Reclassification of Infectious Disease in the Mondo Disease Ontology

Nicole Vasilevsky [1], Sabrina Toro [1], Nico Matentzoglu [2], Dazhi Jiao [3], Melissa Haendel [1], Peter N Robinson [4] and Chris Mungall [5]

[1] *University of Colorado Denver, Anschutz Medical Campus, Aurora, CO USA*

[2] *Semanticly, Athens, Greece*

[3] *Johns Hopkins University, Baltimore, MD, USA*

[4] *Jackson Laboratory, Farmington, CT, USA*

[5] *Lawrence Berkeley National Laboratory, Berkeley, CA, USA*

Abstract

The Mondo Disease Ontology (Mondo) represents cross-species diseases, which integrates several source disease terminologies to represent cross-species diseases, and provides precise semantic mappings to the original sources. Mondo spans both rare and 'common' diseases, as well as monogenic, acquired, neoplasms, infectious diseases, and more. Mondo is a community resource and is continuously updated and iteratively curated. Recent efforts sought to improve the representation of viral infectious diseases in Mondo, to properly represent primary infections, diseases caused by reactivation of a latent virus, such as shingles and diseases caused by aftereffects of a primary infection such as long COVID-19. This included the addition of new classes and new relations (object properties), and the creation of new design patterns. The Mondo ontology files are available for download at: https://github.com/monarch-initiative/mondo.

Keywords

Infectious disease, viral infection, disease ontology, COVID-19, Dead Simple Ontology Design Patterns

1. Introduction

The Mondo Disease Ontology (Mondo) integrates multiple disease terminologies into a coherent logic-based ontology that provides precise semantic mappings between terms [1]. It was created via algorithmic equivalency determination using the k-BOOM algorithm [2], which merged several source terminologies and ontologies covering various aspects of disease across species, including the Online Mendelian Inheritance in Man (OMIM) (Mendelian disease) [3], Orphanet (rare diseases) [4], the neoplasm branch of the National Cancer Institute Thesaurus (NCIt) [5], GARD (rare diseases) [6], Disease Ontology (human diseases) [7], MedGen (genetic diseases) [8], and broad terminologies such as Medical Subject Headings (MeSH) [9], SNOMED Clinical Terminology [10], International Classification of Disease (ICD) [11], ICD-Oncology (ICD-O) [11], and numerous others.

Equivalence mapping between these disease resources is provided as database cross-references on each term using strict semantics, such as MONDO:equivalentTo, or MONDO:relatedTo, which allows for computational integration of associated data (such as phenotype associated with a disease). Dead Simple Design Patterns (DOSDPs) [12] are used to consistently and automatically apply equivalence and subclass axioms to categories of diseases, allowing automatic classification and inferencing. We developed a library of DOSDPs for Mondo and applied these patterns where applicable, to consistently axiomatize disease classes in the ontology.

Mondo currently contains over 20,000 disease terms classified based on anatomical structure affected, genetic etiology (e.g. MONDO:0003847 'Mendelian disease'), origin (e.g. MONDO:0005550 'infectious disease'), etc., under the MONDO:0000001 'disease or disorder' top-level class. The hierarchical classification allows for multi-inheritance, whereby terms can be classified under more than one parent. For example, the term 'respiratory tract infectious disease' (MONDO:0024355) is the child of both 'respiratory system disease' (MONDO:0005087, disease by anatomical system), and 'infectious disease' (MONDO:0005550).

Mondo is an OBO Foundry community-based ontology available in owl, obo, and JSON formats on GitHub (https://github.com/monarch-initiative/mondo), with new versions released monthly. The ontology is entirely manually curated by our expert curation and semantic development team (https://mondo.monarchinitiative.org/pages/contributors/) based on community feedback and iterative review by our primary development team.

The COVID-19 pandemic prompted the addition of new COVID-19 classes to Mondo, including a COVID-19 disease class and associated diseases. The addition of these new terms revealed the need for reclassification and organization of the 'viral infectious disease' hierarchy in Mondo, to more accurately represent the subtypes of viral infections. Here, we report the ongoing work to reorganize the 'viral infectious disease' branch of Mondo. We created a new classification of the infectious disease branch of Mondo, including the separation of the "infectious" and "post-infectious" into separate concepts. We created new terms to represent the concepts of 'primary' infectious disease versus disease arising from reactivation of a latent virus, and leverage the creation of new design patterns to classify subclass terms. Finally, we added COVID-19 specific terms following the newly created classification and design pattern.

2. Infectious Disease in Mondo

2.1   Revised Superclasses of 'viral infectious disease'

Infectious diseases in Mondo are classified by the infectious agent such as **bacterial infectious disease** (MONDO:0005113), **viral infectious disease** (MONDO:0005108), **fungal infectious disease** (MONDO:0002041), and **parasitic infectious disease** (MONDO:0005135), by the anatomical structure infected (e.g **ear infection** (MONDO:0021666), **eye infectious disease** (MONDO:0043885)), by the mode of transmission (eg. **vector-borne diseases** (MONDO:0100120)), and **coinfection** (MONDO:0100128).

Some infections, such as infections by COVID-19, induce not only an immediate infectious disease but also diseases caused by after-effects or chronic complications resulting from the primary infection. We separated these two different concepts into two classes: '**infectious disease**' (MONDO:0005550) and '**post-infectious disorder**' (MONDO:0021669), respectively.   A new higher-level classification was introduced to serve as a grouping class, **'infectious disease or post-infectious disorder'** (MONDO:0100336), under the top-level term 'disease or disorder' (MONDO:0000001) [Figure 1].

We focused on the 'viral infectious disease' branch of the ontology, and added the new grouping class, **'viral disease or post-viral disorder'** (MONDO:0100321) under 'infectious disease and post-infectious diseases' (MONDO:0100336). Children of this class include the existing terms **'viral infectious disease'** (MONDO:0005108), and **'post-viral disorder'** (MONDO:0021674), which groups terms such as 'long COVID-19' (MONDO:0100233) [Figure 1].

2.2 Primary Infection versus Reactivation of Latent Virus

Some viruses, such as herpesvirus and varicella zoster, are known to be able to enter a latent phase after the original infection. The virus, which DNA had been incorporated into the host cell's genome, is dormant during this latent phase but can reactivate later after the original infection leading to a new disease [13]. For example, varicella zoster infection can cause chickenpox after the original infection, and shingles after reactivation of the latent virus.

These concepts were represented in the ontology: **'viral infectious disease'** (MONDO:0005108) was subclassified into **'primary viral infectious disease'** (MONDO:0100329), referring to the original infection, and **'disease arising from reactivation of latent virus'** (MONDO:0100330), which refers to disease that arises from reactivation of a virus from a latent phase to a lytic phase [14].

To leverage the power of consistent automatic classification and inferencing offered by DOSDPs, we created new design patterns using the newly created relations '**disease has primary infectious agent'**(MONDO:0100332) or '**disease caused by reactivation of latent infectious agent**' (MONDO:0100333) that use the taxon of the infectious agent from the NCBI Taxonomy in the equivalence axiom to refer to 'primary' or 'arising from reactivation of latent virus'' infectious diseases, respectively. For example, herpes zoster (synonym: shingles) (MONDO:0005609) has the equivalence axiom:

```
'infectious disease'
 and ('disease caused by reactivation of latent infectious agent' some
'Human alphaherpesvirus 3')
```
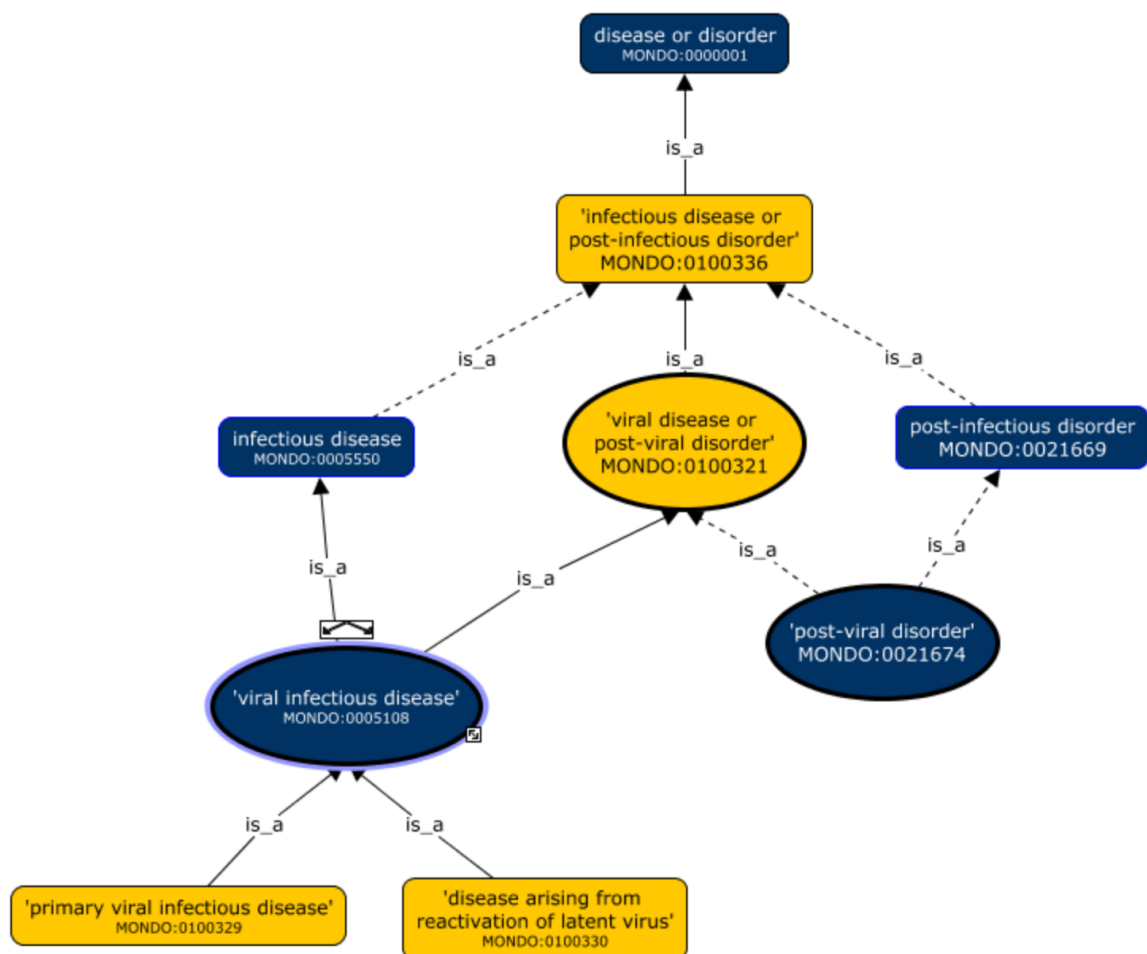

Figure 1: Representation of infectious disease and viral diseases classification in the Mondo Disease Ontology. A new 'infectious disease or post-infectious disorder' (MONDO:0100336) was created as a grouping class of 'infectious disease' (MONDO:0005550) and 'post-infectious disorder' (MONDO:0021669). A child of this new class was created specifically for viral disorder: 'viral disease or post-viral disorder (MONDO:0100321), which subsumes both 'viral infectious disease' (MONDO: 0005108) and 'post-viral disorder' (MONDO:0021674). The terms 'primary viral infectious disease' (MONDO:0100329) and 'disease arising from reactivation of latent virus' (MONDO:0100330) were created as subclasses of 'viral infectious disease'. Yellow filled boxes - newly added terms, blue filled

boxes - pre-existing terms, solid lines - asserted superclass Of relationship, dotted lines - inferred classes. Oval-shaped boxes indicate terms that overlap in Figure 1 and Figure 2.

2.3 Representation of COVID-19 in Mondo

After infection with SARS-CoV-2, patients presented not only symptoms of COVID-19 but also associated long term and/or chronic COVID-19 associated disorders, including the newly described disease 'long COVID-19'. To properly represent these new infectious diseases, terms referring to virus infection or its associated long term diseases in which the virus is SARS-CoV-2 were created: **'COVID-19'** (MONDO:0100096), which is an inferred descendent of 'primary viral infectious disease' (MONDO:0100319), and **'post-COVID-19 disorder'** (MONDO:0100320), an inferred child of 'post-viral disorder' (MONDO:0021674). These 2 terms share the asserted common parent: **'SARS-CoV-2-related disease'** (MONDO:0100318), which is a child of **'viral disease or post-viral disorder'** (MONDO:0100321). The terms **'long COVID-19'** (MONDO:0100233), '**COVID-19-associated MIS-A'** (MONDO:0100319), and '**COVID-19-associated MIS-C'** (MONDO:0100163) were created as children of '**post-COVID-19 disorder** (MONDO:0100320), representing the diseases reported in adults and children post-COVID-19 infection [15-20] [Figure 2].

As described in the previous section, infectious disease terms conform to a design pattern that specifies that the disease is an 'infectious disease' and 'has a primary infectious agent' that is an organism from the NCBI Taxonomy. In this case of these COVID-19 terms, the equivalence axiom for COVID-19 is:

```
'infectious disease' and ('disease has primary infectious agent'
some 'Severe acute respiratory syndrome coronavirus 2')
```
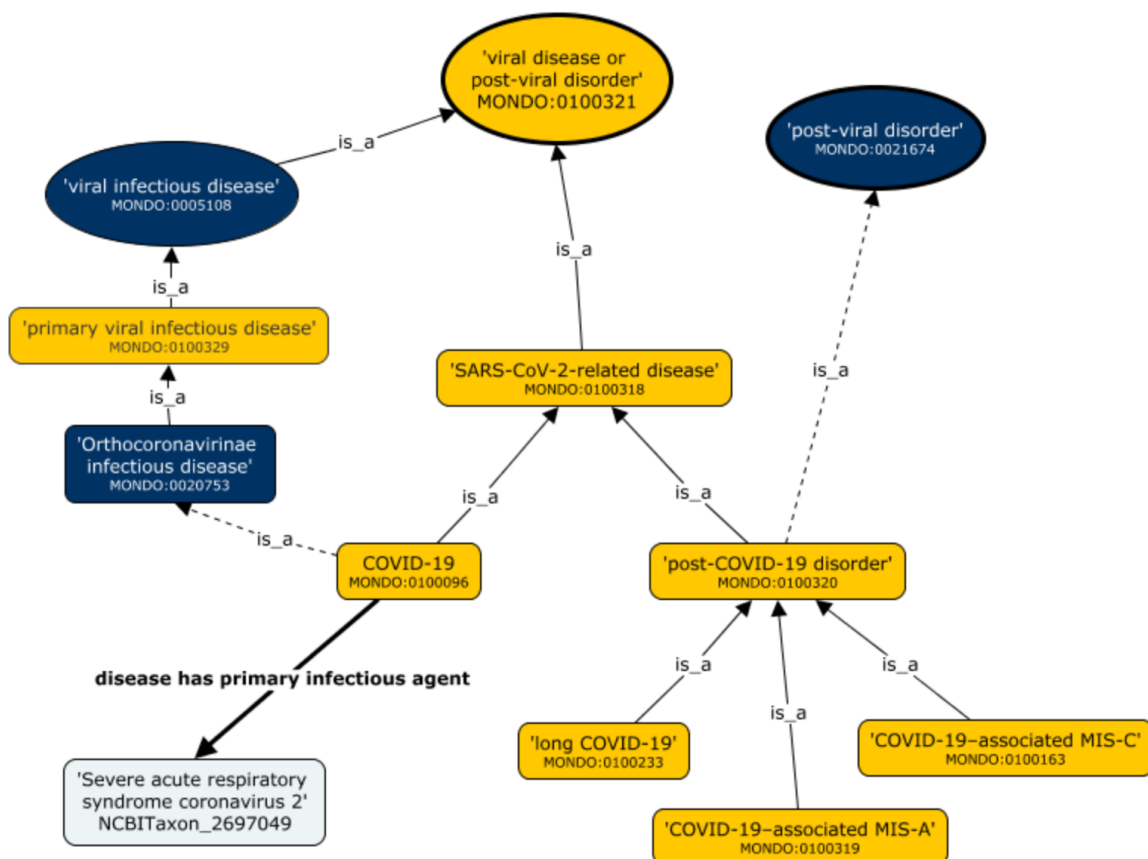

Figure 2: Representation of 'SARS-CoV-2-related disease' classification in Mondo Disease Ontology. 'SARS-CoV-2-related disease' (MONDO:0100318), child of 'viral disease of post-viral disorder' (MONDO:0100321), is a grouping class for 'COVID-19' (MONDO:0100096), inferred 'primary viral infectious disease' (MONDO: 0100329), and 'post-COVID-19 disorder' (MONDO:0100320), inferred

'post-viral disorder' (MONDO:0021674). 'COVID-19' is defined by an equivalence axiom relating this term to the NCBITaxon 'Severe acute respiratory syndrome coronavirus 2' via the 'disease has primary infectious agent' relation. 'post-COVID-19 disorder' encompasses several subclasses: 'long COVID-19' (MONDO:0100233), 'COVID-19-associated MIS-A' (MONDO:0100319), and 'COVID-19-associated MIS-C' (MONDO:0100163). Yellow filled boxes - newly added terms, blue filled boxes - pre-existing terms, solid lines - asserted superclass Of relationship, dotted lines - inferred classes, bold line and text - object property relationship ('disease has primary infectious agent').

3. Conclusion

Here we reported the enhanced representation of viral infectious diseases in Mondo. We updated the ontology to include the concepts of immediate versus the long-term effect of an infection and primary viral infection versus disease arising from reactivation of latent virus. We created new patterns and introduced new relationships to support these patterns allowing the automatic classification of infectious diseases into these new classes. This work is ongoing and focused exclusively on improving the classification of viral infectious diseases. These improvements will serve as a template to further remodel the infectious disease branch in Mondo, and future efforts will review the bacterial, parasitic and fungal infectious disease terms. This model of infectious disease may be useful as a template for other ontologies that capture this type of disease information. Mondo is an open-source ontology and community input is always welcome. Please submit your feedback or requests for changes in our GitHub issue tracker (https://github.com/monarch-initiative/mondo/issues).

4. Acknowledgments

5. References

[1] Vasilevsky N, Essaid S, Matentzoglu N, Harris NL, Haendel M, Robinson P, Mungall CJ. Mondo Disease Ontology: harmonizing disease concepts across the world. ICBO Proceedings 2020. http://ceur-ws.org/Vol-2807/abstractY.pdf

[2] Mungall CJ, Koehler S, Robinson P, Holmes I, Haendel M. k-BOOM: A Bayesian approach to ontology structure inference, with applications in disease ontology construction. bioRxiv. 2019. p. 048843. doi:10.1101/048843

[3] https://omim.org/

[4] https://www.orpha.net/consor/cgi-bin/index.php?lng=EN

[5] https://ncithesaurus.nci.nih.gov/ncitbrowser/

[6] https://rarediseases.info.nih.gov/

[7] Bello SM, Shimoyama M, Mitraka E, Laulederkind SJF, Smith CL, Eppig JT, Schriml LM. Disease Ontology: improving and unifying disease annotations across species. Dis Model Mech. 2018 Mar 12;11(3):dmm032839. doi: 10.1242/dmm.032839. PMID: 29590633; PMCID: PMC5897730.

[8] http://www.medgenehr.com/

[9] https://meshb.nlm.nih.gov/search

[10] https://www.snomed.org/

[11] https://apps.who.int/iris/handle/10665/42980

[12] Osumi-Sutherland, D., Courtot, M., Balhoff, J. et al. Dead simple OWL design patterns. J Biomed Semant 8, 18 (2017). https://doi.org/10.1186/s13326-017-0126-0

[13] Traylen, Christopher M et al. "Virus reactivation: a panoramic view in human infections." Future virology vol. 6,4 (2011): 451-463. doi:10.2217/fvl.11.21

[14] Traylen CM, Patel HR, Fondaw W, Mahatme S, Williams JF, Walker LR, Dyson OF, Arce S, Akula SM. Virus reactivation: a panoramic view in human infections. Future Virol. 2011 Apr;6(4):451-463. doi: 10.2217/fvl.11.21. PMID: 21799704; PMCID: PMC3142679.

[15] Morris SB, Schwartz NG, Patel P, Abbo L, Beauchamps L, Balan S, Lee EH, Paneth-Pollak R, Geevarughese A, Lash MK, Dorsinville MS, Ballen V, Eiras DP, Newton-Cheh C, Smith E, Robinson S, Stogsdill P, Lim S, Fox SE, Richardson G, Hand J, Oliver NT, Kofman A, Bryant B, Ende Z, Datta D, Belay E, Godfred-Cato S. Case Series of Multisystem Inflammatory Syndrome in Adults Associated with SARS-CoV-2 Infection - United Kingdom and United States, March-August 2020. MMWR Morb Mortal Wkly Rep. 2020 Oct 9;69(40):1450-1456. doi: 10.15585/mmwr.mm6940e1. PMID: 33031361; PMCID: PMC7561225.

[16] Whittaker E, Bamford A, Kenny J, Kaforou M, Jones CE, Shah P, Ramnarayan P, Fraisse A, Miller O, Davies P, Kucera F, Brierley J, McDougall M, Carter M, Tremoulet A, Shimizu C, Herberg J, Burns JC, Lyall H, Levin M; PIMS-TS Study Group and EUCLIDS and PERFORM Consortia. Clinical Characteristics of 58 Children With a Pediatric Inflammatory Multisystem Syndrome Temporally Associated With SARS-CoV-2. JAMA. 2020 Jul 21;324(3):259-269. doi: 10.1001/jama.2020.10369. PMID: 32511692; PMCID: PMC7281356.

[17] Ramcharan T, Nolan O, Lai CY, Prabhu N, Krishnamurthy R, Richter AG, Jyothish D, Kanthimathinathan HK, Welch SB, Hackett S, Al-Abadi E, Scholefield BR, Chikermane A. Paediatric Inflammatory Multisystem Syndrome: Temporally Associated with SARS-CoV-2 (PIMS-TS): Cardiac Features, Management and Short-Term Outcomes at a UK Tertiary Paediatric Hospital. Pediatr Cardiol. 2020 Oct;41(7):1391-1401. doi: 10.1007/s00246-020-02391-2. Epub 2020 Jun 12. PMID: 32529358; PMCID: PMC7289638.

[18] Nakra NA, Blumberg DA, Herrera-Guerra A, Lakshminrusimha S. Multi-System Inflammatory Syndrome in Children (MIS-C) Following SARS-CoV-2 Infection: Review of Clinical Presentation, Hypothetical Pathogenesis, and Proposed Management. Children (Basel). 2020 Jul 1;7(7):69. doi: 10.3390/children7070069. PMID: 32630212; PMCID: PMC7401880.

[19] Carfì A, Bernabei R, Landi F; Gemelli Against COVID-19 Post-Acute Care Study Group. Persistent Symptoms in Patients After Acute COVID-19. JAMA. 2020 Aug 11;324(6):603-605. doi: 10.1001/jama.2020.12603. PMID: 32644129; PMCID: PMC7349096.

[20] Nath A. Long-Haul COVID. Neurology. 2020 Sep 29;95(13):559-560. doi: 10.1212/WNL.0000000000010640. Epub 2020 Aug 11. PMID: 32788251.