FAIRifying biomedical ontology synergy

WSBO-2021: Workshop on Synergizing Biomedical Ontologies

Lynn M. Schriml, Ph.D.
University of Maryland School of Medicine
Institute for Genome Science

Funding: NIH/NHGRI Genomic Resource
U41 HG008735-01A1

DISEASE ONTOLOGY

https://www.disease-ontology.org
https://github.com/DiseaseOntology/HumanDiseaseOntology

UNIVERSITY of MARYLAND
SCHOOL OF MEDICINE
INSTITUTE FOR GENOME SCIENCES

DO Disease Ontology

Northwestern University

2003

(ICD, MESH, NCI)
Semantic Relationships, DAG
Persistent Identifiers: DOIDs

2008

DO R01 – DO Classification Website
OBO Foundry domain ontology

Term expansion
Outreach
Genetic Diseases

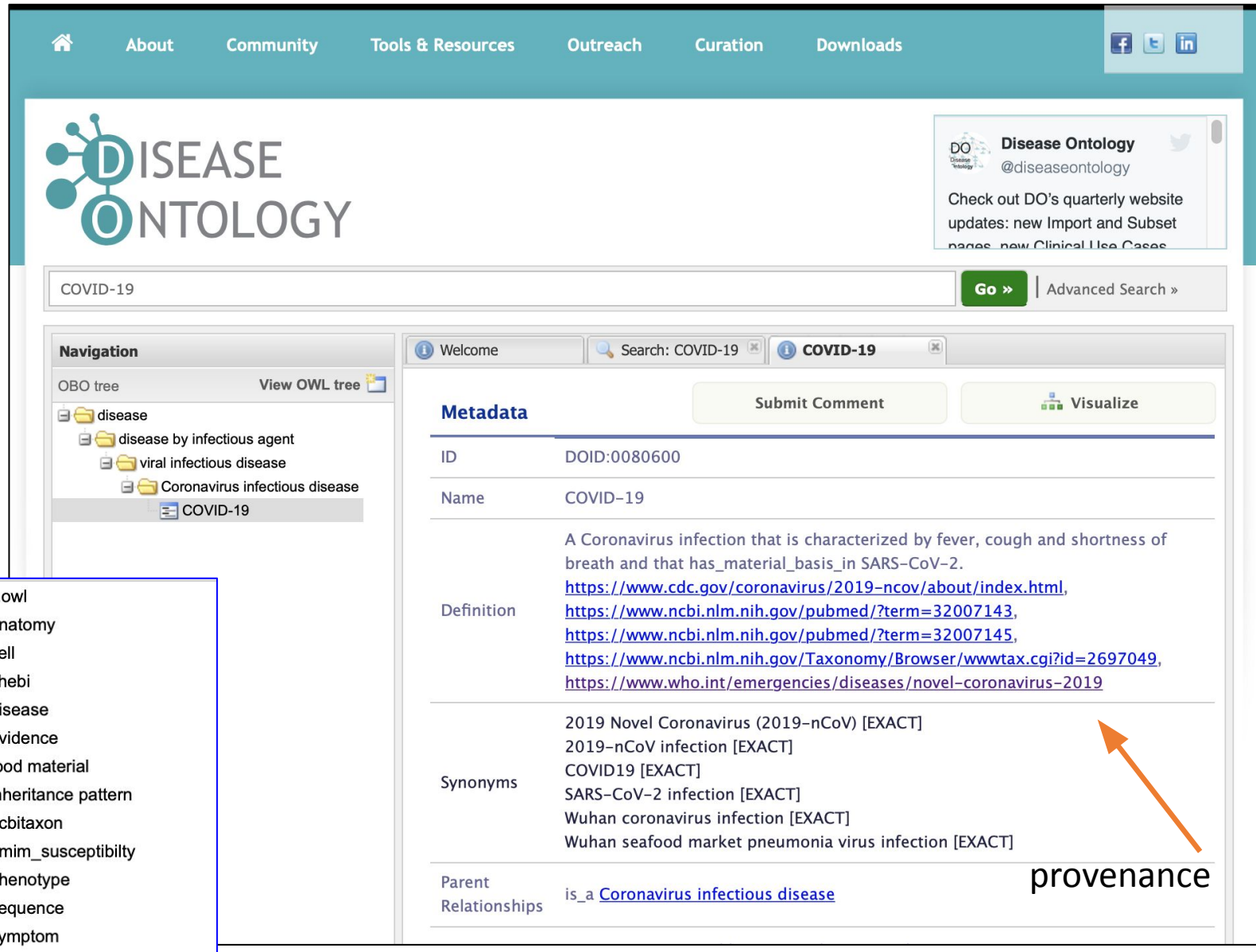
2017

NHGRI U41
Genomic Resource
CCBY → CC0 open license

Complex Disease

2021

10,791 disease terms
++ ~ 1000 new terms/yr



genetic & environmental drivers of disease, phenotypic features

DO xRefs

GARD:2001; MESH:3609; **OMIM**:5467; ICD9CM:2267; ICD10CM:3661; NCI:4732; **Orphanet**:1845; SNOMED: 5070

> 12,200 cited definition sources (PubMed IDs, databases)

Synergy via shared terms and IDs

DO Usage
(>280) databases & tools
(>50) ontologies

DO citing papers (> 780)

Somatic Cancer Pediatric Taskforce

Synergizing efforts

Community

Term requests
(ontology enrichment)

DO GitHub

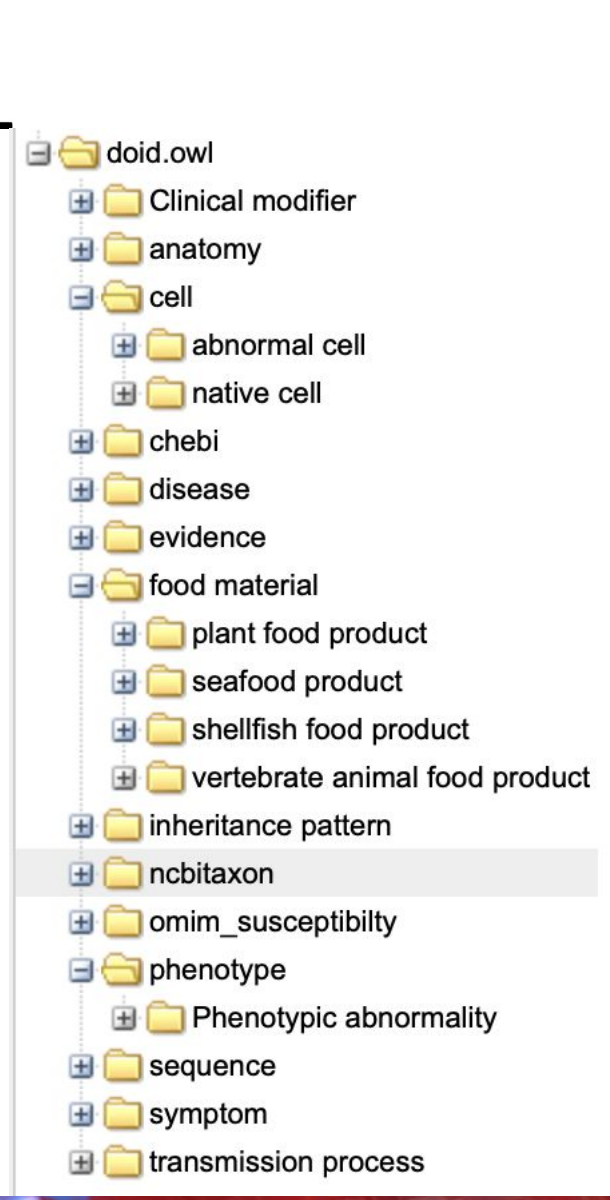
Imports OBO Foundry Ontologies

Axioms

- Codify disease to disease relationships
- Statements asserting SubClass
- Views/Mechanisms of Disease

Automated DO xref assessment

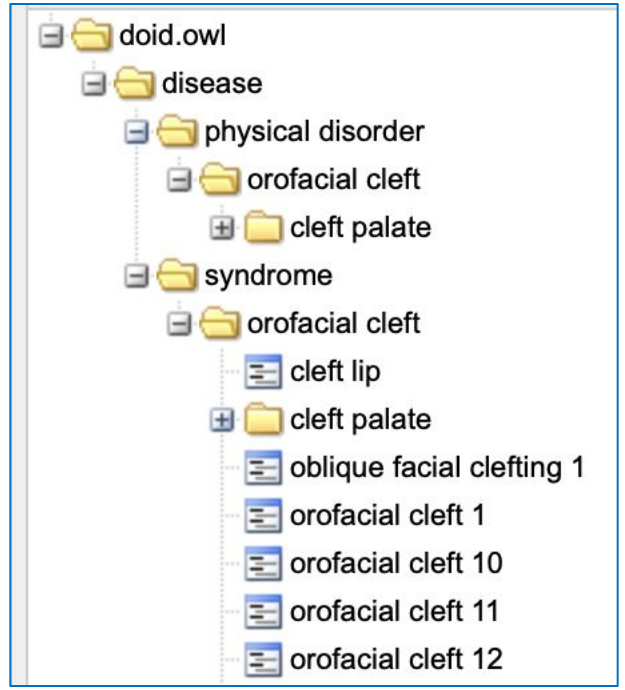
Uberon/Monarch – Bridging

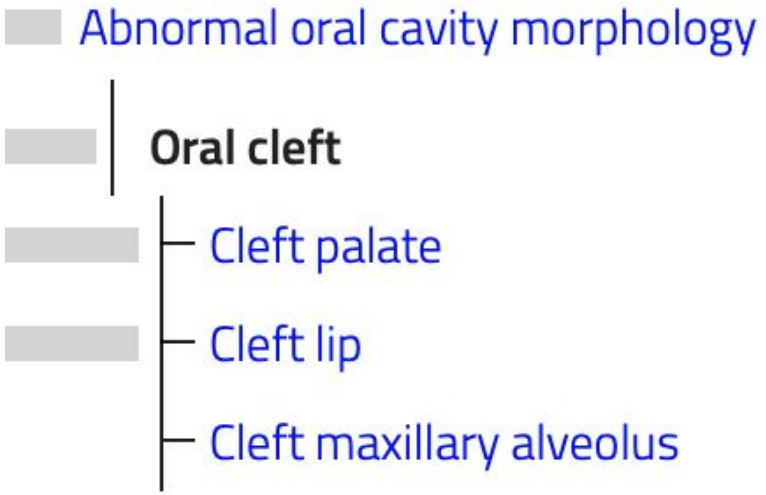


synergy between ontological domains

phenotypes, traits, syndromes, diseases, symptoms

domain fuzziness - recognizing usage differences
domain boundaries are fuzzy, often overlapping as knowledge expands and usage mature





Search: 'cleft lip-palate '
Results: 1,175 entries. « First | ‹ Previous | Next › | Last »

1: # 216100. JUBERG-HAYWARD SYNDROME; JHS
Cytogenetic location: 8p21.1
Matching terms: "lip palate", cleft, lippalate
▶ Phenotype–Gene Relationships ▶ ICD+ ▶ Links

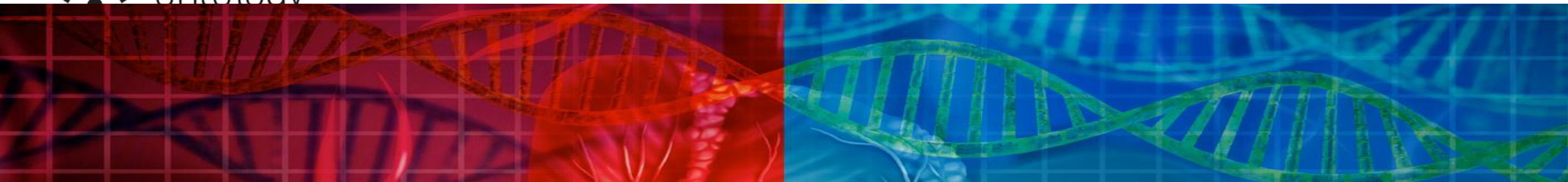
2: 201180. ACROFRONTOFACIONASAL DYSOSTOSIS 1
Matching terms: "lip palate", cleft, lippalate
▶ ICD+ ▶ Links

3: % 119530. OROFACIAL CLEFT 1; OFC1

human phenotype ontology

50 YEARS OMIM
Human Genetics Knowledge for the World

DISEASE ONTOLOGY



synergy within ontological domains

the OBO Foundry's "Scope" principle addresses the **ideal** of "non-overlapping and strictly-scoped content".

Open Biological and Biomedical Ontology (OBO) Foundry

Community development of interoperable ontologies for the biological sciences

Summary

The scope of an ontology is **the extent of the domain** or subject matter it intends to cover. The ontology must have a clearly specified scope and content that adheres to that scope.

Purpose

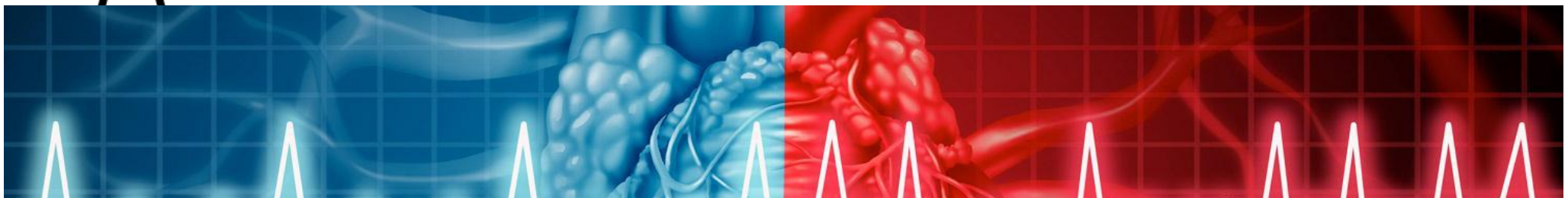
An in-scope ontology prevents overlaps between ontologies (duplication of terms), facilitates user searches for specific content, and **enables quick selection of ontologies of interest**, yet still allows for new terms to be created via combination of existing terms (cross-products).

Ontologies Across Biomedical Domains

OLS: Ontology Lookup Service: 264 ontologies

Expanded usage of a domain through adaptations or derivatives are "part in parcel of" the evolution of biomedical knowledge

Anatomy: 25
Traits: 25
Phenotype: 24
Disease: 19
Cell: 12
Chemical: 10
Protein: 9



The Disease Ontology is utilized to build other biomedical ontologies.

These application ontologies either reuse:

DO's terms or IDs as PURLS (unique URL based IDs, e.g. http://purl.obolibrary.org/obo/DOID_4), map to DOIDs as ontology cross references (xrefs), as synonyms or as annotations.

To date, the Disease Ontology is being utilized across 52 other biomedical ontologies:

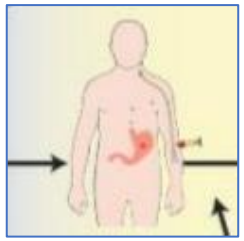
Antibiotic Resistance Ontology, Apollo Structured Vocabulary, BioAssay Ontology, Cardiovascular Disease Ontology, Cell Culture Ontology, Cell Line Ontology, Coronavirus Infectious Disease Ontology, Data Use Ontology, Dermatological Disease Ontology, Diabetes Mellitus Treatment Ontology, Drug-drug Interaction and Drug-drug Interaction Evidence Ontology, Drug Target Ontology, eagle-i resource ontology, eNanoMapper Ontology, Experimental Factor Ontology, FoodOn, Gender, Sex, and Sexual Orientation Ontology, Genomic Epidemiology Ontology, Hearing Impairment Ontology, Human Phenotype Ontology, Human Physiology Simulation Ontology, Hypertension Ontology, Infectious Disease Ontology, Influenza Ontology, Informed Consent Ontology, Mental Disease Ontology, Metabolomics Standards Initiative Ontology, Mondo Disease Ontology, Mouse Pathology Ontology, Neural Electro Magnetic Ontology, Non-Coding RNA Ontology, Obstetric and Neonatal Ontology, Ontology for Biobanking, Ontology for General Medical Science, Ontology for MIRNA Target - microRNA (miR) domain, Ontology for Biomedical Investigations, Ontology of Adverse Events, Ontology of Drug Adverse Events, Ontology of Host-Microbiome Interactions, Ontology of Host Pathogen Interactions, Ontology of Precision Medicine and Investigation, Pre-Eclampsia Ontology, Protein Ontology, Public Health Document Ontology, Quantitative Histopathology Image Ontology, Quantitative Imaging Biomarker Ontology, Sickle Cell Disease Ontology, TOXic Process Ontology, Vaccine Ontology, Translational Medicine Ontology, VEuPathDB ontology, Vaccine Informed Consent Ontology.

Collaboratively developed definition of “disease” (DOID:4 in the Human Disease Ontology)

A disease is a disposition (i) to undergo pathological processes that (ii) exists in an organism because of one or more disorders in that organism.

reuse

(with new ID, term name, revised definition)



Ontology for General Medical Science

OGMS:0000031

Experimental factor ontology

EFO:0000408

disease or disorder

MONDO:0000001

A semi-automatically constructed ontology that merges in multiple disease resources to yield a coherent merged ontology.

database cross reference

A disease is a disposition to undergo pathological processes that exists in an organism because of one or more disorders in that organism. [OGMS:0000031]

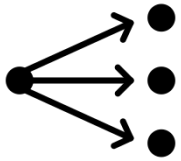
- OGMS:0000031 (MONDO:equivalentTo)
- NCIT:C2991 (MONDO:equivalentTo)
- DOID:4 (MONDO:equivalentTo)

NCIthesaurus: C2991: Disease or Disorder: Any abnormal condition of the body or mind that causes discomfort, dysfunction, or distress to the person affected or those in contact with the person.

Synergy Challenges

- Similar, but unreferenced terms
- Resource interoperability
- Minting new terms vs Term reuse
- Non-reuse of IRIs, definitions
- Scope creep
- **Asynchronous Update Cycles**

  UMLS – bi-annual; ontologies: weekly; monthly; quarterly
- **Lumping and Splitting**
  - Molecular Subtypes; Phenotypic Variation



- Ontologies create their own ID space

- Specific-Use-Cases

Promote best practices for synergizing ontologies

To improve the **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse
of ontology terms and biomedical data

Reference Ontologies

- community engagement
- assess and integrate new domain knowledge
- collaborate on new term requests

Application Ontologies or Reusing terms

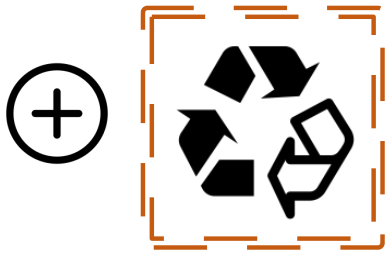
- request terms from reference ontologies
- ‘reuse’ the term ID, label and definition
  from a source ontology
- define source of reused terms

Promote best practices for synergizing ontologies

To improve the **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse
of ontology terms and biomedical data

ENHANCE SYNERGY

(1) Indicate which ontologies to 'Reuse' ?



(2) **Defining the provenance of 'reference ontology' IDs**

Instead of : creating xrefs for 'source biomedical ontologies'

Include: 'reference ontology IRIs'
via an "**imported from**" (IAO_0000412) Annotation Property

(3) (Within Domains) GitHub trigger, new term flag reference ontology to add term