exist, which include [HPO](#), [OMIM](#), [SNOMED CT](#), [ICD](#), [PhenoDB](#), [MedDRA](#), [MedGen](#), [ORDO](#), [DO](#), [GARD](#), etc; however, these sources partially overlap and sometimes conflict, making it difficult to know definitively how they relate to each other. This has resulted in a proliferation of mappings between disease entries in different resources; however mappings are problematic: collectively, they are expensive to create and maintain. Most importantly, the mappings lack completeness as a result, mapping calls are often inconsistent between resources. The UMLS provides intermediate concepts that can be mapped, but these mappings suffer from the same challenges: they are not guaranteed to be one-to-one for evolving disease concepts such as rare disease.

In order to address the lack of a unified disease terminology that provides precise equivalences between diseases, we have created the Mondo Disease Ontology, which provides a logic-based structure for unifying multiple disease resources.

Mondo's development is coordinated with the [Human Phenotype Ontology \(HPO\)](#), which describes the individual concepts that constitute a disease. Like the HPO, Mondo provides a hierarchical structure which can be used for classification and higher level groupings. It provides mappings to other disease resources, but in contrast to other mappings between resources, we annotate each mapping using strict semantics, so that we know when two disease names or identifiers are equivalent to simply being closely related.

For more details, please see these [slides](#).

Mondo Documentation

Mondo Documentation

- About Mondo
- Rare Disease analysis in Mondo
- Build pipeline and Releases
- Changelog
- Developer guide ^
- Editors guide ^
 - Introduction to Mondo editors guide

How to Make a Good Term Request

The following is intended to serve as a guide for anyone who would like to contribute to the Mondo project by making new term requests.

Does the term you are looking for already exist?

<https://mondo.readthedocs.io/en/latest/editors-guide/c-make-good-term-request/>

Overall design principles

Dead Simple Ontology Design Patterns (DOSDP)

Pre-made patterns that specify:

- label
- text definition
- synonyms
- *logical definition*

Diseases series by gene

```
pattern_name: disease_series_by_gene
```

```
description: >-
```

```
  This pattern is for diseases that are caused by a single mutation in a single gene, that have gene-based names, such as new disease terms that are requested by ClinGen, like like MED12-related intellectual disability syndrome.
```

```
Examples: [MED12-related intellectual disability syndrome](http://purl.obolibrary.org/obo/MONDO_0100000), [TTN-related myopathy](http://purl.obolibrary.org/obo/MONDO_0100175), [MYPN-related myopathy](http://purl.obolibrary.org/obo/MONDO_0015023)
classes:
  disease: MONDO:0000001
  gene: SO:0001217
```

```
relations:
```

```
  disease has basis in dysfunction of: RO:0004020
```

```
vars:
```

```
  disease: "'disease'"
  gene: "'gene'"
```

```
name:
```

```
  text: '%s caused by mutation in %s'
  vars:
    - disease
    - gene
```

Disease series by gene

description: This pattern is for diseases that are caused by a single mutation in a single gene, that have gene-based names, such as new disease terms that are requested by ClinGen, like MED12-related intellectual disability syndrome.

Examples: MED12-related intellectual disability syndrome, TTN-related myopathy, MYPN-related myopathy

vars:

```
disease: "'disease'"  
gene: "'gene'"
```

**MED12-related
intellectual disability
syndrome**

Disease: Intellectual
disability syndrome

Gene: MED12

TTN-related myopathy

Disease: myopathy

Gene: TTN

**MYPN-related
myopathy**

Disease: myopathy

Gene: MYPN

Disease series by gene

name:

```
text: '%s caused by mutation in %s'
```

vars:

- disease
- gene

Disease: myopathy

Gene: MYPN

Name: myopathy caused
by mutation in MYPN

annotations:

- `annotationProperty`: `exact_synonym`

```
text: '%s %s'
```

vars:

- gene
- disease

Exact synonym: MYPN
myopathy

Disease series by gene

- **annotationProperty:** exact_synonym

```
text: '%s related %s'
```

```
vars:
```

- gene
- disease

Disease: myopathy

Gene: MYPN

Exact synonym: MYPN
related myopathy

def:

```
text: Any %s in which the cause of the disease is a mutation in the %s gene.
```

```
vars:
```

- disease
- gene

Disease: myopathy

Gene: MYPN

Definition: Any myopathy
in which the cause of the
disease is a mutation in the
MYPN gene.

How to request new terms (Mondo IDs) & changes

1. GitHub tracker: New issue
2. Pick appropriate template
3. Fill in the information that is requested on the template below each header
4. Please include:
 - a. A definition in the proper format
 - b. Sources/cross references for synonyms
 - c. Your ORCID or the URL for your ClinGen working group
 - d. Add any additional comments at the end
5. Nicole will automatically be tagged
6. Please email Nicole or comment on the ticket (Nicole will be emailed) if you have any additional questions or need the ticket is high priority

Recommendations for GitHub tickets/new term requests

We appreciate your contributions to extending and improving Mondo



General Recommendations:

1. New term requests should not match existing terms or synonyms
2. Write a concise definition in the definition field. More info about writing definitions is [here](#)
3. Synonyms - please provide a source/cross-reference
4. Check OMIM for children classes

Formatting:

1. Preferred term labels should be lowercase (unless it is a proper name or abbreviation)
2. Write the request below the prompts on the template so the Markdown formatting displays properly
3. Synonyms should be lowercase (with exceptions above)
4. Definition source - if from PubMed, please use the format PMID:XXXXXX (no space)
5. Include the Mondo ID and label for the parent term
6. List the children terms with Mondo ID and label in a bulleted list

Writing Ontology Definitions

Guidelines for writing definitions in ontologies

Selja Seppälä

Ph.D., University of Geneva, Switzerland

Postdoctoral Associate, University of Florida, Gainesville, United States.

<http://seljaseppala.wordpress.com>

E-mail: sseppala@ufl.edu

Alan Ruttenberg

M.S. Massachusetts Institute of Technology

Director of Clinical Data Exchange, School of Dental Medicine,
University at Buffalo, United States.

<http://alan.ruttenbergs.com>

E-mail: alanruttenberg@gmail.com

Barry Smith

Ph.D., University of Manchester, United Kingdom.

Professor of Philosophy, University at Buffalo, Buffalo, United States.

<http://ontology.buffalo.edu/smith/>

E-mail: phismith@buffalo.edu

STRUCTURE OF A TEXTUAL DEFINITION

When used to refer to the natural language text of a definition, the term 'definition' itself can denote different forms: a sentence and a sentence fragment. Broadly, a definition has the canonical form X is a Y that Zs as in example (1) adapted from the definition of 'ligament' (synonym of 'skeletal ligament') in the Uberon multi-species anatomy ontology (UBERON).

(1) A **ligament** is a dense regular connective tissue connecting two or more adjacent skeletal elements.

Definitions in this form have a three-part structure:

1. a **definiendum** [X], i.e., the defined term;
2. a **definiens** [*a Y that Zs*], i.e., the part that expresses the definition content and that is called a **definition** in dictionaries;
3. a **copula** [*is*] that expresses an equivalence between definiendum and definiens.

Gene-based names

For example, 'GTP cyclohydrolase I deficiency'

****We are able to accomodate gene-based names****



SubClass Of 

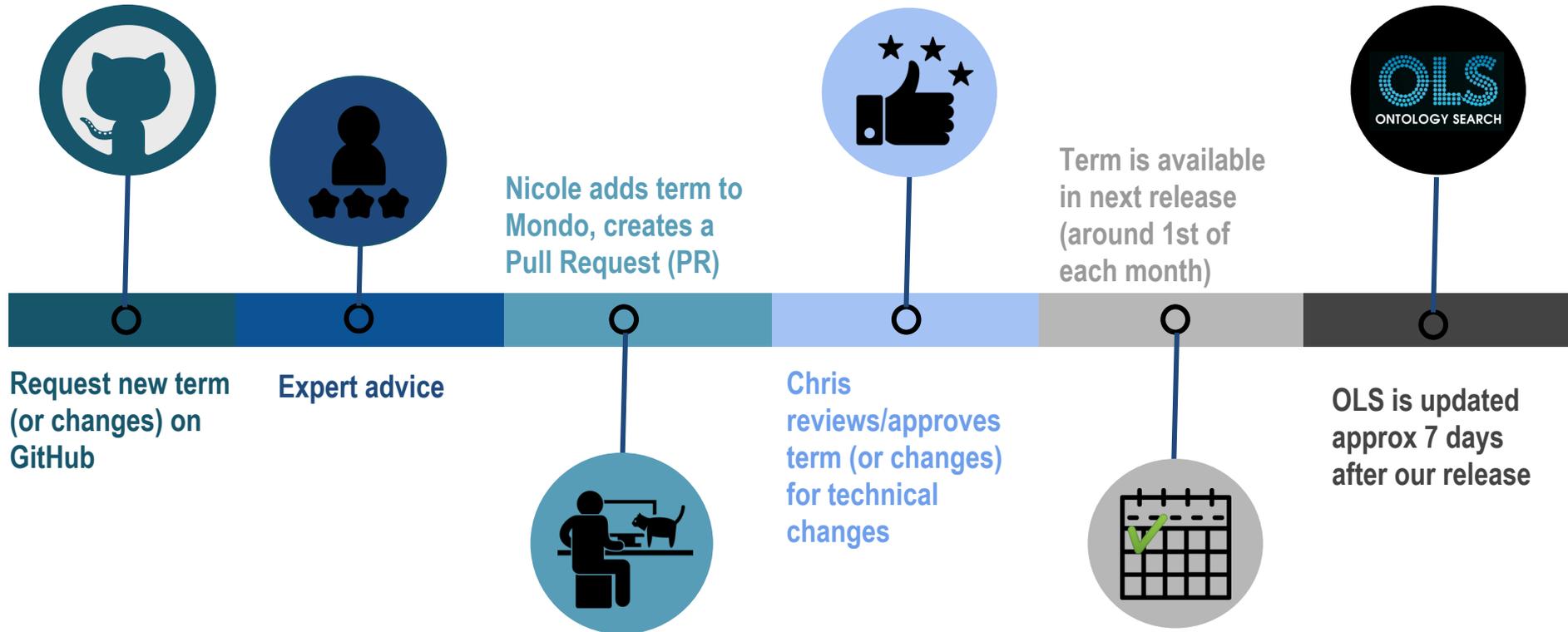
 'disease has basis in dysfunction of' some GCH1



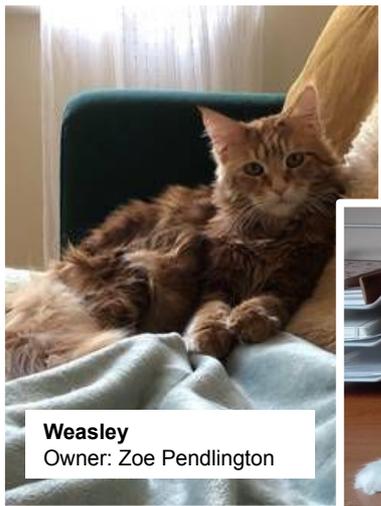
Synonym types

Scope	Exact	An exact match	E.g. hereditary Wilms' tumor exact synonym: familial Wilms' tumor
	Narrow	A more specific term	E.g. asthma narrow synonym: exercise-induced asthma
	Broad	A more general term	E.g. autoimmune hepatitis broad synonym, autoimmune liver disease
	Related	A word or phrase has been used synonymously with the primary term name in the literature, but the usage is not strictly correct	E.g. AGAT deficiency related synonym: disorder of glycine amidinotransferase activity
Type	Excluded	Some synonyms are annotated with EXCLUDE, e.g. "NOS" (not otherwise specified) synonyms. It is useful to have these in the edit version, but these are filtered on release.	
	Deprecated	We may also mark synonyms with DEPRECATED. E.g. all occurrences of "mental retardation" should be "intellectual disability"	

How long does it take to see changes in to Mondo?



Cats of Mondo



Weasley
Owner: Zoe Pendlington



Pepper
Owner: Melissa Haendel



Zisimos
Owner: Nico Matentzoglu



Nin
Owner: Sherri De Coronado



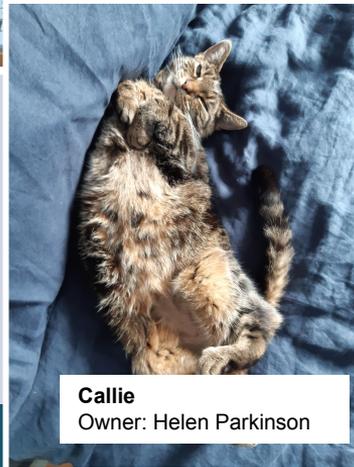
Blue
Owner: Paola Roncaglia



Sapphire
Owner: Nicole Vasilevsky



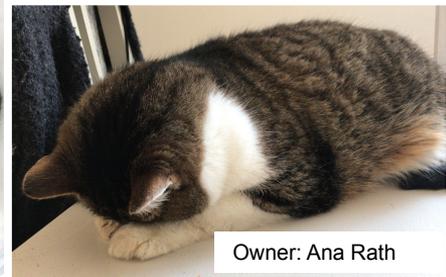
Meow
Owner: Sira Sarntivijai



Callie
Owner: Helen Parkinson



Zed and Trillian
Owner: Leigh Carmody



Owner: Ana Rath

Thank you!



Chris Mungall, LBNL
Email: cjmungall@lbl.gov
Twitter: @chrismungall
GitHub: cmungall



Nicole Vasilevsky, OHSU
Email: vasilevs@ohsu.edu
Twitter: @n_vasilevsky
GitHub: nicolevasilevsky

Dogs of Mondo



Luna
(grandma)-owner Anne Pariser



Neville



Copa and Molly
owner, Leigh Carmody

Funding:

Monarch is supported generously by a NIH Office of the Director Grant #5R24OD011883, as well as by NIH-UDP: HHSN268201350036C, HHSN268201400093P, NCI/Leidos #15X143.

Big thanks to our contributors!



Broad Institute

Samantha Baxter
Andrew Grant
Jessica Hekman
Madeline Hughes
Kate Megquier
Kathy Reinold
Rebecca Siegert

CHOP

Colin Ellis
Allison Heath
Ingo Helbig
Avi Kelman

CoRDS-Sanford

Austin Letcher

ClinGen

Larry Babb
Taylor Bingaman
Marina DiStefano
Jenny Goldstein
Brooke Palus
Heidi Rehm
Erin Riggs
Tam Sneddon
Courtney Thaxton
Matt Wright

EBI

Mélanie Courtot
Simon Jupp
David
Osumi-Sutherland
Zoë Pendlington
Paola Roncaglia

GARD

Gioconda Alyea
PJ Brooks
Maria Della Rocca
Janine Lewis
Anne Pariser
Andrea Storm

Monarch Initiative

Melissa Haendel
Leigh Carmody
Shahim Essaid
Nomi Harris
Nico Matentzoglu
Julie McMurry
Moni Munoz-Torres
Peter Robinson
Kent Shefchek
Anne Thessen
Aaron Zhang

NCIt

Gilberto Fragoso
Bron Kisler

NIH NCATS

Alice Chen
Eric Sid

NCBI

Donna Maglott

Johns Hopkins

Christopher Chute

NIH NHGRI

Robert Fullem
Morgan Similuk

NORD

Vanessa Boulanger

OMIM

Joanna Amberger
Ada Hamosh

Orphanet

Marc Hanauer
Annie Olry
Ana Rath

University of Colorado

Tiffany Callahan