Coordinating Coronavirus Research: The Virus Infectious Disease Ontology

John Beverley[1], Shane Babcock[2], Barry Smith[3], Yongqun He[4], Eric Merrell[1,3], Lindsay Cowell[5], Regina Hurley[6], Sebastian Deusing[7]

[1] *Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA*
[2] *Department of Philosophy, Niagara University, Lewiston, NY, USA*
[3] *Department of Philosophy, University at Buffalo, Buffalo, NY, USA*
[4] *University of Michigan Medical School, Ann Arbor, MI, USA*
[5] *University of Texas Southwestern Medical School, Dallas TX, USA*
[6] *Department of Philosophy, Northwestern University, Evanston, IL, USA*
[7] *Loyola University, Chicago, Il, USA*

Abstract

The COVID-19 pandemic prompted immense work on the investigation of the SARS-CoV-2 virus. Ontologies – structured, controlled, vocabularies – are designed to support consistency of interpretation, and thereby to prevent the development of data silos. This paper describes how ontologies are serving this purpose in the virus research domain, following the principles of the Open Biological and Biomedical Ontology (OBO) Foundry and drawing on the resources of the Infectious Disease Ontology (IDO) Core. We report the development of the Virus Infectious Disease Ontology (VIDO), a reference ontology extending IDO Core covering viral infectious diseases. We examine newly minted terms, showcase ontology term reuse, and illustrate the use of VIDO by annotating selections from recent life science research on SAR-CoV-2, in the interest of supporting machine learning projects.

Keywords

Virus, SARS-CoV-2, coronavirus, COVID-19, applied ontology, infectious disease ontology, CIDO

1. Introduction

The volume of data collected by life-science researchers, the speed at which it is generated, range of its sources, quality, accuracy, and need for assessment of usefulness, results in complex, multidimensional datasets [1,2], often annotated using discipline- or institution-specific terminologies and coding systems that lead to data silos. These undermine interoperability, meta-data analysis, pattern identification, and discovery across disciplines [3]. The value of cross-discipline meta-data analysis is evident during the present pandemic [4,5].

Ontologies – interoperable, logically well-defined, controlled vocabularies representing common entities and relations across disciplines using consensus terminologies – constitute a well-known solution to the formation of data silos. The need for rapid analysis of evolving datasets representing coronavirus research motivated the development of the Virus Infectious Disease Ontology (VIDO) comprised of textual definitions for terms and relations and logical axioms supporting automated consistency checking, querying over datasets, and interoperability with other ontologies. VIDO is an extension of the widely used Infectious Disease Ontology Core (IDO Core) [7], comprised of terminological content common to all investigations of infectious disease. VIDO extends IDO with terms specific to the domain of infectious diseases caused by viruses and provides a foundation for ontologies representing specific viral infectious

ICBO 2022, September 25-28, 2022, Ann Arbor, MI, USA
EMAIL: johnbeverley2021@u.northwestern.edu (A.1); babcock8@buffalo.edu (A.2); ifomis@gmail.com (A.3); yongqunh@med.umich.edu (A.4); Lindsay.Cowell@utsouthwestern.edu (A.5); ReginaHurley2023@u.northwestern.edu (A.6); sduesing@luc.edu (A.7)
ORCID: 0000-0002-1118-1738 (A. 1); 0000-0003-0798-114X (A. 2); 0000-0003-1384-116X (A. 3); 0000-0001-9189-9661 (A.4); 0000-0003-1617-8244 (A.5)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3805/ICBO-2022_paper_3877.pdf

diseases such as the Coronavirus Infectious Disease Ontology (CIDO), which is itself an extension of VIDO covering coronavirus infectious diseases such as COVID-19. The most recent version of VIDO can be found at the National Center for Biomedical Ontology Bioportal [8].

VIDO was developed in collaboration with relevant domain experts and drawing on the expertise of the IDO developers to ensure alignment with OBO Foundry principles [9], supporting interoperability with Foundry ontologies [10]. Development of VIDO is transparent, with discussions available on GitHub [11]. VIDO term additions are driven by the needs of researchers investigating viruses and nearby domains, and so remains sensitive to evolving knowledge.

2. Methods

VIDO and is formally represented in OWL 2 Web Ontology Language. VIDO was developed using the Protégé-OWL editor and tested against automated reasoners such as HermiT and Pellet [12,13]. Axioms underwriting VIDO were translated into a decidable fragment of first-order logic readable by the Mace4 [14] model checker, which allowed for manual graphical inspection of classes of models constrained by the asserted axioms. An automated proof-checker Prover9 bundled with Mace4 was used to validate expected theorems while refining axiom models.

2.1 OBO Foundry Alignment

Ontologies are widely used in bioinformatics, supporting data standardization, integration, sharing, reproducibility, and automated reasoning. The Gene Ontology (GO), for example, maintains species-neutral annotations of gene products and functions, and – since its inception in 1998 – has inspired an explosion of biomedical ontologies covering all domains of the life sciences [15,16,17]. These early developments led to worries, however, that data silos – the very problem ontologies were designed to address – might reemerge as researchers developed ontologies using concepts local to their discipline. By 2007, the Open Biomedical and Biological Ontologies (OBO) Foundry was created to provide guidance for ontology developers and promote alignment and interoperability. OBO Foundry design principles require ontologies: use a well-specified syntax unambiguous with a common space of identifiers, be openly available in the public domain, be modularly developed with a specified scope in a collaboration with ontologists covering nearby domains and conform to a common top-level architecture. The OBO library consists of over 250 ontologies, including some externally developed ontologies such as the NCI Thesaurus [18] and the NCBI Taxonomy [19], and some constructed *ab initio* to satisfy OBO principles. At its core is Basic Formal Ontology (BFO), a top-level ontology covering general classes such as material entity, quality, process, and role [20,21,22], which provides the architecture referred to in the last Foundry principle. BFO is, moreover, an ISO/IEC approved standard 21838-2 [23].

Where BFO is domain-neutral, OBO Foundry ontologies represent types of entities in more specific domains, using terms such as disease, cell division, surgical procedure, and so forth. Each domain ontology is constructed using a methodology for formulating definitions through a process of downward population from BFO. The resulting alignment with BFO, and the conformance to OBO Foundry principles, fosters integration across ontologies. VIDO was designed with alignment and conformance in mind. Development followed metadata conventions adopted by many OBO Foundry ontologies. These conventions require that every term introduced into the ontology has a unique IRI, textual definitions, definition source, designation of term editor(s), and preferred term label. In the interest of coordinating development with existing OBO ontologies, VIDO developers imported terms where possible from OBO library ontologies and constructed logical definitions using imported terms. Development was guided by best practices for definition construction [24]. New terms were introduced when needed, after consultation with domain experts, relevant literature, and examination of the OBO library.

2.2 Hub and Spokes Approach

VIDO follows the "hub and spokes" methodology [25] for ontology development. VIDO is a spoke ontology, extending from the Infectious Disease Ontology Core (IDO Core) as its hub. IDO Core is an OBO ontology consisting of terms, relations, natural language definitions and associated logical axioms representing phenomena common across infectious diseases research [26]. IDO Core has long provided a base from which more specific infectious disease ontologies extend and has been recently updated to keep pace with scientific and top-level architecture changes [27]. Extensions of IDO Core covering specific infectious diseases are created, first, by importing needed terms from IDO Core and other OBO Foundry ontologies, and second, by constructing the domain-specific terms where needed to adequately characterize entities in the relevant domain. Examples include the Brucellosis Infectious Disease Ontology (IDOBRU) the Influenza Infectious Disease Ontology (IDOFLU), and more recently the Coronavirus Infectious Disease Ontology (CIDO), each semantically interoperable with other OBO library ontologies [28,29,30].

VIDO was designed to occupy the ontological space between such virus-specific ontologies and IDO Core. As a result, more specific virus-related ontologies such as CIDO and IDOFLU are being curated to extend directly from VIDO, rather than IDO Core.

3. Results

VIDO takes IDO Core as its starting point, but also imports terms relevant to the domain of viruses from other OBO Foundry ontologies, such as GO, the Ontology for General Medical Science (OGMS) and the Ontology for Biomedical Investigation (OBI) [31].

3.1 Virus

Like IDO Core, VIDO imports from OGMS:

> disorder $=_{def}$ Material entity that is a clinically abnormal part of an extended organism.

A part of a material entity is "clinically abnormal" if it is not expected in the life plan for entities of the relevant type and is causally linked to elevated risk – that is, risk exceeding some threshold – of illness, death, or disfunction [32]. Organism is imported from OBI:

> organism $=_{def}$ Object that is an individual living system, such as animal, plant, bacteria, or virus, that is capable of replicating or reproducing, growth and maintenance in the right environment. An organism may be unicellular or made up, like humans, of many billions of cells divided into specialized tissues and organs.

Here we run into the first of several ontological puzzles that emerged while developing VIDO. On the one hand, this definition aligns with common use of the term "organism" among researchers insofar as instances are cellular entities [33,34]. On the other hand, the textual definition includes viruses among instances, which are in every case acellular. Debates [35] among ontology developers over organism have resulted in deprecation of the OBI term in favor of a term from the Common Anatomy Reference Ontology [36]: organism or virus or viroid. This avoids the preceding worries but reveals two more. First, introducing disjunctive classes suggests closure over instances. We should, however, avoid suggesting that classes – especially in biological domains – are settled results of scientific discoveries. Second, this disjunctive class leads naturally to debates over whether viruses are alive, since it classifies viruses alongside paradigmatic living entities. Decades of discussion have not resolved this question [37,38,39], and it is not obvious we need an answer for the purposes of ontological modelling. Rather than introduce an *ad hoc* disjunctive class, IDO Core and VIDO developers collaborated to add the following disjoint sibling class of organism to IDO Core:

> acellular structure $=_{def}$ Object consisting of an arrangement of interrelated acellular parts

forming an acellular biological unit that is able to initiate replication of the structure in a host.

Imported to VIDO as the parent class of the term virus. The term virus is imported from the NCBITaxon [40], alongside other terms representing entities investigated by virologists, such as prion, viroid, and satellite.

The NCBITaxon provides an exhaustive list of life science terms but has its limitations. First, with respect to virus terms NCBITaxon appears to align with the International Committee on Taxonomy of Viruses (ICTV). The ICTV, however, lacks systematic classification criteria and consequently leaves several viruses unclassified [41,42]. Second, when NCBITaxon is combined with automated importing tools such as Ontofox [43], ontology developers (for example, the developers of the Schistosomiasis Ontology [27]) sometimes import entire ICTV hierarchies –from kingdom to species – resulting in large, unwieldy, taxonomies obscuring classes of interest. Third, NCBITaxon provides few textual definitions for terms. To align with OBO Foundry metadata conventions and best practices, definitions are needed for virus and its subclasses.

Standard definitions of "virus" provide a starting point, but caution is needed. Viruses are often described as obligate pathogens [44,45], since virus replication requires host machinery for production and assembly of viral components. However, defining a class virus solely in terms of what viruses typically do runs the risk of overlooking what viruses are, materially speaking. Compare: *Homo sapiens* are obligate aerobes, but this is no definition of the class. Insofar as we are defining the material entity virus, better to attend to genetic and structural components common to all viruses, and best to define the material entity in a way that captures obligate pathogenicity. VIDO defines:

virus $=_{def}$ Acellular structure with RNA or DNA genetic material which uses host metabolic resources for RNA or DNA replication.

Subclasses of virus in VIDO are imported from NCBITaxon in alignment with the Baltimore Classification [46], which groups viruses into seven exhaustive classes based on genetic structure. For example, a subclass of virus:

positive-sense single-stranded RNA virus $=_{def}$ Virus with genetic material encoded in single-stranded RNA that can be translated directly into proteins.

Which provides the parent class for:

coronavirus $=_{def}$ Positive-sense single-stranded RNA virus with a helically symmetrical nucleocapsid, lipid bilayer viral envelope, and surface spike peplomers.

Figure 1 illustrates how the Baltimore Classification appears in the Protégé visualization of the class positive-sense single-stranded RNA virus, supplemented by a textbook image of the seven viral replication pathways underwritten by virus genetic differences.

Figure 1: Baltimore Classification in Protégé Editor

By incorporating the Baltimore Classification, VIDO provides developers of more specific virus classes within a succinct, navigable, ontological structure which refers to viral replication pathways, and so to virus pathogenicity.

VIDO subclasses of virus include those common in virology research, such as bacteriophage – viruses that infect bacteria – virophage – viruses that infect viruses – oncovirus – viruses that cause cancer – and mycovirus – viruses that infect fungi. One subclass of virus imported from NCBITaxon is the VIDO term:

virion $=_{def}$ Virus that is in its assembled state consisting of genomic material (DNA or RNA) surrounded by coating molecules.

Some researchers use “virion” and “virus” synonymously [47]. Some define “virion” so that instances only exist outside host cells [48] or distinguish virions outside host cells from those inside host cells, calling the former “mature virions.” Some claim “virion” is best understood as analogous to a sperm cell [49,50]. Ontologically speaking, one might model the relationship between a virus and its virion in a variety of ways: virion is to a virus as human infant is to human, or as human student is to human, or as human gamete is to human. Treating virions as akin to gametes is uncommon among researchers. Between the remaining options, we adopt the first, treating virion as a type of virus, since adopting the alternative would suggest a virion is just a virus in a specific context, with a specific role.

Some viruses do not replicate faithfully, perhaps resulting in genetically distinct mutants or – in extreme cases – an inactive aggregate of virion components. Virus mutations may undermine host immune system recognition of viral threats, as evidenced by the difficulty in developing vaccines for certain influenza strains. Too many mutations, however, and a virus may lose its ability to replicate, an observation used in development of treatments for polio and hepatitis C which exacerbate respective virus mutations [51,52]. VIDO provides the term:

> disordered virus $=_{def}$ Acellular structure having some clinically abnormal arrangement of viral components (e.g. viral capsid, viral DNA/RNA)

Viruses falling in this class may be associated with diseases much different from those of the clinically normal variety. Terms for viral components are imported to VIDO. From GO, VIDO imports viral nucleocapsid, viral capsid, capsomere, from the Chemical Entities of Biological Interest (ChEBI) ontology [53] the terms nucleic acid and ribonucleic acid, and from the Protein Ontology [54] the terms protein and viral protein are imported.

3.2 Infectious Structure

The term “pathogen” is indexed to species or to stages in the developmental cycle of a species. A given virus may engage in mutual symbiosis with one species, while exhibiting pathogenic behavior towards others [55,56]. Mature plants are often susceptible to different pathogens than developing plants [57,58]. We capture virus pathogenicity in VIDO in steps. From IDO Core, we import dispositions borne by pathogens and infectious agents, as follows:

> pathogenic disposition $=_{def}$ Disposition borne by a material entity to establish localization in or produce toxins that can be transmitted to an organism or acellular structure, either of which may form disorder in the entity or immunocompetent members of the entity’s species.

> infectious disposition $=_{def}$ Pathogenic disposition borne by a pathogen to be transmitted to a host and then become part of an infection in that host or immunocompetent members of the same species as the host.

The class infectious agent in IDO Core is a subclass of organism, and so cannot include instances of virus. To address this issue, the term infectious structure was developed to parallel the IDO Core term infectious agent and is a logically defined subclass of acellular structure. The term infectious disposition bridges infectious acellular structures and infectious organisms since instances of each bear an infectious disposition. Moreover, the logical definitions of infectious structure and infectious agent are such that, though the former is a defined subclass of acellular structure and the latter a subclass of organism, they are both inferred subclasses of pathogen.

Establishment of localization used in pathogenic disposition is characterized using the GO term establishment of localization in host representing tethering or adhesion to a host, while “formation of disorder” abbreviates two imported IDO Core terms: appearance of disorder, which is a process that results in formation of a disorder. The definition of pathogenic disposition is meant to

reflect a temporal ordering between establishment of localization and appearance of disorder. This is reflected explicitly in the logical axioms associated with the class. Similarly, in the definition of infectious disposition there is an intended temporal ordering between transmission to a host – represented by pathogen transmission process imported from the Pathogen Transmission Ontology [59] – and becoming part of an infection – represented by the IDO Core process of establishing an infection. A pathogen bearing an infectious disposition that generates disorder in a host will have been transmitted to the host prior to establishing localization in the host and will have established an infection prior to the appearance of disorder.

The complexity of the definitions of pathogenic disposition and infectious disposition reflect the variety of pathogen examples in contemporary literature. Consider S. aureus, an opportunistic pathogen [60] in humans. We count S. aureus as a pathogen, even when it does not realize disorder in a host, since it is nevertheless disposed to localize in a human host and generate disorder if given the opportunity. This is a disposition of S. aureus in BFO terms because it is an "internally-grounded" property of the entity [61]. That is, it is part of the material basis of S. aureus to generate disorder in human hosts if given the chance. This is analogous to the way salt has a disposition to dissolve, based on its lattice structure, independently of whether it ever realizes this disposition. Opportunistic pathogens are not pathogens because of an opportunity; they are pathogens because they are disposed to localize and cause disorder in a host.

Consider now, C. botulinum, a pathogen which produces toxin spores sometimes ingested by humans. This bacterium is a pathogen for adult humans since the toxins often result in disorder when ingested. Furthermore, C. botulinum may cause infection in human infants if, say honey colonized by C. botulinum is ingested. The sugar content of honey inhibits C. botulinum growth, but in the low-oxygen, low-acid intestines of human infants, spores can localize, grow, and produce toxins resulting in disorder. Thus, C. botulinum counts as a human infant pathogen. Nevertheless, because C. botulinum is not itself disposed to invade or be transmitted to human infants, we do not say the bacterium is infectious [62]. Being part of an infection is not itself sufficient to be counted as infectious. Pathogens bearing an infectious disposition must be disposed to both transmit and become part of an infection. Many opportunistic pathogens are not infectious.

The immunocompetence and species membership clauses in the respective definitions of infectious disposition and pathogenic disposition are included in IDO Core to address instances where mutations in hosts may block realization of disorder or infection. In such cases, an infectious pathogen may nevertheless be transmissible and cause disorder or infection in others. For example, HIV-1 is a pathogen that may localize in a host with CCR-5 mutations [63] that block the virus from attaching to host cells, and so block pathogenesis to AIDS. Similarly, P. falciparum may be transmitted to a host with a sickle-cell trait that blocks manifestation of the disease malaria [64,65]. However, P. falciparum and HIV-1 count as pathogens even if they do not result in the formation of disorders for hosts with a sickle-cell trait or CCR-5 mutation, respectively. Each pathogen may be transmitted to immunocompetent members of the same species as the host, and so counts as bearing instances of infectious disposition and pathogenic disposition. Note, that P. falciparum and HIV-1 do not result in the formation of disorders in hosts with sickle-cell traits or CCR-5 mutations should not suggest there are no clinical abnormalities associated with these traits or mutations. Individuals with, say, CCR-5 mutations may exhibit clinical abnormalities [66], and so do exhibit disorders. But these disorders are due to the CCR-5 mutation rather than the HIV-1 infection.

3.3 Host

IDO Core and VIDO prioritize neither host nor pathogen in representation of pathogens and associated diseases, adopting the Damage Response Framework (DRF) for guidance in development of relevant terms [67,68]. According to the DRF,

pathogenesis results from interactions between both host and pathogen interacting primarily through host damage, which is a function of the intensity and degree of host response and pathogen factors. Host and pathogen interactions thus influence manifestations of signs, symptoms, and disease. IDO Core defines terms reflecting the DRF:

host role$=_{def}$ Role borne by an acellular structure containing a distinct material entity, or organism whose extended organism contains a distinct material entity, realized in use of that structure or organism as a site of reproduction or replication.

pathogen host role $=_{def}$ Host role borne by an organism having a pathogen as part of its extended organism.

Symptomatic cases of virus infection can be represented by importing terms from the Symptom Ontology, such as dry cough, fever, taste alteration, smell alteration, among others [69]. Given the importance of asymptomatic carriers in viral infection spread, moreover, attention is also given to:

symptomatic carrier role $=_{def}$ Pathogen host role borne by an organism whose extended organism contains a pathogen bearing an infectious disposition towards the host, and the host has manifested symptoms of the infectious disease caused by the pathogen.

asymptomatic carrier role $=_{def}$ Pathogen host role borne by an organism whose extended organism contains a pathogen bearing an infectious disposition towards the host, and the host has no symptoms of the infectious disease caused by the pathogen.

subclinical infection $=_{def}$ Infection that is part of an asymptomatic carrier.

The term subclinical infection reflects standard – if somewhat obscure – use of the terms "subclinical" and "asymptomatic" while allowing for cases of clinically abnormal infections without symptoms. VIDO defines subclinical virus infection as an infection caused by some virus part of an asymptomatic carrier. These remarks bring us full circle to the term disorder introduced above, since clinical abnormality is associated with disorder. When that disorder stems from infection it counts as an:

infectious disorder $=_{def}$ Disorder that is part of an extended organism which has an infectious pathogen part, that exists as a result of a process of formation of disorder initiated by the infectious pathogen.

And when the adverted pathogen is a virus, it falls in the VIDO class:

virus disorder $=_{def}$ Infectious disorder that exists as a result of a process of formation of disorder initiated by a virus.

Medical researchers draw a distinction between symptoms and signs, which OBO Foundry ontologies respect (from OGMS):

symptom $=_{def}$ Process experienced by the patient which can only be experienced by the patient, that is hypothesized to be clinically relevant.

qualitative sign $=_{def}$ Abnormal observable quality of a part of a patient that is hypothesized to be clinically relevant.

processual sign $=_{def}$ Abnormal processual entity occurring in a patient that is hypothesized to be clinically relevant.

An asymptomatic carrier infected with SARS-CoV-2 likely exhibits signs indicating that the infection is clinically abnormal, such as ground-glass opacities or positive PCR test results. Such asymptomatic carriers exhibit an instance of the VIDO class virus disorder which is the material basis of a viral disease:

infectious disease $=_{def}$ Disease whose physical basis is an infectious disorder.

viral disease $=_{def}$ Infectious disease inhering in a virus disorder that is a disorder due to the presence of the virus.

This result aligns with the CDC's case criteria adopted on April 5, 2020 which indicate that the

presence of the SARS-CoV-2 genome or relevant antigens in an individual is sufficient to count as a case of COVID-19, asymptomatic or not [70,71].

3.4 Viral Pathogenesis

Accurately representing viral pathogenesis is of great importance, in particular, as researchers are still working to understand how SARS-CoV-2 infections cause such a wide range of signs and symptoms across demographics. Representing disease pathogenesis is aided by importing terms from relevant OBO Foundry ontologies, to define:

viral pathogenesis $=_{def}$ Pathogenesis process realization of an infectious disposition inhering in a virus or virus population, having at least the proper process parts:
(1) pathogen transmission,
(2) establishment of localization in host,
(3) process of establishing an infection, and
(4) appearance of a virus disorder.

Where pathogenesis is imported from GO:

pathogenesis $=_{def}$ Process that generates the ability of a pathogen to induce disorder in an organism.

As defined, pathogenesis is a success term, in that it encompasses formation of disorder in an entity. This is reflected in (1)-(4) of the viral pathogenesis definition and motivated by the GO Consortium focus on canonical biological processes [72]. This is not to say all virus infections result in successful pathogenesis. An individual may be infected by a virus without a relevant disorder. Absent the disorder, there is no material basis for the associated disease. Consequently, this would not count as an instance of viral pathogenesis.

Viral pathogenesis involves transmission of virions. From PTRANS [59] is imported:

pathogen transmission process $=_{def}$ Process during which a pathogen is transmitted directly or indirectly to a new host.

Role terms – reflecting "externally-grounded" realizable entities acquired based on circumstance, such as the role a student takes on in university – needed to characterize transporters are imported from IDO Core:

pathogen transporter role $=_{def}$ Role borne by a material entity in or on which a pathogen is located, from which the pathogen may be transmitted to a new host.

A subclass – fomite role, roughly, a pathogen transporter role borne by non-living entities – plays an important function in virus transmission.

Viral pathogenesis involves replication in a host. The term virus replication is defined in VIDO as a subclass of the IDO Core term replication, specifically:

virus replication $=_{def}$ Replication process in which a virus containing some portion of genetic material inherited from a parent virus is replicated.

Instances of virus pathogenesis have virus replication parts. Viral replication typically occurs within an:

incubation process $=_{def}$ Process beginning with the establishing of an infection in a host and ending with the onset of symptoms by the host during which pathogens are multiplying in the host.

Which occupies an incubation interval and may precede a communicability interval. The corresponding process during which viral hosts bear a contagiousness disposition has proper part some latency process which itself has an eclipse process part:

communicability interval $=_{def}$ One-dimensional temporal region during which a pathogen host bears a contagiousness disposition.

latency process $=_{def}$ Process beginning with the establishing of an infection in a host and ending when the host becomes contagious, during which pathogens are multiplying in the host.

eclipse process $=_{def}$ Process beginning with the establishment of a virus in a host and ending with the first appearance of a virion following viral

release, during which an infecting virus is uncoating to begin genome replication.

The last being specific to viruses, so specific to VIDO. The remaining terms are found in IDO Core. Viral dormancy is a virus-specific term from VIDO occurring over a:

viral dormancy interval $=_{def}$ One-dimensional temporal region on which a virus is no longer replicating but remains within a host cell and which may be reactivated to begin replication again.

With participants found from familiar viruses such as varicella zoster and herpes simplex.

Generative stage is imported from IDO Core, defined as a temporal subdivision of a developmental process, as well as:

virus generative stage $=_{def}$ Infectious structure generative stage that is a temporal subdivision of a virus developmental process.

Subclasses of which include the stages through which viruses may proceed during replication:

virus attachment stage $=_{def}$ Virus generative stage during which a virion protein binds to molecules on the host surface or host cell surface projection.

virus penetration stage $=_{def}$ Virus generative stage during which a virion or viral nucleic acid breaches the barriers of a host.

3.5 Bridge to CIDO

VIDO serves as a bridge between IDO Core and the IDO extension ontologies representing specific viral diseases. An extension of importance during the pandemic is the recently developed Coronavirus Infectious Disease Ontology (CIDO). CIDO provides semantic resources needed for representing coronavirus genome, surveillance, vaccine, and host data. The ontology has been used to annotate data pertaining to 136 known anti-coronavirus drugs [73], to identify 110 candidate drugs for potential COVID-19 drug repurposing [74,75], and to provide input to machine learning efforts [4] in identifying potential COVID-19 vaccines. Several members of both the IDO Core and the VIDO development teams are members of the CIDO development team working to ensure alignment among these ontologies. Terms reflecting common features of coronaviruses can be imported from OBO ontologies to characterize features of coronaviruses, such as the viral genome including a five-prime nucleotide cap, or the common glycoprotein spikes found in the viral envelope [76,77]. Some terms are specializations of terms from the Protein Ontology [54], including membrane protein (SARS-CoV-2) and spike glycoprotein (SARS-CoV-2). Worth noting is that while CIDO is not the only ontology developed to support curation of COVID-19 data [78,79,80,81], most alternatives are stand-alone initiatives, and so subject to the silo problems typically found in ontologies developed outside the scope of the OBO Foundry and with no attention to its principles.

3.6 Annotations

The extent of coverage in VIDO and related ontologies can be illustrated by annotation of coronavirus research articles.

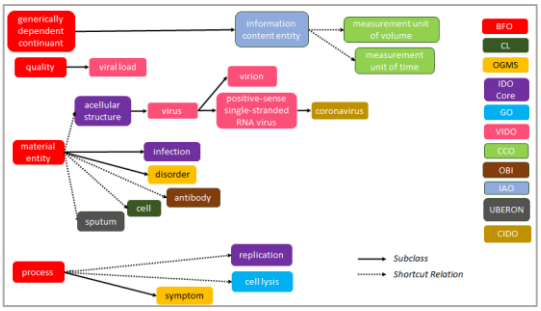

Figure 2: Relationships in the VIDO ecosystem

Consider the following overview of SARS-CoV-2 pathogenesis (color coded as in Figure 5):

Following **replication**, **cell lysis** of **SARS-CoV-2** **coronavirus** **virions** causes **host cells** to **release** **molecules** which **function** to warn nearby **cells**. When **recognized** by **epithelial cells**, **endothelial**, and **alveolar macrophages**, **proteins** such as **IL-6**, **IP-10**, and **MCPI**, are **released** which **attract** **T cells**, **macrophages**, and **monocytes** **to** the **site of infection**, **promoting** **inflammation**. In **disordered immune systems**, **immune cells** **accumulate**

> **in** the **lungs**, then **propagate to** and **damage** other **organs**. In **normal immune systems**, **inflammation attracts T cells** which **neutralize** the **virus at** the **site of infection**. **Antibodies circulate**, preventing **SARS-CoV-2 infection**, and **alveolar macrophages recognize SARS-CoV-2** and **eliminate virions** via **phagocytosis** [82].

In more a more ontologically oriented language, we speak of the relevant part of a host's immune response as being disposed to manifest a response that eliminates SARS-CoV-2 infection, while SARS-CoV-2 has a disposition to block manifestation of this immune system response. Consider next a color-coded selection from the Lancet [83] concerning SARS-CoV-2:

> "The **viral load**s **in throat swab**s and **sputum sample**s peaked **at** around **5-6 days after symptom onset**, ranging from around **10^4** to **10^7 copies per mL during** this **time**."

SARS-CoV-2 infected hosts contain the highest concentration of SARS-CoV-2 virions – the viral load – during the incubation interval [84]. Viral load is a common measurement of the proportion of virions to fluid, and for SARS-CoV-2 is frequently measured from host sputum. VIDO also provides resources for annotating virus quantification:

> viral load $=_{def}$ Quality inhering in a portion of fluid that is the proportion of virions to volume of that portion of fluid.

Our color-coding of this passage models term reuse across existing ontologies. For example, developers can use VIDO terms alongside terms from the Common Core Ontology [85] such as is measured by, measurement information content entity, has integer value, uses measurement unit, and milliliter measurement unit. Other virus quantification metric terms, such as multiplicity of viral infection - the ratio of virions to susceptible cells in a target area – can be found in VIDO as well.

4. Discussion

VIDO enables extensive representation of virus-related research. The very scope of VIDO provides challenges, however. Often, terms were developed then presented to domain specialists for vetting, after which they were refined through discussion. As in the case of all scientific ontologies, refinement will continue as research advances, and further collaborators are welcome and necessary.

The existence of IDO Core extensions covering infectious disease-causing entities other than viruses suggests a need for the creation of reference ontology extensions of IDO covering bacteria, fungi, and parasites. The methodology illustrated in the development of VIDO provides a recipe for such reference ontology creation. Additionally, the methodology illustrated in the development of coronavirus-specific terms outlined above provides a recipe for the creation of novel virus-specific ontologies, namely, by extending them from existing virus ontologies such as CIDO where possible. Adoption of these methodologies by developers during ontology construction will significantly reduce the labor involved in ontology creation.

VIDO represents a substantial effort to characterize viruses in a collaborative, computationally tractable manner. Ontologies like these are crucial in the contemporary Big Data era and will provide researchers needed resources for gathering and coordinating increasingly important life science data.

5. Acknowledgments

JB, SB supported by NIH/NLM T15LM012495-01 Buffalo Research Innovation in Genomic and Healthcare Technology (BRIGHT) Education Training Programs, 2020-2021. BS's contributions were supported by the NIH under NCATS 1UL1TR001412 (Buffalo Clinical and Translational Research Center). VIDO is available under the Creative Commons Attribution 4.0 license and an up-to-date version of the ontology.

Many thanks to Asiyah Yu Lin for assistance in VIDO development; to Darren Natale for helpful critical feedback on an earlier draft. Figures were designed by or in consultation with Rain Yuan.

6. References

[1] Arp R, Smith B, Spear A. Building Ontologies with Basic Formal Ontology. MIT Press, Cambridge, MA, 2015.

[2] He, Y. et al. CIDO, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis. Scientific data. 7 (2020), doi:https://doi.org/10.1038/s41597-020-0523-6

[3] Taylor CF, et al. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. Nat Biotechnol. 2008; 26(8):889–96. doi: 10.1038/nbt.1411

[4] Ong, O., Wong, M., et. al. (2020). COVID-19 Coronavirus Vaccine Design using Reverse Vaccinology and Machine Learning. bioRxiv. doi: https://doi.org/10.1101/2020.03.20.000141.

[5] Liu, Y. Wang, W. et. al. (2020). Ontological and Bioinformatic Analysis of Anti-Coronavirus Drugs and their Implications for Drug Repurposing against COVID-19. OSF Preprint. doi: 10.20944/preprints202003.0413.v1.

[6] Beverley, J., et. el. The Virus Infectious Disease Ontology. 2020. URL: https://bioportal.bioontology.org/ontologies/VIDO.

[7] Cowell LG, Smith B. Infectious Diseases Ontology. In: Sintchenko V, editor. Infectious Disease Informatics. New York, NY: Springer; 2010. p. 373-95.

[8] Musen M, Shah N, Noy N, Dai B, Dorf M, Griffith N, et al. BioPortal: Ontologies and Data Resources with the Click of a Mouse. AMIA Annu Symp Proc. 2008:1223–4. PMID: 18999306

[9] Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007; 25:1251–1255. doi:10.1038/nbt1346.

[10] The Open Biological and Biomedical Ontology Foundry. 2022. http://obofoundry.org/.

[11] Infectious Disease Ontology Extensions. Github. 2022. https://github.com/infectious-disease-ontology-extensions

[12] Shearer R, Motik B, Horrocks I, editors. HermiT: A Highly-Efficient OWL Reasoner. OWLED; 2008.

[13] Evren Sirin BP, Bernardo Grau C, Kalyanpur Aditya, Yarden Katz. Pellet: A practical OWL-DL reasoner Web Semantics: Science, Services and Agents on the World Wide Web. 2007; 5(2):51–3.

[14] McCune, W. Prover9 and Mace4. 2018. https://www.cs.unm.edu/~mccune/prover9/

[15] The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing. Nucleic Acids Res. 2019; 47:D330–D338. doi:10.1093/nar/gky1055.

[16] Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics 25, 25-29 (2000).

[17] Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007; 25:1251–1255. doi:10.1038/nbt1346.

[18] National Cancer Institute Thesaurus. 2022. https://ncithesaurus.nci.nih.gov/ncitbrowser/

[19] National Library of Medicine Taxonomy. 2020. https://www.ncbi.nlm.nih.gov/taxonomy

[20] Smith, B. Classifying Processes: An Essay in Applied Ontology. Ratio.2012. (4):463-488.

[21] Smith, B. On Classifying Material Entities in Basic Formal Ontology. Interdisciplinary Ontology: Proceedings of the Third Interdisciplinary Ontology Meeting. Tokyo: Keio University Press. 2012. 1-13.

[22] Spear, A., Ceusters, W., Smith, B. Functions in Basic Formal Ontology. Applied Ontology. 2016. (11):103-128.

[23] Information Technology – Top-level Ontologies (TLO) – Part 2: Basic Formal Ontology (BFO). 2021. https://www.iso.org/standard/74572.html

[24] Seppala, S., Ruttenberg, A., Schreiber, Y., Smith, B. (2016). Definitions in Ontologies. Cahiers de Lexicologie. 109(2):175-207.

[25] Janssen, T. et. al., Introduction: Ontologies, Semantic Technologies, and Intelligence. Ontologies and Semantic Technologies for Intelligence. 2010.
[26] Babcock S, Beverley J, Cowell LG, Smith B. The Infectious Disease Ontology in the age of COVID-19. J Biomed Semantics. 2021 Jul 18;12(1):13. doi: 10.1186/s13326-021-00245-1. PMID: 34275487; PMCID: PMC8286442.
[27] Babcock, S. Cowell, L. Beverley, J. Smith, B. (2020). Supplementary Documentation to [26].
[28] Lin Y, Xiang Z, He Y. Brucellosis ontology (IDOBRU) as an extension of the infectious disease ontology. J Biomed Semant. 2011; doi: 10.1186/2041-1480-2-9.
[29] Lin Y, Xiang Z, He Y. Ontology-based representation and analysis of host-Brucella interactions. J Biomed Semant. 2015; doi: 10.1186/s13326-015-0036-y.
[30] Luciano J, Schriml L, Squires B, Scheuermann R. The Influenza Infectious Disease Ontology (I-IDO). The 11th Annual Bio-Ontologies Meeting, ISMB. 2008.
[31] Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, et al. The Ontology for Biomedical Investigations. PLOS ONE. 2016; 11(4):e0154556. doi: 10.1371/journal.pone.0154556.
[32] Scheuermann, R., Ceusters, W., Smith, B. (2009). Towards an Ontological Treatment of Disease and Diagnosis. AMIA Joint Summit on Translational Science. 116-20. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041577/
[33] Foulger, R. E., Osumi-Sutherland, D. et. al. (2015). Representing Virus-Host Interactions and other Multi-Organism Processes in the Gene Ontology. BMC Microbiology. 15. 146. doi:10.1186/s12866-015-0481-x
[34] Hulo, C., Masson. P., et. al. (2017). The Ins and Outs of Eukaryotic Viruses: Knowledge Base and Ontology of a Viral Infection. PLoS One. 12:2. doi: 10.1371/journal.pone.0171746
[35] Core Ontology for Biology and Biomedicine. 2018. https://github.com/OBOFoundry/COB/issues/6
[36] Common Anatomy Reference Ontology. 2022. https://bioportal.bioontology.org/ontologies/CARO
[37] Raoult, D. (2014). How the Virophage Compels the Need to Readdress the Classification of Microbes. Virology. Doi: 10.1016/j.virol.2014.11.014. https://www.sciencedirect.com/science/article/pii/S0042682214005157
[38] Roault, D. Forterre, P. (2008). Redefining Viruses: Lessons from Mimivirus. Nature Reviews Microbiology. 6, 315-9. https://www.nature.com/articles/nrmicro1858
[39] Koonin, E. V., Starokadomskyy, P. (2016) Are Viruses Alive? The Replicator Paradigm Sheds Decisive Light on an Old but Misguided Question. Studies in History and Philosophy of Science. 59. 125-34. doi: 10.1016/j.shpsc.2016.02.016
[40] Federhen S. The NCBI Taxonomy Database. Nucleic Acids Res. 2012; 40:D136-D143. doi:10.1093/nar/gkr1178.
[41] Mahmoudabadi, G., Philops, R. (2018). A Comprehensive and Quantitative Exploration of Thousands of Viral Genomes. eLife. doi: 10.7554/eLife.31955
[42] Kuhn, J. (2020). Virus Taxonomy. Reference Modules in Life Sciences. doi: 10.1016/B978-0-12-809633-8.21231-4
[43] Xiang, Z., Courtot, M., Brinkman, R. R., Ruttenberg, A. & He, Y. OntoFox: webbased support for ontology reuse. BMC research notes 3:175, 1-12, (2010) doi:17560500-3-175 doi:10.1186/1756-0500-3-175
[44] Dimmock, N.J., et. al. (2007). Introduction to Modern Virology. 6th edition. Blackwell Publishing.
[45] Cann, A. (2016). Principles of Molecular Virology. Academic Press.
[46] Baltimore, D. (1971). Expression of Animal Virus Genomes. Bacteriological Reviews. 35, 235-41.
[47] Maier, H., Bickerton, E. Britton, P. (2015). Coronavirus: An Overview of their Replication and Pathogenesis. Coronaviruses. 1282: 1-23. doi:10.1007/978-1-4939-2438-7_1.

[48] Bauman, R. (2017). Microbiology with Disease Taxonomy. Pearson Publishing.
[49] Claverie, J. (2006). Viruses take Center Stage in Cellular Evolution. Genome Biology. 7(6). 110. doi:10.1186/gb-2006-7-6-110.
[50] Forterre, P. 2010. Defining Life: The Virus Viewpoint. Origins of Life Evolution Biosphere. 40(2). 151-60. doi:10.1007/s11084-010-9194-1.
[51] Crotty, S. (2001). RNA Virus Error Catastrophe: Direct Molecular Test by Using Ribavarin. Proceedings of the National Academy of Sciences. 98(12). 6895-6900 doi:10.1073/pnas.111085598.
[52] Pfeiffer, J. (2003). A Single Mutation in Poliovirus RNA-Dependent RNA Polymerase Confers Resistance to Mutagenic Nucleotide Analogs via Increased Fidelity. Proceedings of the National Academy of Sciences. 100(12). 7289-7294 doi:10.1073/pnas.1232294100
[53] Hastings, J. et al. 2016. ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic acids research. 44, D1214-19, doi:10.1093/nar/gkv1031
[54] Protein Ontology. 2022. https://bioportal.bioontology.org/ontologies/PR
[55] Barr, J. J., Rita, A., et. al. (2013). Bacteriophage Adhering to Mucus Provide a Non-Host Derived Immunity. PNAS. 110:26, 10771-6. doi:1305923110
[56] Meyer, J. R. (2013). Sticky Bacteriophage Protect Animal Cells. PNAS. 110: 23. 10475-6. doi:1307782110.
[57] Karasov TL, Chae E, Herman JJ, Bergelson J. 2017. Mechanisms to mitigate the trade-off between growth and defense. Plant Cell 29, 666–680. Doi:10.1105/tpc.16.00931
[58] Yassour M, et al. 2016. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. Sci. Transl. Med. 8, 343ra381
[59] Pathogen Transmission Ontology. 2022. https://bioportal.bioontology.org/ontologies/PTRANS
[60] Bauman, R. (2017). Microbiology with Disease Taxonomy. Pearson Publishing.
[61] Goldfain A, Smith B, Cowell LG. Dispositions and the infectious disease ontology. In: Galton A, Mizoguchi R, editors. Formal Ontology in Information Systems: Proceedings of the 6th International Conference (FOIS 2010). Amsterdam: IOS Press; 2010. p. 400-413.
[62] Ananthanarayan, R. & Paniker, J. (2005). Textbook of Microbiology. 7th Edition.
[63] Samson, M. (1996). Resistance to HIV-1 Infection in Caucasian Individuals bearing Mutant Alleles of the CCR-5 Chemokine Receptor Gene. Nature. 382(6593): 722-5. doi:1038/382722a0.
[64] Tiffert, T. et. al. (2005). The hydration state of human red blood cells and their susceptibility invasion by Plasmodium falciparum. Blood, 105 (12) pp. 4853-4860.
[65] Goldfain A, Smith B, Cowell LG. Towards an ontological representation of resistance: the case of MRSA. J Biomed Inform. 2011; 44:35-41. doi:10.1016/j.jbi.2010.02.008.
[66] Vangelista L, Vento S. The Expanding Therapeutic Perspective of CCR5 Blockade. Front Immunol. 2018 Jan 12;8:1981. doi: 10.3389/fimmu.2017.01981.
[67] Pirofski, L. Casadevall, A. (2020). Pathogenesis of COVID-19 from the Perspective of the Damage-Response Framework. MBio. 11(4). 1-12. doi: . https://doi.org/ 10.1128/mBio.01175-20.
[68] Morrison, T., Garsin, D. (2020). Pathogenesis of COVID-19 from the Perspective of the Damage-Response Framework. Host-Microbe Biology. doi:10.1128/mBio.01175-20.
[69] Wang, D. et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. Jama, doi:10.1001/jama.2020.1585 (2020).
[70] Centers for Disease Control and Prevention. (2020). Coronavirus Disease 2019 (COVID-19) 2020 Interim Case Definition, Approved April 2, 2020.
[71] Council of State of Territorial Epidemiologists. (2020). Standardization Surveillance Case Definition and National Notification for 2019 Coronavirus Disease.

[72] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25(1):25–9.
[73] Sayers S, Li L, Ong E, Deng S, Fu G, Lin Y, et al. Victors: a web-based knowledge base of virulence factors in human and animal pathogens. Nucleic Acid Res. 2019; 47:D693-D700. doi:10.1093/nar/gky999.
[74] Zhou, Y. (2020). Network-based Drug Repurposing for Novel Coronavirus 2019-nCoV/SARS-CoV-2. Cell Discovery. 6(14).
[75] Hoffmann, M. et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 181, 271-280 e278, doi:10.1016/j.cell.2020.02.052 (2020).
[76] Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. Annu Rev Virol. 2016;3(1):237-261. doi:10.1146/annurev-virology-110615-042301
[77] Schoeman, D., Fielding, B.C. Coronavirus envelope protein: current knowledge. Virol J 16, 69 (2019).
[78] WHO COVID-19 Rapid Version CRF. 2020. https:// bioportal.bioontology.org/ ontologies/ COVIDCRFRAPID
[79] COVID-19 Surveillance Ontology. 2020. https:// bioportal.bioontology.org/ontologies/COVID19
[80] Linked COVID-19 Data Ontology. 2020. https://github.com/Research-Squirrel-Engineers/COVID-19
[81] COVID-19 Research Knowledge Graph. 2020. https://github.com/nasa-jpl-cord-19/covid19-knowledge-graph
[82] Ye, Q., Wang, B. & Mao, J. The pathogenesis and treatment of the `Cytokine Storm' in COVID-19. J Infect, doi:10.1016/j.jinf.2020.03.037 (2020).
[83] Pan, Y. et. al. (2020). Viral Load of SARS-CoV-2 in Clinical Samples. The Lancet. 20(4):411-2. doi: https://doi.org/10.1016/S1473-3099(20)30113-4
[84] Gandhi, M., Yokoe, D. S. & Havlir, D. V. Asymptomatic Transmission, the Achilles' Heel of Current Strategies to Control Covid-19. The New England journal of medicine, doi:10.1056/NEJMe2009758 (2020).
[85] The Common Core Ontologies. 2022. https://github/com/CommonCoreOntology/CommonCoreOntologies