Coordinating Coronavirus Research: The COVID-19 Infectious Disease Ontology

John Beverley[1,2,4*], Shane Babcock[2,3], Gustavo Carvalho[1], Lindsay Cowell[5], Sebastian Duesing[4], Regina Hurley[1], Barry Smith[2,6]

[1]Department of Philosophy, Northwestern University, Evanston, IL, USA
[2]National Center for Ontological Research
[3]Department of Philosophy, Niagara University, Lewiston, NY, USA
[4]School of the Art Institute of Chicago, Chicago, Il, USA
[5]University of Texas Southwestern Medical School, Dallas TX, USA
[6]Department of Philosophy, University at Buffalo, Buffalo, NY, USA

*Corresponding Author
johnbeverley2021@u.northwestern.edu

Abstract

The COVID-19 pandemic prompted immense investigation of the SARS-CoV-2 virus. Rapidly, accurately, and easily interpreting generated data is of fundamental concern. Ontologies – structured, controlled, vocabularies – support interoperability, and prevent the development of data silos which undermine interoperability. The Open Biological and Biomedical Ontology Foundry serves to ensure ontologies remain interoperable through adherence by its members to core ontology design principles. For example, the Infectious Disease Ontology (IDO) Core includes terminological content common to investigations of all infectious diseases. Ontologies covering more specific infectious diseases, in turn, extend from IDO Core, such as the Coronavirus Infectious Disease Ontology (CIDO). The growing list of virus specific IDO extensions has motivated construction of a reference ontology covering content common to viral infectious disease investigations: the Virus Infectious Disease Ontology (VIDO) (https://bioportal.bioontology.org/ontologies/VIDO). Additionally, the present pandemic has motivated construction of a more specific extension of CIDO, covering terminological content specific to the pandemic: the COVID-19 Infectious Disease Ontology (IDO-COVID-19) (https://bioportal.bioontology.org/ontologies/IDO-COVID-19). We report here the development of VIDO and IDO-COVID-19. More specifically, we examine newly minted terms for each ontology, showcase reuse of terms from existing OBO ontologies, motivate choice-points for ontological decisions based on research from relevant life sciences, apply ontology terms to explicate viral pathogenesis, and discuss the annotating power of virus ontologies for use in machine learning projects.

Introduction

Information emerging from life science research has increasingly been recorded stored in databases.

The sheer volume of data collected by 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 [30], often annotated in specific terminologies and coding systems by

researchers in distinct disciplines. The resulting *data silos* [1] undermine interoperability, meta-data

1

analysis, reproducibility, pattern identification, and discovery across disciplines [12, 13]. The value of cross-discipline meta-data analysis is obvious, and evident in the present pandemic. Prostate cancer researchers [4,5] have leveraged existing research on enzymes crucial in host cell penetration by SARS-CoV-2 to explain differences in disease severity across sex [2,3]. Immunologists have combined insights from research on SARS-CoV-1 and MERS-CoV with chemical compound profile data, to identify drug and vaccine options for SARS-CoV-2 [6,7,8]. Pediatric researchers observing that children have fewer nasal epithelia susceptible to SARS-CoV-2 infection than adults, have suggested this difference partially explains symptom disparities between the groups [9,10]. Researchers [11] across the life sciences clearly recognize the need for coordinated data-driven efforts during the current crisis.

Ontologies - interoperable, logically well-defined, controlled vocabularies representing common entities and relations across disciplines - is a well-known solution to data silo problems. The present need for rapid analysis of evolving datasets representing coronavirus research motivates, moreover, the development of virus, coronavirus, and SARS-CoV-2 specific vocabularies. To these ends, we have developed the Virus Infectious Disease Ontology (VIDO; https://bioportal.bioontology.org/ontologies/VIDO) and the COVID-19 Infectious Disease Ontology (IDO-COVID-19; https://bioportal.bioontology.org/ontologies/IDO-COVID-19). Each is a structured vocabulary, with textual definitions for terms and relations, as well as logical axioms expressed in the OWL 2 Web Ontology Language (https://www.w3.org/TR/owl2-overview/), a World Wide Web Consortium (https://www.w3.org/) language developed for the semantic web. The formal representations of these ontologies support automated consistency checking, querying over relevant datasets, and interoperability with existing data on the semantic web. VIDO is an extension of the widely-used Infectious Disease Ontology Core (IDO Core; https://bioportal.bioontology.org/ontologies/IDO), comprised of terminological content common

to all investigations of infectious disease. VIDO extends to the domain of infectious diseases caused by viruses, and is comprised of terminological content common to investigations of viral disease, classification, epidemiology, replication, vaccinology, and rational drug design. VIDO provides a carefully curated foundation for ontologies representing specific viral infectious diseases such as IDO-COVID-19, an extension of VIDO to the disease COVID-19 and its causative virus SARS-CoV-2.

VIDO and IDO-COVID-19 are available under the Creative Commons Attribution 4.0 license (https://creativecommons.org/licenses/by/4.0/) and up-to-date versions of each are found at the National Center for Biomedical Ontology (NCBO) Bioportal [104], the Ontobee repository (http://www.ontobee.org/), and the Ontology Lookup Service (https://www.ebi.ac.uk/ols/index). Each ontology was developed in collaboration with relevant domain experts, such as immunologists, virologists, ontologists, and logicians, and aligns with principles outlined by the OBO Foundry [15], thereby supporting interoperability with existing Foundry ontologies [14]. Development of the ontologies is transparent, with discussions available on GitHub (https://github.com/) as part of an organization covering IDO Core extension ontologies (https://github.com/infectious-disease-ontology-extensions). VIDO and IDO-COVID-19 term additions are driven by the needs of researchers investigating viruses, COVID-19, and nearby domains. Consequently, neither is claimed to be exhaustive of its respective domain, and each remains sensitive to evolving knowledge.

Methods

OWL, Protégé, Mace4, and Prover9

VIDO and IDO-COVID-19 are underwritten by the OWL 2 Web Ontology Language used to represent ontologies in the semantic web. OWL is an expansion of the Resource Description Framework (RDF; https://www.w3.org/TR/rdf-primer/) and RDF Schema which represent data as sets of subject-predicate-object directed graphs, and which can be queried using the SPARQL

Protocol and RDF Query (SPARQL; https://www.w3.org/TR/sparql11-query/). OWL supplements these languages by allowing for description of classes, members of classes, relations among individuals, and annotation properties. Formally, OWL is a decidable fragment of first-order logic, meaning there is an algorithm which can determine the truth-value for any statement expressed in the language in a finite number of steps [105]. Restricting expressions to a decidable language allows automated consistency and satisfiability checking.

VIDO and IDO-COVID-19 were developed using the Protégé-OWL editor (https://protege.stanford.edu/) and tested against automated reasoners such as HermiT [17] and Pellet [18]. Additionally, logical axioms underwriting these ontologies were translated into a syntax readable by the Mace4 (https://www.cs.unm.edu/~mccune/prover9/) model checker, which allowed for manual graphical inspection of classes of models constrained by the asserted axioms. An automated proof-checker Prover9 (https://www.cs.unm.edu/~mccune/prover9/) bundled with Mace4 was used to validate expected theorems while refining axiom models.

Alignment with OBO Foundry Ontologies

Ontologies are widely used in bioinformatics and biomedical data standardization, supporting data integration, sharing, reproducibility, and automated reasoning. The Gene Ontology (GO; https://bioportal.bioontology.org/ontologies/GO), for example, maintains species-neutral annotations of gene products and functions, and – since its inception in 2000 – has inspired an explosion of biomedical ontologies covering domains of relevant chemicals, diagnostic tools and methods, human diseases, and their causes [20,21,22]. These early developments led to worries, however, that data silos – the very problem ontologies were designed to address – might reemerge [1] as researchers developed ontologies using concepts local to their discipline. By 2007, the Open Biomedical and Biological Ontologies (OBO) Foundry [14] was created to provide guidance for

ontology developers and promote alignment and interoperability while structuring data. The OBO Foundry design principles require ontologies:

- Use a well-specified syntax
- Share a common space of identifiers
- Be openly available in the public domain for reuse
- Be developed in a collaborative effort with ontologies covering nearby domains
- Be developed in a modular fashion
- Have a clearly specified scope
- Use common unambiguously defined relations between their terms
- Conform to a common top-level architecture

The OBO library (http://obofoundry.org/) presently consists of over 200 ontologies, including some externally developed ontologies such as the NCI Thesaurus (https://ncithesaurus.nci.nih.gov/ncitbrowser/) and the NCBI Taxonomy (https://www.ncbi.nlm.nih.gov/taxonomy), and some constructed *ab initio* to satisfy OBO Foundry principles. At its core is Basic Formal Ontology (BFO; https://bioportal.bioontology.org/ontologies/BFO), which is an ISO/IEC approved standard 21838-2 (https://www.iso.org/standard/74572.html; https://standards.iso.org/iso-iec/21838/-2/ed-1/en/). BFO is a top-level ontology covering general classes such as *material entity*, *quality*, *process*, *function*, and *role* [1,23,24,25], and provides the architecture adverted in the last Foundry principle.

Where BFO is domain-neutral, OBO Foundry ontologies are domain-specific, meaning they 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, foster integration across ontologies. VIDO and IDO-COVID-19 were designed with alignment and conformance in mind. In the interest of reusability, development of each ontology followed metadata conventions exhibited by many OBO

Foundry ontologies [16]. These conventions require every term introduced to the ontology have a unique IRI, include textual definitions, definition source, designation of term editor(s), editor(s) and preferred term label. In the interest of coordinating development with existing OBO ontologies, VIDO and IDO-COVID-19 developers imported terms where possible from existing OBO library ontologies, and constructed logical definitions using imported terms. Development was guided, moreover, by best practices for definition construction [1,33]. New primitive terms were introduced when needed, after consultation with domain experts, relevant literature, and careful examination of the OBO library to avoid redundancy. VIDO and IDO-COVID-19 development was coordinated over near daily video conferencing and Slack (https://slack.com/) communication involving globally dispersed developers, punctuated by presentation of results to ontologists and domain specialists, where critical comments were solicited and guided subsequent refinements to the ontologies.

'Hub' and 'Spokes' Approach

VIDO and IDO-COVID-19 follow the 'hub' and 'spoke' methodology [31,64,106] for ontology development. For example, VIDO - a 'spoke' - is extended from the Infectious Disease Ontology Core (IDO Core; https://bioportal.bioontology.org/ontologies/IDO) – a 'hub' – which 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, has been updated [31] to keep pace with scientific and top-level architecture changes [32]. 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; https://bioportal.bioontology.org/ontologies/IDOBRU) the Influenza Infectious Disease Ontology (IDOFLU;

https://bioportal.bioontology.org/ontologies/FLU), and more recently the Coronavirus Infectious Disease Ontology (CIDO; https://bioportal.bioontology.org/ontologies/CIDO), each a member OBO library and semantically interoperable with other library ontologies [27,28,29,30]. Several extensions are illustrated in **Figure 1**.
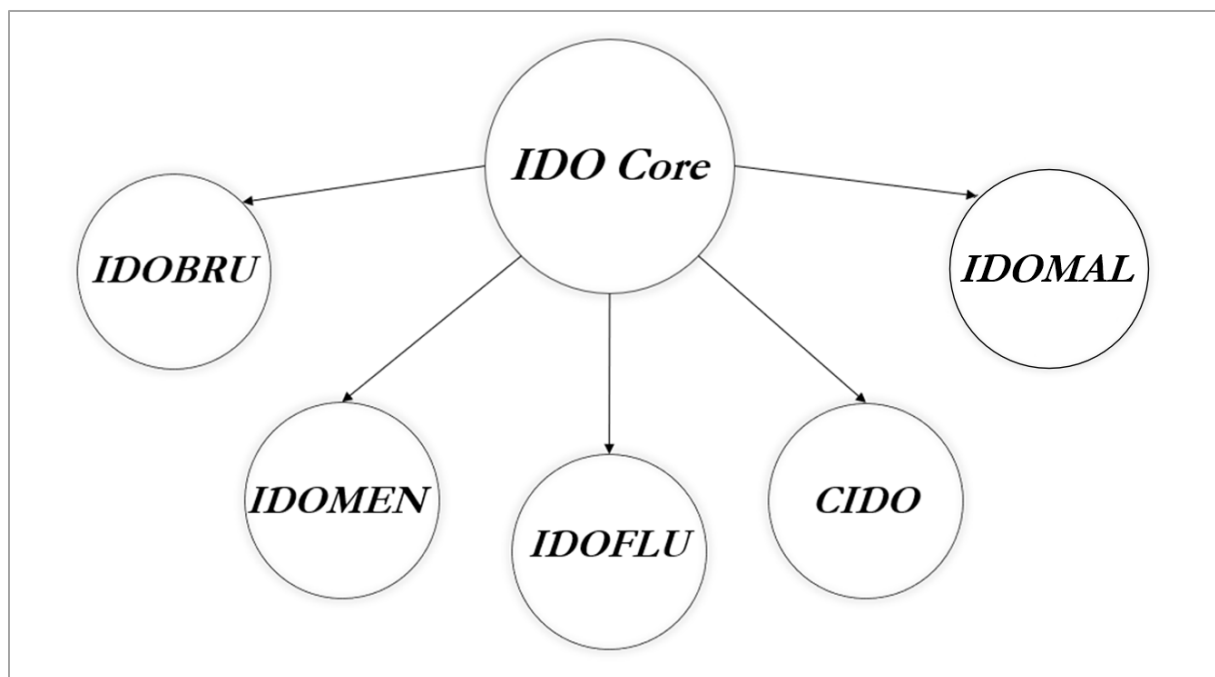

Figure 1: IDO Core and Sample Extensions - the Brucellosis Infectious Disease Ontology (IDOBRU), the Meningitis Infectious Disease Ontology (IDOMEN), the Influenza Infectious Disease Ontology (IDOFLU), the Coronavirus Infectious Disease Ontology (CIDO), the Malaria Infectious Disease Ontology (IDOMAL)

VIDO was designed to occupy ontological space between virus-specific ontologies and IDO Core, the result being, for example, CIDO and IDOFLU extending directly from VIDO. Additionally, the COVID-19 Infectious Disease Ontology is designed to extend from CIDO, the latter covering the domain of coronaviruses generally, and the former covering the domain of SARS-CoV-2 and its associated infectious disease COVID-19.

Results

The Virus Infectious Disease Ontology

Acellular Structure. While VIDO takes IDO Core as its starting point, terms from several other OBO Foundry ontologies were relevant to the domain of viruses, such as the Gene Ontology (GO),

the Ontology of General Medical Science (OGMS;

https://bioportal.bioontology.org/ontologies/OGMS) and the Ontology of Biomedical

Investigation (OBI; https://bioportal.bioontology.org/ontologies/OBI) [34]. **Figure 2** illustrates

high-level relationships of ontologies we discuss in what follows. Arrows represent subclass

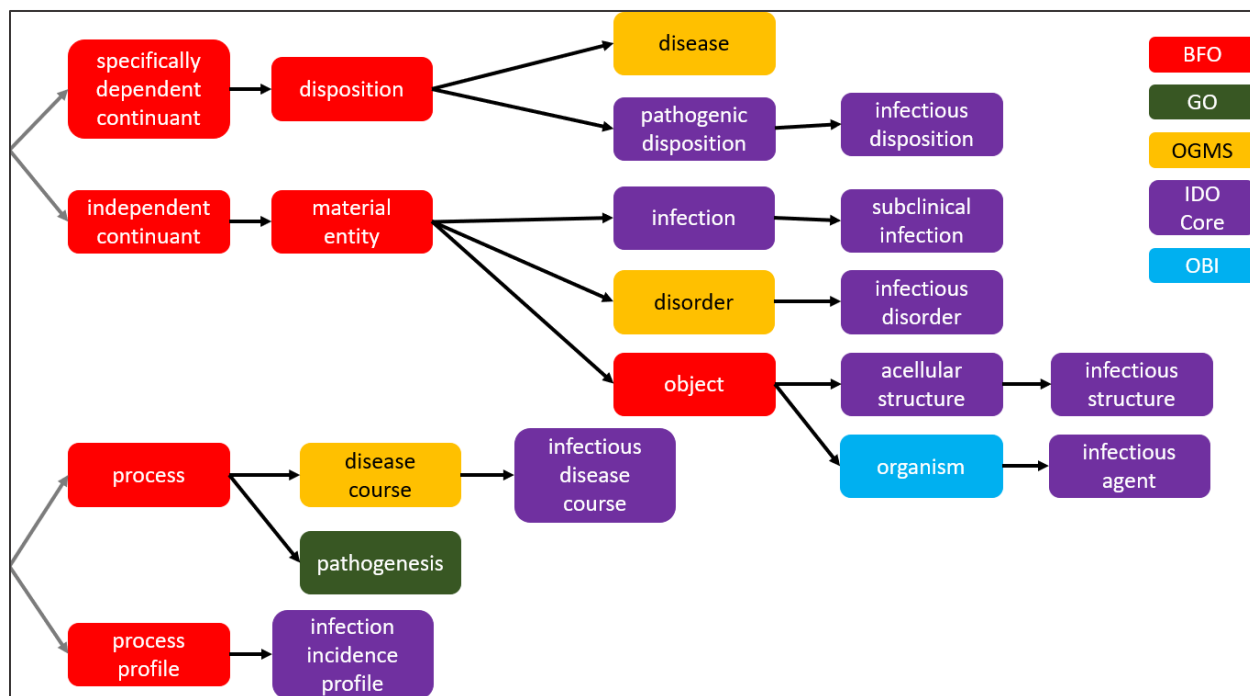relationships. For example, *disease* is a subclass of *disposition*.

Figure 2: Relationships among BFO, GO, OGMS, IDO Core, and OBI

From OGMS, VIDO imports:

   disorder =def Material entity that is a clinically abnormal part of an extended organism.

Where a material entity part 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 – exceeding some threshold – of illness,

death, or disfunction [35]. *Extended organism* is imported from OGMS as well:

   extended organism =def Object aggregate consisting of an organism and all material entities
                located within the organism, overlapping the organism, or occupying
                sites formed in part by the organism.

Which relies on the OBI term:

> *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. The definition of *organism* in OBI includes viruses. This notably stands in contrast to how the term "organism" is used by researchers, who often use it to refer to cellular entities [36,37]. More worrisome, the textual definition asserts instances of *organism* are cellular. Viruses, however, are acellular. Debates (https://github.com/OBOFoundry/COB/issues/6) among ontology developers over *organism* have resulted in deprecation of the OBI term in favor of a nearby term from the Common Anatomy Reference Ontology (https://bioportal.bioontology.org/ontologies/CARO): *organism or virus or viroid.* This avoids the preceding worries but reveals two more. First, introducing disjunctive classes suggests closure over instances, i.e. instances are all and only organisms, viruses, or viroids. We should, however, avoid suggesting classes – especially in biological domains – are settled results of scientific discoveries [1]. Second, this disjunctive class lends itself naturally to intractable, unnecessary debates over whether viruses are alive since it classifies viruses alongside paradigmatic living entities. Decades of discussion has not resolved this admittedly fascinating question [38,39,40,41,42,43], and it is not obvious we need an answer for the purposes of ontological modelling; we should avoid prompting it.

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:

> *acellular structure* =def Object consisting of an arrangement of interrelated acellular parts forming an acellular biological unit.

Which is imported to VIDO, as the parent class of the term *virus*. The term *virus* is, in turn, imported from the NCBITaxon [44] (https://bioportal.bioontology.org/ontologies/NCBITAXON),

alongside other taxon terms representing entities investigated by virologists, e.g. *prion*, *viroid*, and *satellite*.

Virus. While exhaustive of the biological domain, the NCBITaxon provides minimal ontological structure to virus terms, following the International Committee on Taxonomy of Viruses (ICTV). An issue worth noting is that ICTV guidance lacks systematic classification criteria and consequently leaves several viruses unclassified [45,46]. Another is that when NCBITaxon is combined with automated importing tools such as the widely-used Ontofox (http://ontofox.hegroup.org/) [47] the result may be importing of an entire ICTV hierarchy - stretching from kingdom to species - resulting in large, unwieldy, taxonomies obscuring classes of interest. Case in point, the IDO Core extension IDOSCHISTO imports the entire taxonomy from kingdom to respective pathogen using OntoFox, making the ontology rather formidable to navigate [32].

The NCBITaxon itself provides few textual definitions for terms. To align with OBO Foundry metadata conventions [16] and best practices [1,33], textual definitions and logical axioms are needed for *virus* and subclasses. Standard definitions of "virus" provide a starting point for the former, but caution is needed. Viruses are often described – and sometimes defined - as obligate pathogens [48,49], since virus replication requires host machinery for production and assembly of viral components. However, defining a class *virus* in this manner would run the risk of confusing what viruses are – materially speaking - with what viruses typically do. *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. With this in mind, VIDO defines:

> *virus* =def Acellular structure with RNA or DNA genetic material which uses host metabolic resources for RNA or DNA replication.

And rather than import in accordance with the ICTV taxonomy, subclasses of *virus* are imported from NCBITaxon corresponding to a more elegant categorization of viruses: the Baltimore

Classification [50], which groups viruses based on features of genetic structure. There are accordingly seven, exhaustive, classes we import as subclasses of *virus*, of most relevance:

> *positive-sense single-stranded RNA virus* =def Virus with genetic material encoded in single-stranded RNA that can be translated directly into proteins.

Relevant because it includes the subclass:

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

Figure 3 illustrates the Baltimore Classification in Protégé, supplemented by a standard visual summary of the seven viral replication pathways underwritten by virus genetic differences.
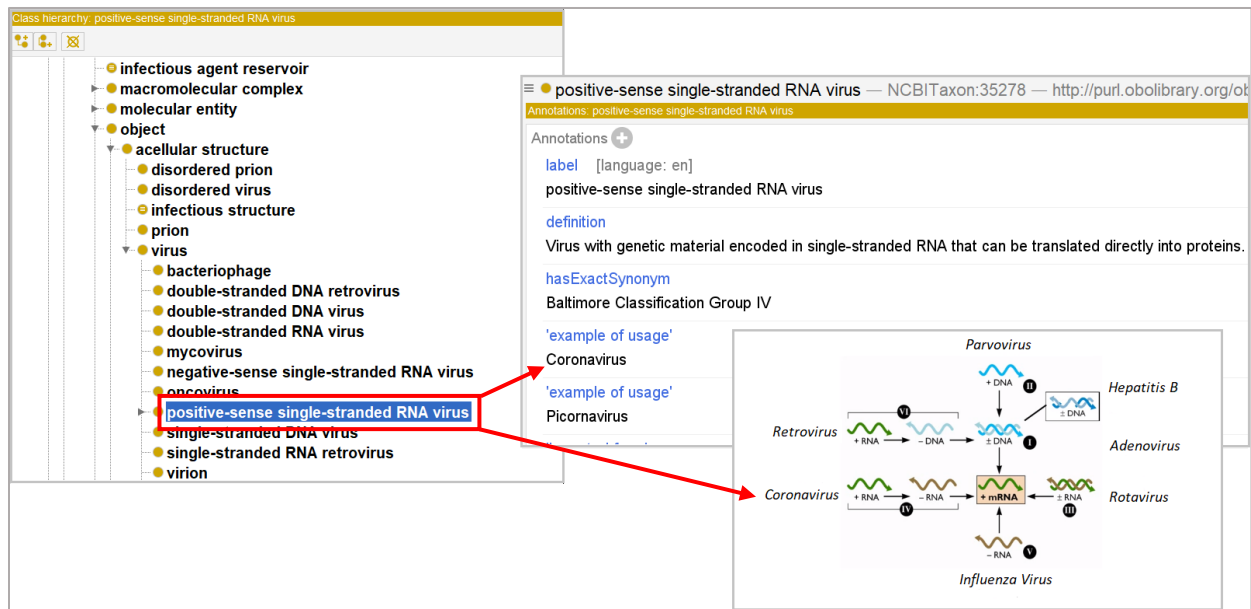

Figure 3: Baltimore Classification in Protégé Editor

More generally, VIDO using the Baltimore Classification provides developers of more specific virus ontologies a succinct, navigable, ontological structure which refers to viral replication pathways, and so the obligate pathogenicity of viruses.

Further 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. A subclass of *virus* imported from

NCBITaxon worth discussing is the VIDO term:

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

Since "virion" is used in divergent ways among researchers. Some use "virion" and "virus"

synonymously [51]. Some define "virion" so that instances only exist outside host cells [52] 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 [53,54]. Ontologically

speaking, one might model the relationship between a virus and its virion in a variety of ways, e.g.

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 gametes is uncommon among researchers, so we put that

option aside. Between the remaining options, we adopt the first, treating *virion* as a type of *virus*, a

stage in virus development following assembly of viral components. Adopting the alternative would

suggest a virion is just a virus in a specific context, with a specific role.

Incidentally, some viruses do not replicate successfully, perhaps resulting in genetically

distinct mutants or – in extreme cases – an inactive aggregate of virion components. Virus mutations

potentially undermine the ability of a host's immune system to recognize that virus as a threat, as

evidenced the difficulty in developing vaccines for certain influenza strains. Too many mutations,

however, and a virus may lose its pathogenic and infectious dispositions, an observation used in

development of treatments for *polio* and *hepatitis C* which exacerbate respective virus mutations

[89,90]. VIDO provides a relevant term for tracking such differences:

> *disordered virus* =def Acellular structure having some arrangement of viral components (e.g.
> viral capsid, viral DNA/RNA), that is clinically abnormal

Viruses falling in this class may even be associated with diseases much different from those of the

clinically normal variety. Related, VIDO imports virus components from: GO such as *nucleocapsid*,

capsid, *capsomere*, *viral envelope*, the Chemical Entities of Biological Interest [55,56] ontology (https://bioportal.bioontology.org/ontologies/CHEBI) such as *nucleic acid*, *ribonucleic acid*, and *epitope*, and the Protein Ontology (https://bioportal.bioontology.org/ontologies/PR) such as *protein* and *viral protein*.

Infectious Structure. IDO Core provides terms needed to represent pathogens, where "pathogen" should be understood as indexed either to a species or to stages in the developmental cycle of a species. With respect to the former, some viruses may engage in mutual symbiosis with one species, while exhibiting pathogenic behavior towards others [57,58]. With respect to the latter, mature plants are often susceptible to different pathogens than developing plants [59,60,61,62]. We capture virus pathogenic behavior in VIDO in steps. We import from IDO:

> *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.

Borne by instances of the class *pathogen*, and:

> *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.

IDO was initially designed around the class *infectious agent*, instances bearing an *infectious disposition*. The class *infectious agent*, however, is a subclass of *organism*, and so cannot include instances of *virus*. To address this issue, the term *infectious structure* was developed in consultation with IDO Core developers 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 the same infectious disposition.

The complexity of the definitions of *pathogenic disposition* and *infectious disposition* reflect the variety of pathogen examples in contemporary literature. Three preliminaries are in order before examining illustrative examples. First, note the term *establishment of localization* used in *pathogenic*

disposition is imported from GO, and is tethering or adhesion to a host, while 'formation of disorder' abbreviates two imported IDO Core terms: *appearance of disorder,* a *process* that *results in the formation of* a *disorder.* Second, there is an implicit temporal ordering in the textual definition of *pathogenic disposition*, which is reflected explicitly in the associated logical definition. Similarly, there is an implicit temporal ordering in the definition of *infectious disposition* between transmission to a host – represented by *pathogen transmission process* imported from the Pathogen Transmission Ontology (https://bioportal.bioontology.org/ontologies/PTRANS) - and becoming part of an infection – represented by the IDO Core *process of establishing an infection*. A pathogen bearing an *infectious disposition* will typically be transmitted to the host prior to establishing localization in the host and will typically establish infection prior to the appearance of disorder. Lastly, the term:

> *infection* =def Material entity part of an organism whose extended organism has some pathogen as part, which participates in the formation of the material entity by infecting the organism.

Is imported from IDO Core as well.

Preliminary comments in hand, the need for such complex definitions is best illustrated by examples. Consider, *s. aureus* is an opportunistic pathogen [63] in humans, becoming harmful to its host under changes in its environment. 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* – following BFO – because it is an 'internally-grounded' property of the entity [64]. 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, whether it ever realizes this disposition. Salt thrown in unsaturated water had a disposition to dissolve before it was immersed; just because an environmental change triggered manifestation does not mean salt lacked the disposition. Similarly,

for opportunistic pathogens, which are not pathogens because of an opportunity, but rather because they are disposed to localize and cause disorder in a host whether they get the opportunity or not.

Consider now, *c. botulinum*, a pathogen which produces toxin spores sometimes ingested by humans. This bacterium counts as a pathogen for adult humans since the toxins often result in disorder when ingested. That said, *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. This of course counts *c. botulinum* as a human infant pathogen. We do not, however, take the further step and say *c. botulinum* bears an *infectious disposition* towards human infants, however, since it is not disposed to invade or be transmitted to them [121]. We leave open whether *c. botulinum* may become part of an infection in an infant. Compatibility with either characterization stems from the fact that being part of an infec*tion* is not itself sufficient to be counted as infec*tious*. Pathogens bearing an *infectious disposition* must be disposed to *both* transmit *and* become part of an infection. Many opportunistic pathogens for example, are not infectious.

Immunocompetence and species membership clauses in the respective definitions of *infectious disposition* and *pathogenic disposition* are included to address instances where mutations in hosts may block realization of disorder or infection. In such cases, a 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 [65] 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 [66,67]. In each case, however, the relevant pathogen may be transmitted to immunocompetent members of the same species as the host. Our definitions count these entities as pathogens, as they should be. It is worth noting, moreover, our claim that *p. falciparum* and HIV-1 count as a pathogens even if they do not result in the formation of

disorders for hosts with a sickle-cell trait or CCR-5 mutation, respectively, does not entail there are *no* clinical abnormalities associated with these traits or mutations. Individuals, for example, with CCR-5 mutations do exhibit clinical abnormalities, and so disorders. Importantly, however, this is not *because* of the HIV-1 virus. Rather, it is because of the genetic mutation. On the other hand, fitness pressure due to the presence of *p. falciparum* results in the presence of sickle-cell trait in a population. Consequently, the clinical abnormality associated with the sickle-cell trait is – in a broad sense – because of *p. falciparum*. Similar remarks would apply to the relationships between CCR-5 mutations and *y. pestis* or *v. major*, if suggested selection pressure explanations involving these pathogens are true [108].

Altogether, *infectious dispositions* are realized in localization in a host, transmission to a host, and generation of infection and disorder in a host or immunocompetent member of the host's species, and *infectious structures* – such as viruses – bear this disposition. SARS-CoV-2, for example, is disposed – as a matter of its material composition – to be transmitted to hosts, localize, result in disorder and infection. 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*, as they should be.

Pathogen Host. We have mentioned "host" at several points in the preceding, and collaboration with the IDO Core team resulted in a ready import for this term to VIDO. Construction of the term was guided by recent shifts in researcher focus on host-pathogen interactions. Until recently, microbiologists [71,72], immunologists, virologists, and others studying pathogenesis have engaged in either host-centered or pathogen-centered pathogenesis research [68,69,70]. Each approach has its merits and has led to impressive research results. Nevertheless, emphasizing one aspect of host-pathogen interactions at the expense of the other may leave valuable questions unanswered. Emphasis, for example, solely on pathogenic factors of SARS-CoV-2 will provide only a partial

explanation of various pathogenesis paths observed in clinical settings; focusing solely on host immune response is likely similarly limiting. VIDO prioritizes neither host nor pathogen in its representation of viruses and associated diseases, instead adopting the Damage Response Framework (DRF) in development of pathogen and host terms [76,77,78,79], which recognizes the importance of both to pathogenesis:

(I)   Pathogenesis results from interactions between host and pathogen, and attributable to neither alone
(II)   Host and pathogens interact primarily through damage to the host
(III)   Host damage is a function of the intensity and degree of host response and pathogen factors, each determined by genetic and phenotypic profiles

Host and pathogen engage in – metaphorically – a tug of war, the result of which influences manifestations of signs, symptoms, and disease.

These reflections bring us to the definitions relevant to hosts developed with IDO Core collaborators, then imported to VIDO:

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 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.

Where symptomatic cases of virus infection can be represented by importing from the OBO Foundry Symptom Ontology (https://bioportal.bioontology.org/ontologies/SYMP) terms for *dry cough*, *fever*, *taste alteration*, *smell alteration*, among others [80]. Given the importance of asymptomatic carriers in viral infection spread, moreover, special attention should be given to:

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 not somewhat obscure – use of the terms "subclinical" and "asymptomatic" while nevertheless allowing for cases in which hosts with clinically abnormal infections exhibit no symptoms. For VIDO, this term is straightforwardly extended to *subclinical virus infection*, which is an infection caused by a virus that is part of an asymptomatic carrier.

Note, an individual exhibiting no symptoms of infection may nevertheless exhibit *signs* of infection. Medical researchers draw a distinction between symptoms and signs, which OBO Foundry ontologies respect with the following imported from OGMS [35]:

> *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.

For example, an asymptomatic carrier infected with SARS-CoV-2 likely exhibits signs indicating the infection is clinically abnormal, such as ground-glass opacities or positive PCR test results. These remarks bring us full circle to the term *disorder* imported to VIDO, 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 organism whose extended organism has some infectious pathogen as part, which participates in the formation of the infection.

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.

Which can be straightforwardly extended to virus disorders involving specific viruses.

Viral Disease. A given *virus disorder* will be the material basis of some associated *viral disease* which may be realized in some associated *viral disease course*. For example, an asymptomatic carrier of SARS-

CoV-2 counts as both a carrier and as having the associated disease. This result aligns, moreover, with the CDC's case criteria adopted on April 5th, which indicates that the presence of the SARS-CoV-2 genome in an individual is sufficient to count as a case of COVID-19, antigen presence is sufficient to count as likely COVID-19, and – generally speaking – asymptomatic cases should be counted as instances of the disease [81,82]. To represent diseases related to viruses, from IDO Core, VIDO imports:

> *infectious disease* =def Disease whose physical basis is an infectious disorder.
>
> *infectious disease course* =def Disease course that is the realization of an infectious disease.

Where *disease* and *disease course* are themselves imported from OGMS. From these starting points, VIDO developers define:

> *viral disease* =def Infectious disease inhering in a virus disorder that is a disorder due to the presence of the virus.
>
> *viral disease course* =def Infectious disease course whose physical basis is a virus disorder that is clinically abnormal in virtue of the presence of the relevant virus.

Once again illustrating a simple recipe for extending an ontology to a more specific domain.

Viral Epidemiology. Changes in viral disease and infection incidence are among the targets of epidemiological investigation. To that end, VIDO imports from IDO Core:

> *infectious disease incidence* =def Quality that inheres in an organism population and is the number of realizations of an infectious disease for which the infectious disease course begins during a specified period.
>
> *infectious disease incidence rate* =def Quality that inheres in an organism population and is the infectious disease incidence proportion per unit time.
>
> *infectious disease incidence proportion* =def Quality that inheres in an organism population and is the proportion of members of the population not experiencing an infectious disease course at the beginning of a specified period of time and in whom the infectious disease begins during the specified period.

Each of which is a quality inhering in an:

>   *organism population* =def Aggregate of organisms of the same species.

These qualities underwrite the VIDO terms:

>   *viral disease epidemic* =def Process of viral disease realizations in which there is a statistically significant increase in the infectious disease incidence of a population.

>   *viral disease pandemic* =def Process in which multiple viral disease epidemics of the same type of viral disease unfold over overlapping periods of time and affect organism populations located in different geographic regions, including different countries and continents.

Note, each is a subclass of a respective 'infectious disease' class imported from IDO Core. This is important, since "epidemic" and "pandemic" refer to diseases spread by infectious entities. Cancers of various sorts are widespread diseases but are not infectious, and so do not count as sustaining a 'cancer pandemic' [109]. AIDS, on the other hand, is considered a pandemic, in part due to the transmissibility of HIV [110]. For simplicity here we focus on infectious disease, but VIDO imports several other important epidemiological terms, such as *infection prevalence*, *infectivity*, and *infectious disease mortality rate*. Each are qualities inhering in some material entity, though not always in some organism population. For example, *infectivity* is a quality that inheres in instances of *pathogen*. Additionally, VIDO imports terms such as:

>   *infection incidence* =def Quality that inheres in an organism population and is the number of organisms in the population that become infected with a pathogen during a specified period.

On which instances of *infectious disease incidence* depend since infectious disease realizations require infection. **Figure 4** illustrates relationships among instances of *infection*, *infectious incidence*, *viral disease epidemic*, and *viral disease pandemic* over time.
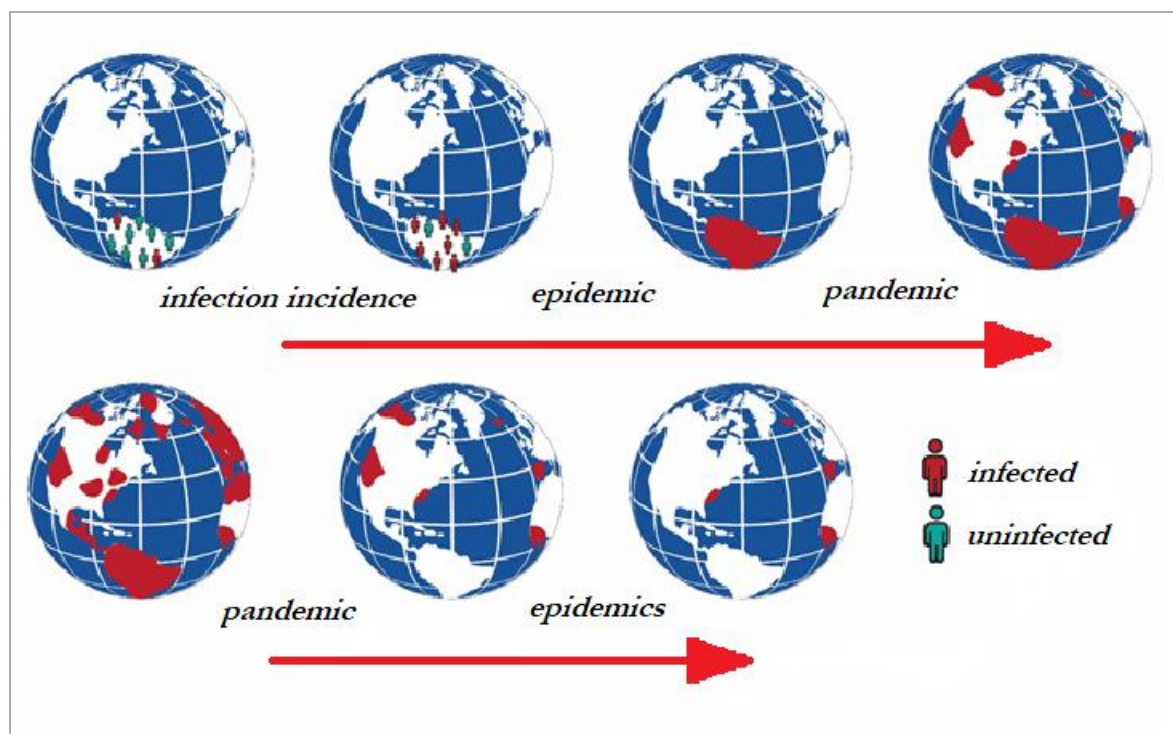
Figure 4: Relationships among infection incidence in a population, epidemics, and pandemic

In words, when an infection incidence in a population increases beyond a certain threshold in a geographic region, this may signal an epidemic in the region. When epidemics emerge in distinct geographic regions, this may signal the emergence of a pandemic. Over time, a pandemic may involve more or fewer geographic regions, and remain a pandemic. However, once the number of epidemics decreases below a certain threshold, there is no longer a pandemic. Similarly, the distribution of infections among members of a population in a geographic region may change while sustaining an epidemic, but once the infection incidence falls below a certain threshold, there is no longer an epidemic.

Collaboration with IDO Core developers resulted in the introduction of subclasses of BFO's *process profile*, essentially, a class of processes tracking changes in specific qualities in material entities over time [1,23]. For example, a patient's temperature will likely fluctuate over time, as will many other qualities of the patient. The specific fluctuations of temperature in the patient over time is a

process profile, a common abstraction used in clinical diagnosis. Changes in qualities of clinical interest

may follow several patterns, each of which can be defined as a subclass of *process profile*. A patient's

temperature may exhibit a linear increase, followed by a linear decrease. Similarly, there are *process

profile* instances of irregular patterns, or as is most relevant to us here, cyclical patterns. *Influenza*, for

example, exhibits complex seasonal patterns [111], which we may characterize in VIDO by defining:

> *viral disease incidence profile* =def Infectious disease incidence profile involving a series of determinate infectious disease incidence qualities caused by a specific virus in a population over time.

> *viral disease proportion profile* =def Infectious disease proportion profile that inheres in an organism population and is the infectious disease incidence caused by a specific virus per unit time.

> *viral disease rate profile* =def Infectious disease incidence rate profile that inheres in an organism population and is the infectious disease incidence proportion caused by a specific virus per unit time.

Among other classes paralleling the IDO Core qualities for organism populations and pathogens

introduced above.

VIDO's Relationship to CIDO

VIDO was developed as a bridge between IDO Core and extension ontologies representing specific

diseases and specific causative pathogens. An extension of importance during the pandemic is the

recently developed Coronavirus Infectious Disease Ontology (CIDO;

https://bioportal.bioontology.org/ontologies/CIDO). Developed by Oliver He and his team,

CIDO provides semantic resources needed for representing coronavirus genome, surveillance,

vaccine, and host data. CIDO has been used to annotate 136 known anti-coronavirus drugs [8],

identify 110 candidate drugs [7] for COVID-19 drug repurposing [107], and provides input to

machine learning efforts [6] in identifying potential COVID-19 vaccines. Several members of both

and IDO and VIDO development teams are also members of the CIDO development team

working to ensure alignment among these ontologies, and adherence to OBO Foundry principles.

Coronavirus Disease. CIDO can straightforwardly extend from VIDO by adopting terms such as the following:

> *coronavirus disease* =def Viral disease inhering in a coronavirus disorder.

> *coronavirus disease course* =def Viral disease course that is the realization of some coronavirus disease and has as a participant a coronavirus.

And more generally, CIDO can be populated by starting with a given virus term from VIDO, then creating a subclass of that term restricted to members of the species coronavirus and associated causative diseases. Following representation of the Baltimore Classification in VIDO, for example, a subclass for *positive-sense single-stranded RNA virus* reflecting instances of *coronavirus* which can be imported from the NCBITaxon, and for which a definition was generated above. Features common to coronaviruses, can be used imported from other OBO ontologies to characterize the virus species, such as that the viral genome including a five-prime nucleotide cap, or the common glycoprotein spikes found in the viral envelope [112,113].

Bridge to IDO-COVID-19. Following the 'hub' and 'spoke' approach to ontology development, CIDO in turn acts as a 'hub' for extension to more specific coronaviruses and associated diseases. Whereas CIDO includes terminological content covering existing and novel coronaviruses in general, and so provides resources for high-level comparison of coronavirus biological profiles, more specific extensions delve deeper into the transmission, genome, epidemiology, treatment, and so forth, for given species of coronavirus. **Figure 5** illustrates various links between IDO Core, VIDO, CIDO, and IDO-COVID-19, among others. Several such links have been examined; we examine the rest as we turn to the star of the present pandemic: SARS-CoV-2.
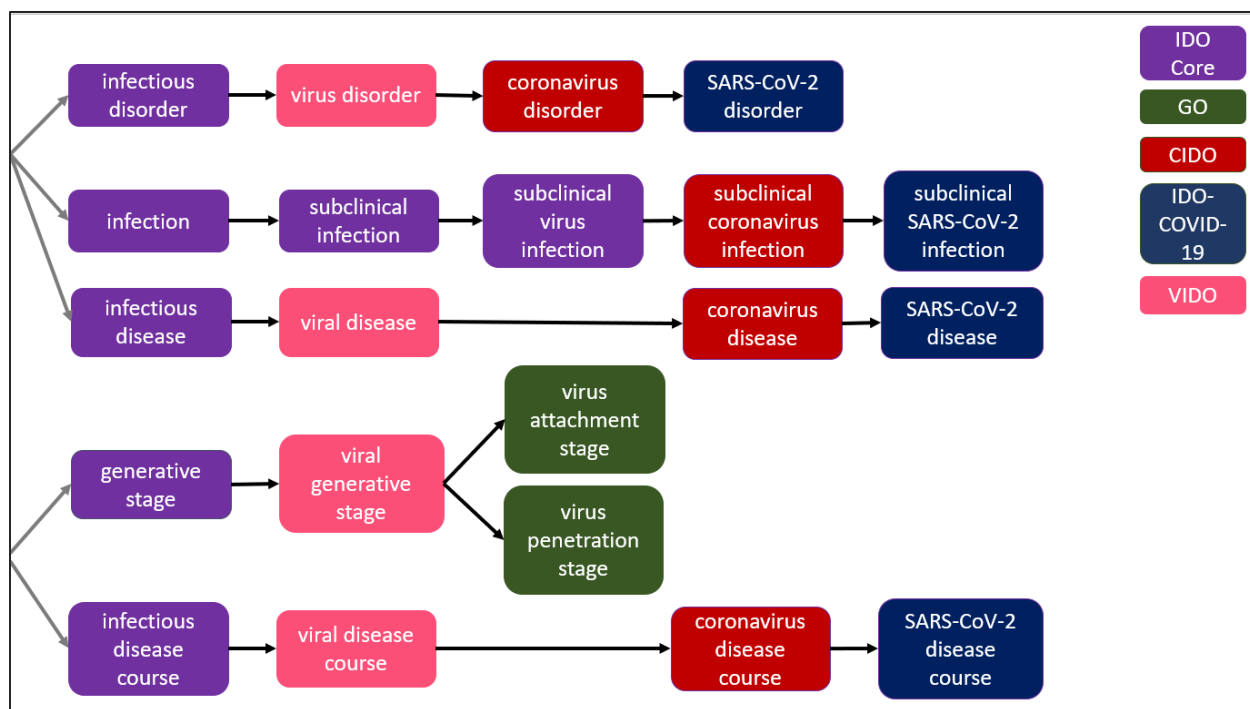
Figure 5: Relationships among GO, IDO Core, VIDO, CIDO, and IDO-COVID-19

The COVID-19 Infectious Disease Ontology

SARS-CoV-2 Pathogenesis. The starting point for IDO-COVID-19 is pathogenesis to COVID-19

caused by SARS-CoV-2. Picking up from CIDO, IDO-COVID-19 introduces the classes:

> *COVID-19* =def Coronavirus disease inhering in a SARS-CoV-2 disorder, and which is realized in some COVID-19 disease course.

> *COVID-19 disease course* =def Coronavirus disease course that is the realization of some COVID-19 disease and has participant SARS-CoV-2.

Representing COVID-19 pathogenesis is of importance during the current pandemic, as researchers

are still working to understand how SARS-CoV-2 infections cause such a wide range of signs and

symptoms across demographics. Evidence suggests SARS-CoV-2 may evade and hijack host

immune response as it spreads [73] resulting in an eventual overreacting immune response that may

kill the host [74]. It is, moreover, worth noting that explaining this phenomenon invariably involves

attention to both host and pathogen contributions to pathogenesis [75], supporting adoption of the

Damage Response Framework as guiding ontological representation of pathogenesis to COVID-19.

24

Representing COVID-19 pathogenesis in IDO-COVID-19 requires importing relevant terms from VIDO, CIDO, and relevant OBO Foundry ontologies, to define terms such as:

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

Instances of *SARS-CoV-2 pathogenesis* are in turn asserted as part of some *COVID-19 disease course*. The term *coronavirus pathogenesis* is imported from CIDO, and is itself a subclass of the VIDO term *viral pathogenesis*, which is in turn a subclass of the term:

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

Imported from the Gene Ontology. 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 *SARS-CoV-2 pathogenesis* definition. This is not to say all SARS-CoV-2 infections result in successful pathogenesis. An individual may be infected by SARS-CoV-2, but this need not result in a relevant disorder. Absent the relevant disorder, there is no appropriate material basis for COVID-19. Consequently, this would not count as an instance of *SARS-CoV-2 pathogenesis*, as the process part (4) would be missing. Commitment to pathogenesis as a success term is analogous to the GO Consortium focus canonical biological processes [103]. **Figure 6** illustrate how these processes are temporally related, and **Figure 7** illustrates mereologically relationships.
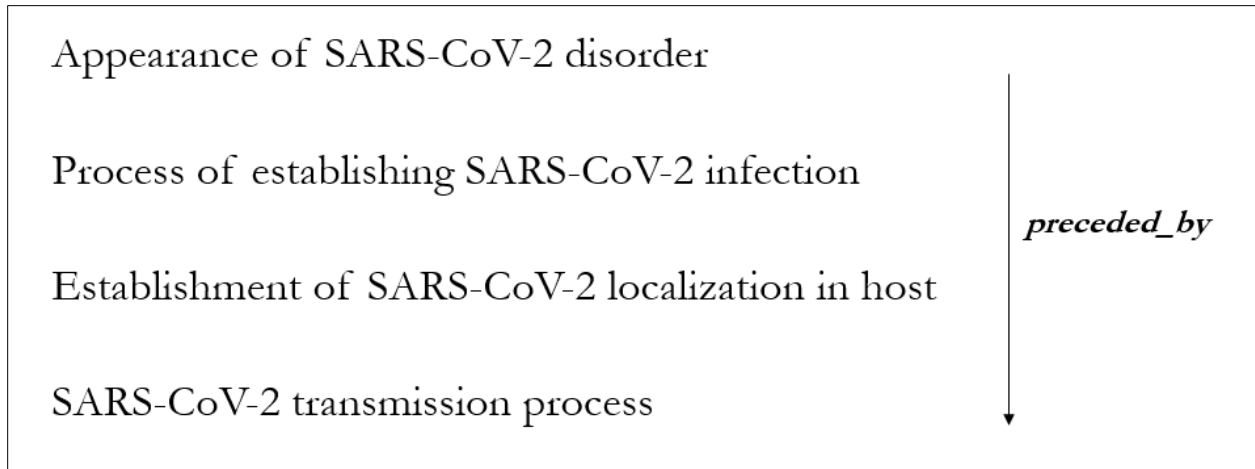
Appearance of SARS-CoV-2 disorder

Process of establishing SARS-CoV-2 infection

preceded_by

Establishment of SARS-CoV-2 localization in host

SARS-CoV-2 transmission process

Figure 6: Sub-processes of SARS-CoV-2 Pathogenesis Temporally Ordered

SARS-CoV-2 pathogenesis

part_of

Appearance of SARS-CoV-2 disorder

part_of

Process of establishing SARS-CoV-2 infection

part_of

Establishment of SARS-CoV-2 localization in host

part_of

SARS-CoV-2 transmission process
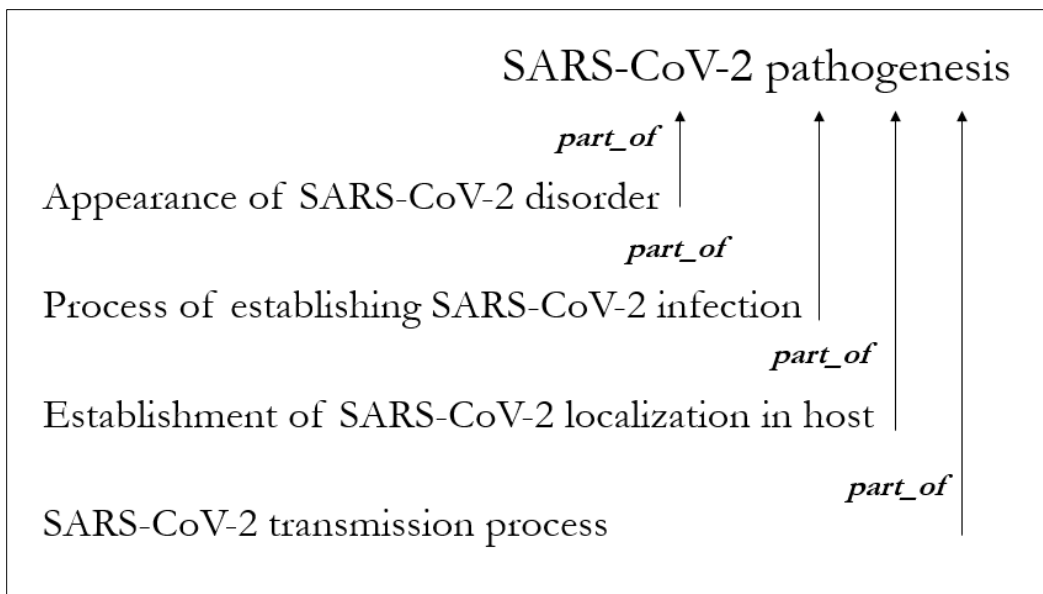
Figure 6: Sub-processes of SARS-CoV-2 Pathogenesis Mereologically Ordered

Just as important as it is to represent SARS-CoV-2 pathogenesis to COVID-19, adequate representation of the target domain requires representation of pathogenesis to *acute respiratory distress syndrome* (ARDS), which has been one of the leading causes of death in those infected by SARS-CoV-2 [119,120]:

> *acute respiratory distress syndrome* =def Progressive and life-threatening pulmonary distress in the absence of an underlying pulmonary condition, usually following major trauma or surgery.

From the OBO Experiment Factor Ontology (https://bioportal.bioontology.org/ontologies/EFO).

SARS-CoV-2 pathogenesis as defined involves transmission of SARS-CoV-2 virions. From PTRANS (https://bioportal.bioontology.org/ontologies/PTRANS) is imported:

> *pathogen transmission process* =def Process during which a pathogen is transmitted directly or indirectly to a new host.

From which SARS-CoV-2 specific terms can be constructed. Role terms – reflecting 'externally-grounded' realizable entities in BFO, or 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, such as:

> *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 respiratory droplet carrying SARS-CoV-2 expelled from an infected individual may be inhaled by an organism, leading to establishment of SARS-COV-2 in parts of the organism permitting the virus to replicate. IDO-COVID-19 imports needed terms to capture transmission from the Environmental Ontology (https://bioportal.bioontology.org/ontologies/ENVO):

> *respiratory droplet* =def Respiratory secretion composed of bounded portion of liquid which maintains its shape due to surface tension

Combining with the term *fomite* from IDO Core, to create:

> *respiratory droplet SARS-CoV-2 fomite* =def Respiratory droplet fomite with SARS-CoV-2 part

Instances of which are common targets of research in transmission studies.

Knowledge of transmission steps supports strategies designed to break the transmission chain. The OBO library already includes an ontology – APOLLO-SV - which contains terms useful in representing various transmission control strategies (https://bioportal.bioontology.org/ontologies/APOLLO-SV). Given the variety of strategies

employed during the pandemic, IDO-COVID-19 imports liberally from this ontology. Two

examples are:

> *contact tracing* =def Infectious disease control strategy that identifies and treats contacted organisms in a host population

> *quarantine control strategy* =def Infectious disease control strategy whereby asymptomatic carriers who have had contact with pathogens are prevented from having contact with other susceptible organisms

Other APOLLO-SV classes are needed to adequately capture these strategies, in particular, classes

populated from each of the Information Artifact Ontology

(https://bioportal.bioontology.org/ontologies/IAO) parents imported by APOLLO-SV, *action*

specification, *objective specification*, and *plan specification*, the first and second parts of the third.

SARS-CoV-2 Replication. SARS-CoV-2 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.

And instances of *viral disease course* and *virus pathogenesis* have as respective parts *virus replication*. CIDO

and IDO-COVID-19 introduce expected terms, the latter extended from the former:

> *SARS-CoV-2 replication* =def Coronavirus replication in which SARS-CoV-2 is replicated.

SARS-CoV-2 replication 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 has an associated *incubation interval* which precedes the:

> *communicability interval* =def One-dimensional temporal region during which a pathogen host bears a contagiousness disposition.

An *incubation process* has proper part some:

> *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.

Which itself has as proper part an:

> *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 specifically a VIDO term. The remaining terms are imported to

IDO-COVID-19 from IDO Core. Incubation is another relevant virus-specific term from VIDO:

> *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.

Exhibited by familiar viruses such as *varicella zoster* and *herpes simplex*.

Careful curation of these terms is important given that evidence of confirmed cases of

SARS-CoV-2 infection after previous apparent virus clearance have emerged [93,94,95]. There is

presently insufficient evidence to determine whether these are novel reinfection by SARS-CoV-2 or

reemergence of a dormant SARS-CoV-2 infection. IDO-COVID-19 provides resources to represent

either hypothesis, the latter with a *viral dormancy process* in which instances of *SARS-CoV-2* participate.

IDO-COVID-19 imports the newly minted *generative stage* 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 various stages through which viruses may proceed during a given

replication. Of importance here, are terms imported from GO:

> *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.

From which the following are defined in IDO-COVID-19:

> *SARS-CoV-2 attachment stage* =def Virus attachment stage during which SARS-CoV-2 bonds with a host cell.

> *SARS-CoV-2 penetration stage* =def Virus penetration stage during which SARS-CoV-2 penetrates a host cell.

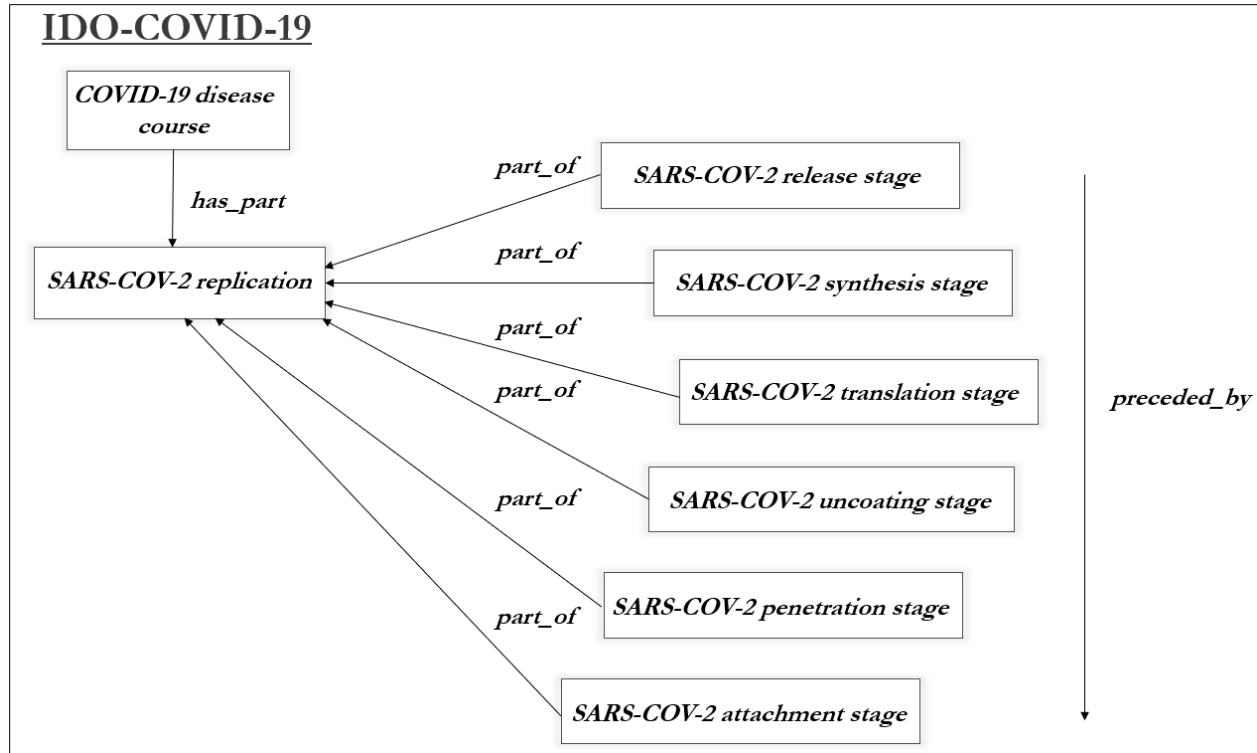**Figure 7** illustrates relationships among stages.

Figure 7: SARS-CoV-2 Replication Generative Stages

SARS-CoV-2 Susceptibility. Not all cells are susceptible to SARS-CoV-2 infection. In those cases of successful infection, the virus attaches to alveolar epithelial cell with a spike surface glycoprotein, by way of these host cell's angiotensin-converting enzyme 2 (ACE2) receptors [85,86,87]. ACE2 receptors appear crucial for SARS-CoV-2 attachment, suggesting the need for:

> *SARS-CoV-2 adhesion susceptible cell* =def Virus adhesion susceptible cell bearing a SARS-CoV-2 adhesion disposition.

Cells lacking ACE2 receptors seem protected from attachment by SARS-CoV-2. As mentioned in the introduction, pediatric researchers have suggested lower levels of ACE2 receptor bearing cells in the nasal epithelium of children may explain why disease severity is much lower for children than for adults [9,10]. The referenced adhesion disposition in the above definition is defined as:

> *SARS-CoV-2 adhesion disposition* =def Virus adhesion disposition borne by a functional receptor that is the disposition to participate in a SARS-CoV-2 attachment process.

Where the adverted functional receptor material base is imported to IDO-COVID-19 from the Protein Ontology (https://bioportal.bioontology.org/ontologies/PRO):

> *ACE2* =def Protein complex consisting of an N-terminal peptidase M2 domain and a C-terminal collectrin renal amino acid transporter domain, which is attached to surface of alveolar, enterocyte cells, arterial and venous endothelial cells, and cortical neurons.

A *SARS-CoV-2 attachment stage* is frequently followed by a penetration stage, involving penetration susceptible cells. More specifically, transmembrane protease serine 2 (TMPRSS2) aids in cleaving host cells in anticipation of SARS-CoV-2 fusing with the cell membrane [88], then introducing viral genomic RNA into the cytoplasm. This similarly suggests a need to define *SARS-CoV-2 penetration susceptible cells* in terms of:

> *SARS-CoV-2 penetration disposition* =def Virus penetration disposition borne by a functional receptor complex that is the disposition to participate in a SARS-CoV-2 penetration process

Where in this case the functional receptor material base is TMPRSS2, also imported to IDO-COVID-19 from the Protein Ontology. Reflection on other stages suggest corresponding terms, since following penetration SARS-CoV-2 genome translation and virion assembly begins in the endoplasmic reticulum, forming virions then packaged into vesicles, sent to the host Golgi apparatus, and fused with the host cell membrane to exit the host.

IDO-COVID-19 terms reflecting stages of the replication cycle for SARS-CoV-2 also provide targets for regulation of that cycle, important to vaccine, drug, and treatment options. Examples of negative regulation relevant here are:

> *negative regulation of SARS-CoV-2 attachment* =def Negative regulation of coronavirus replication process that stops, prevents, or reduces the frequency of some SARS-CoV-2 attachment stage.

> *negative regulation of SARS-CoV-2 penetration* =def Negative regulation of coronavirus replication that stops, prevents, or reduces the frequency of some SARS-CoV-2 penetration stage.

Vaccine trials illustrate the importance of replication [96]. Moderna Therapeutics announced on July 27th the third phase of its clinical trials for the vaccine mRNA-1273, which will test the vaccine in 30,000 U.S. participants. [97,98]. The mRNA vaccine works by introducing a small segment of synthesized SARS-CoV-2 RNA which triggers the immune system to generate viral proteins, which facilitate recognition and elimination of the virus before it spreads throughout the host. On July 27th, Pfizer launched a combined second and third phase clinical trial examining another potential vaccine BNT162b2 exploring diverse populations in areas with high SARS-CoV-2 transmission rates from 39 U.S. states, Brazil, Argentina, and Germany [99]. BNT162b2 differs from mRNA-1273 in that the vaccine prompts host cells to produce the entire spike protein – rather than only part of it – which researchers believe will provide protection in more diverse populations. The University of Oxford is entering a third phase of clinical trials with a viral vector vaccine ChAdOx1 nCoV-19, which transfers the SARS-CoV-2 spike protein the virus uses to invade cells to an attenuated adenovirus, which often causes the common cold, in hopes of triggering immune response to the presence of the spike protein [100]. In each case, vaccines regulating the spread of SARS-CoV-2 in hosts can be represented in part by negative regulation classes.

Annotations. Cells infected with SARS-CoV-2 eventually trigger an immune response. The details are complex, as one should expect, but can be represented in IDO-COVID-19 by importing where possible and defining where needed. We here illustrate the extent of coverage. Consider the following overview of SARS-CoV-2 pathogenesis, where words in **bold** reflect terms included in IDO-COVID-19 and related ontologies:

> **Cell lysis** of **SARS-CoV-2** causes **host cells** to undergo **pyroptosis**, releasing **ATP**, **nucleic acids**, **ASC oligomers**, and other **molecules** whose **function** is to warn nearby **cells**. When recognized by **epithelial cells**, **endothelial**, and **alveolar macrophages**, a

cascade of pro-inflammation **cytokines** and **chemokines** are released. These **proteins**, which include **IL-6**, **IP-10**, **MCPI**, among others [91], attract **T cells**, **macrophages**, and **monocytes** to the **site of infection**, promoting **inflammation**. A feedback loop emerges, whereby inflammation is promoted and promotes further inflammation. In **disordered immune systems**, **immune cells** accumulate in the **lungs**, which are damaged by **inflammation**. At such point, a "cytokine storm" [92] propagates to and damages other **organs**. In **normal immune systems**, **inflammation** attracts **T cells** which neutralize **SARS-CoV-2** at the **site of infection**. **Antibodies** circulate, preventing **SARS-CoV-2 infection**, and **alveolar macrophages** recognize **SARS-CoV-2** and eliminate **virions** via **phagocytosis** [114].

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.

Similarly, SARS-CoV-2 transmission can be described in a few ways, and easily represented in IDO-COVID-19. Consider, where again IDO-COVID-19 terms are in **bold**:

> **SARS-CoV-2** is **directly** or **indirectly transmitted** from a **reservoir** through a **portal of exit** that is **part of** that **reservoir** to a **portal of entry** that is **part of** some **host**. **SARS-CoV-2 participates in** an **establishment of localization in host**, then **participates in** a **process of establishing an infection**.

Related, as indicated earlier asymptomatic SARS-CoV-2 infection is believed crucial for the virus spread [83]. "Viral shedding" occurs during the *incubation period*, which begins with the establishment of an infectious virus in a host and ends with the onset of symptoms. For SARS-CoV-2 infection hosts contain the highest concentration of SARS-CoV-2 virions, i.e. the *viral load*, during this time. Viral load is a common measurement of the proportion of virions to fluid (often in milliliters), and for SARS-CoV-2 is frequently measured from host sputum. VIDO and IDO-COVID-19 provide the 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

VIDO, moreover, imports from the Uber-Anatomy Ontology (https://bioportal.bioontology.org/ontologies/UBERON) the term *sputum*, the term *information bearing entity* from the Information Artifact Ontology

(https://bioportal.bioontology.org/ontologies/IAO), and *is measured by*, *measurement information content entity*, *has integer value*, *uses measurement unit*, and *milliliter measurement unit*, from the Common Core Ontology (https://github/com/CommonCoreOntology/CommonCoreOntologies) to represent research such as – quoted from the *Lancet* [84] with terms from IDO-COVID-19 and nearby ontologies in **bold**:

> "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**."

Similarly, VIDO provides terms for other common virus quantification metrics, such as *multiplicity of viral infection*, the ratio of virions to susceptible cells in a target area.

VIDO, CIDO, and IDO-COVID-19 are presently being used to annotate approximately 400 articles in the National Library of Medicine (https://www.nlm.nih.gov/) COVID corpus, which report COVID-19 clinical trial, epidemiological, and pathogenesis data. The resulting 'gold standard' annotated corpus will be used to train algorithms for use in automated annotating tasks. These algorithms will then be used to identify patterns in rapidly evolving datasets concerning COVID-19.

Discussion

VIDO and IDO-COVID-19 enable representation of various points of virus-related research. Each fits within a broader ontological framework, and indeed fits around an existing coronavirus ontology CIDO. Each complies with OBO Foundry principles, and has been reviewed by several OBO Foundry members, as well as relevant domain experts. Terms needed were imported where possible to avoid redundancy, and where needed, introduced after careful development.

The very scope of VIDO provides challenges, however, as does the specificity of IDO-COVID-19. Viruses are – simply put – mysterious and complex. For this reason, since inception attempts have been made to foster community-driven development of VIDO and IDO-COVID-19. The development team for each ontology spanned disciplines in life science. Additionally, to ensure the computational viability of the formal representation of each ontology, specialists in logic were

included. Often, terms were developed then presented to domain specialists for vetting, after which they were – more often than not - refined, through a process of collective, reflective, equilibrium. This aside, there is no doubt more refinement to be done, and we are eager to continue work with ontology developers and domain experts in the community to get the details right.

From the other direction, the specificity of IDO-COVID-19 requires deep knowledge of viral epidemiology, replication, among other areas. Given the urgency of the present pandemic, experts on these topics are understandably busy, and so perhaps disinclined to aid in ontological analysis over relevant terms. Matters are made more difficulty by the novelty of the pandemic since research for guidance is limited. That said, our team leveraged the expertise of those able to provide it, and each developer has developed competence in the growing literature on COVID-19.

As a final note, we recognize IDO-COVID-19 is not the only ontology initiative developed to support curation of COVID-19 data. Others include:

- The WHO COVID-19 Rapid Version CRF, which provides a semantic data model for the RAPID version (23 March 2020) of the WHO's COVID-19 case record form [115]
- The COVID-19 Surveillance Ontology supports COVID-19 surveillance in primary care by facilitating the monitoring of COVID-19 cases and related respiratory conditions using data from multiple brands of computerized medical record systems [116]
- The Linked COVID-19 Data Ontology uses RDF to present COVID-19 datasets from the European Centre for Disease Prevention and Control, John Hopkins University and the Robert Koch-Institut [117].
- The NASA Jet Propulsion Laboratory's COVID-19 Research Knowledge Graph builds a knowledge graph from the COVID-19 Open Research Dataset [118]

Each, however, is a stand-alone initiative, and so each is subject to the silo problems typically found in ontologies developed outside the scope of OBO Foundry principles. IDO-COVID-19 is not susceptible to these issues, and to that extent is superior to these alternative initiatives.

The successful development of these ontologies reveals more work to be done. VIDO is a reference ontology meant to bridge IDO and virus-specific extensions. Extensions of IDO cover other infectious disease-causing entities. This suggests a need for reference ontology extensions of

IDO covering bacteria, fungi, and parasites. The methodology illustrated here in the development of VIDO provides a recipe for such development. IDO developers working with domain specialists to straightforwardly extend terms where possible, import where needed from the OBO Foundry, and define new terms in consultation with experts should be standard procedure. Similarly, development of IDO-COVID-19 as extending from CIDO should guide needed alignments of existing pathogen-specific ontologies extending from IDO. These efforts are not, of course, as easy as following a simple recipe. But the methodology presented here, and alignment with OBO promoted here, will relieve some of the labor involved in ontology development.

Altogether: VIDO and IDO-COVID-19 represent substantial efforts to characterize viruses in general and SARS-CoV-2 in particular, in a collaborative, computationally tractable manner. Ontologies like these are crucial in the era of 'Big Data' and will provide researchers needed resources for gathering and coordinating increasingly important life science data [101,102].

Acknowledgments

Support

Author Contributions

Wrote the paper: JB, SB, BS. Ontology development: JB, SB, BS, LC, SD, GC, RH.

References

1. Arp R, Smith B, Spear A. *Building Ontologies with Basic Formal Ontology*. Cambridge, MA: MIT Press; 2015.
2. Mollica, V., Rizzo, A., Massari, F. (2020). *The Pivotal Role of TMPRSS2 in Coronavirus Disease 2019 and Prostate Cancer*. Future Oncology. doi: 10.2217/fon-2020-0571.
3. Sharma, G. et. al. (2020). *Sex Differences in Mortality from COVID-19 Pandemic*. JACC Case Reports. 2(9). DOI: 10.1016/j.jaccas.2020.04.027.
4. Jin. J., et. al. (2020). *Gender Differences in Patients with COVID-19: Focus on Severity and Mortality*. Frontiers in Public Health. doi.org/10.3389/fpubh.2020.00152.
5. Stopsack, K., Mucci, L., et. al. (2020). *TMPRSS2 and COVID-19: Serendipity or Opportunity for Intervention?* Cancer Discovery. Doi: 10.1158/2159-8290.CD-20-0451.

6. 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. PMID: 32511333. PMCID: PMC7239068.
7. 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. 2020030413. doi: 10.20944/preprints202003.0413.v1.
8. 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.
9. Dong, Y., et. al. (2020). *Epidemiological Characteristics of 2143 Pediatric Patients with 2019 Coronavirus Disease in China*. Pediatrics. doi: 10.1542/peds.2020-0702.
10. Bunyavanich, S. (2020). *Nasal Gene Expression of Angiotensin-Converting Enzyme 2 in Children and Adults*. JAMA. 323(23):2427-9. doi:10.1001/jama.2020.8707. https://jamanetwork.com/journals/jama/fullarticle/276652
11. Griffiths, E.J., et. al. (2020). The PHA4GE SARS-CoV-2 Contextual Data Specification for Open Genomic Epidemiology. *Preprints* 2020080220 (doi: 10.20944/preprints202008.0220.v1).
12. Taylor CF, Field D, Sansone SA, Aerts J, Apweiler R, Ashburner M, et al. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. Nat Biotechnol. 2008; 26(8):889–96. Epub 2008/08/09. nbt.1411 [pii] doi: 10.1038/nbt.1411
13. Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, et al. (2016) The Ontology for Biomedical Investigations. PLoS ONE 11 (4): e0154556. doi:10.1371/journal.pone.0154556
14. The Open Biomedical Ontologies Foundry. http://obofoundry.org/. Accessed 27 Apr 2020.
15. 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.
16. Schober D, Smith B, Lewis SE, Kusnierczyk W, Lomax J, Mungall C, et al. Survey-based naming conventions for use in OBO Foundry ontology development. BMC Bioinformatics. 2009; 10:125. Epub 2009/04/29. 1471-2105-1
17. Shearer R, Motik B, Horrocks I, editors. HermiT: A Highly-Efficient OWL Reasoner. OWLED; 2008.
18. 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.
19. Homer, H., Selman, A.L. Computability and Complexity Theory: Texts in Computer Science. 2011. 2nd Edition. Springer.
20. 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.
21. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* 25, 25-29 (2000).
22. 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.
23. Smith, B. Classifying Processes: An Essay in Applied Ontology. Ratio.2012. (4):463-488.
24. 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.
25. Spear, A., Ceusters, W., Smith, B. Functions in Basic Formal Ontology. Applied Ontology. 2016. (11):103-128.
26. Cowell LG, Smith B. Infectious Diseases Ontology. In: Sintchenko V, editor. *Infectious Disease Informatics*. New York, NY: Springer; 2010. p. 373-95.
27. 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.
28. 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.
29. Luciano J, Schriml L, Squires B, Scheuermann R. The Influenza Infectious Disease Ontology (I-IDO). *The 11th Annual Bio-Ontologies Meeting, ISMB*. 2008, 20 July; Toronto, Canada.
30. He, Y. *et al.* CIDO, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis. *Scientific data*. 2020. (7):181.
31. Babcock, S. Cowell, L. Beverley, J. Smith, B. (2020). *The Infectious Disease Ontology in the Age of COVID-19*. OSF Preprint. https://osf.io/2w865/
32. Babcock, S. Cowell, L. Beverley, J. Smith, B. (2020). Supplementary Documentation to [20].
33. Seppala, S., Ruttenberg, A., Schreiber, Y., Smith, B. (2016). *Definitions in Ontologies*. Cahiers de Lexicologie. 109(2):175-207.
34. 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.
35. 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/

36. 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. ncbi.nlm.nih.gov/pmc/articles/PMC4517558/

37. 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. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5313201/

38. 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

39. Roault, D. Forterre, P. (2008). *Redefining Viruses: Lessons from Mimivirus*. Nature Reviews Microbiology. 6, 315-9. https://www.nature.com/articles/nrmicro1858

40. 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, https://www.sciencedirect.com/science/article/pii/S1369848616300103

41. Claverie, J.M. (2008) *Encyclopedia of Virology 3rd Edition*

42. Krupovic, M., et. al. (2016). *A Classification System for Virophages and Satellite Viruses*. Archives of Virology. 161, 233-47. Doi: 10.1007/s00705-015-2622-9; https://link.springer.com/content/pdf/10.1007/s00705-015-2622-9.pdf

43. Rodrigo, A. L., Rodrigues, A. C., et. al. (2019). *Giant Viruses and Their Virophage Parasites*. Reference Module in Life Sciences. Doi: 10.1016/B978-0-12-809633-8.20924-2. https://www.sciencedirect.com/science/article/pii/B9780128096338209242

44. Federhen S. The NCBI Taxonomy Database. *Nucleic Acids Res*. 2012; 40:D136-D143. doi:10.1093/nar/gkr1178.

45. Mahmoudabadi, G., Philops, R. (2018). *A Comprehensive and Quantitative Exploration of Thousands of Viral Genomes*. eLife. doi: 10.7554/eLife.31955. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5908442/

46. Kuhn, J. (2020). *Virus Taxonomy*. Reference Modules in Life Sciences. doi: 10.1016/B978-0-12-809633-8.21231-4. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7157452/

47. Xiang, Z., Courtot, M., Brinkman, R. R., Ruttenberg, A. & He, Y. OntoFox: webbased support for ontology reuse. *BMC research notes* **3:175**, 1-12, doi:17560500-3-175 [pii], 10.1186/1756-0500-3-175 (2010).

48. Dimmock, N.J., et. al. (2007). *Introduction to Modern Virology*. 6th edition. Blackwell Publishing.

49. Cann, A. (2016). *Principles of Molecular Virology*. Academic Press.

50. Baltimore, D. (1971). *Expression of Animal Virus Genomes*. Bacteriological Reviews. 35, 235-41.

51. 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.

52. Bauman, R. (2017). *Microbiology with Disease Taxonomy*. Pearson Publishing.

53. Claverie, J. (2006). *Viruses take Center Stage in Cellular Evolution*. Genome Biology. 7(6). 110. doi: 10.1186/gb-2006-7-6-110.

54. Forterre, P. (2010). *Defining Life: The Virus Viewpoint*. Origins of Life Evolution Biosphere. 40(2). 151-60. doi: 10.1007/s11084-010-9194-1.

55. Degtyarenko K, Matos P, Ennis M, et al. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res*. 2008; 36:D344-D350. doi: 10.1093/nar/ gkm791.

56. Hastings, J. *et al.* ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic acids research* **44**, D1214-1219, doi:10.1093/nar/gkv1031 (2016).

57. 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. https://www.pnas.org/content/110/26/10771

58. Meyer, J. R. (2013). *Sticky Bacteriophage Protect Animal Cells*. PNAS. 110: 23. 10475-6. Doi: 1307782110. https://www.pnas.org/content/110/26/10475

59. Karasov TL, Chae E, Herman JJ, Bergelson J. 2017. Mechanisms to mitigate the trade-off between growth and defense. Plant Cell 29, 666–680. (10.1105/tpc.16.00931)

60. 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

61. Simon AK, Hollander GA, McMichael A. 2015. Evolution of the immune system in humans from infancy to old age. Proc. R. Soc. Lond. B 282, 20143085

62. Critchlow JT, Norris A, Tate AT. The legacy of larval infection on immunological dynamics over metamorphosis. *Philos Trans R Soc Lond B Biol Sci*. 2019;374(1783):20190066. doi:10.1098/rstb.2019.0066

63. Bauman, R. (2017). *Microbiology with Disease Taxonomy*. Pearson Publishing.

64. 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.

65. 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.

66. T. Tiffert, et. al. (2005). The hydration state of human red blood cells and their susceptibility invasion by *Plasmodium falciparum*. Blood, 105 (12) pp. 4853-4860.

67. 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.

68. Osterhaus, A. D., Fouchier, R. A. & Kuiken, T. The aetiology of SARS: Koch's postulates fulfilled. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **359**, 1081-1082, doi:10.1098/rstb.2004.1489 (2004).

69. Falkow, S. Molecular Koch's postulates applied to microbial pathogenicity. *Rev Infect Dis* **10 Suppl 2**, S274-276, doi:10.1093/cid/10.supplement_2.s274 (1988).

70. Rivers, T. M. Viruses and Koch's Postulates. *Journal of bacteriology* **33**, 1-12 (1937).

71. Finlay, B. B. & Falkow, S. Common themes in microbial pathogenicity revisited. *Microbiol. Mol. Biol. Rev.* 61, 136–169 (1997).

72. Hacker, J., Blum–Oehler, G., Mühldorfer, I. & Tschäpe, H. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol. Microbiol.* 23, 1089–1097 (1997)

73. Kikkert, M. Innate Immune Evasion by Human Respiratory RNA Viruses. *J Innate Immun* **12**, 4-20, doi:10.1159/000503030 (2020).

74. Tay, M. Z., Poh, C. M., Renia, L., MacAry, P. A. & Ng, L. F. P. The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol*, doi:10.1038/s41577-020-0311-8 (2020).

75. 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.

76. Methot, P., Alizon, S. (2014). *What is a Pathogen? Towards a Process View of Host-Parasite Interactions*. Virulence. 5(8). 775-85. doi: 10.4161/21505594.2014.960726.

77. Casadevall, A., Pirofski, L. (2003). *The Damage-Response Framework of Microbial Pathogenesis*. Nature Reviews Microbiology. 1, 17-24.

78. 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.

79. Yu, H., Huang, H., et. al. (2020). *Ontology-based Systematic Classification and Analysis of Coronaviruses, Hosts, and Host-Coronavirus Interactions towards Deep Understanding of COVID-19*. arXiv preprint. arXiv:2006.00639.

80. 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).

81. Centers for Disease Control and Prevention. (2020). *Coronavirus Disease 2019 (COVID-19) 2020 Interim Case Definition, Approved April 2, 2020*.

82. Council of State of Territorial Epidemiologists. (2020). *Standardization Surveillance Case Definition and National Notification for 2019 Coronavirus Disease (COVID-19)*.

83. 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).

84. 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.

85. Kuba, K. *et al.* A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nature medicine* **11**, 875-879, doi:10.1038/nm1267 (2005).

86. Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature microbiology*, doi:10.1038/s41564-020-0688-y (2020).

87. Chen, Y., Liu, Q. & Guo, D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J Med Virol* **92**, 418-423, doi:10.1002/jmv.25681 (2020).

88. 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).

89. 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.

90. 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

91. Chousterman, B. G., Swirski, F. K. & Weber, G. F. Cytokine storm and sepsis disease pathogenesis. *Semin Immunopathol* **39**, 517-528, doi:10.1007/s00281017-0639-8 (2017).

92. 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).

93. Chen, D., Wenxiong, X. (2020). *Recurrence of Positive SARS-CoV-2 RNA in COVID-19: A Case Report*. International Journal of Infectious Disease. 93: 297-99. doi: 10.1016/j.ijid.2020.03.003.

94. Ibarrondo, F., Fulcher, J. (2020). *Rapid Decay of Anti-SARS-CoV-2 Antibodies in Persons with Mild COVID-19*. The New England Journal of Medicine. Correspondence. DOI: 10.1056/NEJMc2025179.

95. Long, Q., Tang, X. (2020). *Clinical and Immunological Assessment of Asymptomatic SARS-CoV-2 Infections*. Nature Medicine. DOI https://doi.org/10.1038/s41591-020-0965-6.

96. *Pre-Exposure Prophylaxis with Hydroxychloroquine for High-Risk Healthcare Workers during the COVID-19 Pandemic*. Barcelona Institute for Global Health, Hospital of Barcelona, Laboratorios Rubio.

97. *Moderna Announces Phase 3 COVE Study of mRNA Vaccine against COVID-19 (mRNA-1273) Begins*. Press Release (2020); https://investors.modernatx.com/news-releases/news-release-details/moderna-announces-phase-3-cove-study-mrna-vaccine-against-covid.

98. A Study to Evaluate Efficacy, Safety, and Immunogenicity of mRNA-1273 Vaccine in Adults Aged 18 Years and Older to Prevent COVID-19, ClinicalTrials.gov Identifier: NCT04470427, https://clinicaltrials.gov/ct2/show/NCT04470427.

99. *Pfizer and Biotech Choose Lead mRNA Vaccine Candidate against COVID-19 and Commence Pivotal Phase 2/3 Global Study*. Press Release. (2020). https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-choose-lead-mrna-vaccine-candidate-0

100. Folegatti, P. Ewer, K. (2020). *Safety and Immunogenecity of the ChAdOx1 nCoV-19 Vaccine against SARS-CoV-2: A Preliminary Report of a Phase 1/2 , Single-blind, Randomized Controlled Trial*. The Lancet. Doi; 10.1016/S0140-6736(20)31604-4.

101. Apolinario-Arzube Ó. et al. (2020) CollaborativeHealth: Smart Technologies to Surveil Outbreaks of Infectious Diseases Through Direct and Indirect Citizen Participation. In: Silhavy R. (eds) Applied Informatics and Cybernetics in Intelligent Systems. CSOC 2020. Advances in Intelligent Systems and Computing, vol 1226. Springer, Cham. https://doi.org/10.1007/978-3-030-51974-2_15

102. Bernasconi, A.; Canakoglu, A.; Masseroli, M.; Pinoli, P.; Ceri, S. A Review on Viral Data Sources and Integration Methods for COVID-19 Mitigation. Preprints 2020, 2020080133 (doi: 10.20944/preprints202008.0133.v1).

103. 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. Epub 2000/05/10. PMID: 10802651.

104. 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. Epub 2008/11/13. PMID: 18999306

105. Krotzsch, M. et. al. A Description Logic Primer. 2013. Arxiv. arXiv:1201.4089

106. Janssen, T. et. al., Introduction: Ontologies, Semantic Technologies, and Intelligence. Ontologies and Semantic Technologies for Intelligence. 2010.

107. Zhou, Y. (2020). *Network-based Drug Repurposing for Novel Coronavirus 2019-nCoV/SARS-CoV-2*. Cell Discovery. 6(14).

108. Faure E, Royer-Carenzi M (2008). "Is the European spatial distribution of the HIV-1-resistant CCR5-Δ32 allele formed by a breakdown of the pathocenosis due to the historical Roman expansion?". Infection, Genetics and Evolution. 8 (6): 864–874. doi:10.1016/j.meegid.2008.08.007

109. A. M., Dumar (2009). Swine Flu: What You Need to Know. Wildside Press LLC. p. 7.

110. Cohen, MS; Hellmann, N; Levy, JA; DeCock, K; Lange, J (April 2008). "The spread, treatment, and prevention of HIV-1: evolution of a global pandemic". The Journal of Clinical Investigation. 118 (4): 1244–54. doi:10.1172/JCI34706

111. Tamerius J, Nelson MI, Zhou SZ, Viboud C, Miller MA, Alonso WJ. Global influenza seasonality: reconciling patterns across temperate and tropical regions. *Environ Health Perspect*. 2011;119(4):439-445. doi:10.1289/ehp.1002383

112. 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

113. Schoeman, D., Fielding, B.C. Coronavirus envelope protein: current knowledge. *Virol J* **16,** 69 (2019). https://doi.org/10.1186/s12985-019-1182-0

114. Tay, M.Z., Poh, C.M., Rénia, L. *et al.* The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol* **20,** 363–374 (2020). https://doi.org/10.1038/s41577-020-0311-8

115. *WHO COVID-19 Rapid Version CRF*. https:// bioportal.bioontology.org/ ontologies/ COVIDCRFRAPID. Accessed 27 Apr 2020.

116. *COVID-19 Surveillance Ontology*. https:// bioportal.bioontology.org/ontologies/COVID19. Accessed 27 Apr 2020.

117. *Linked COVID-19 Data Ontology.* https://github.com/Research-Squirrel-Engineers/COVID-19. Accessed 27 Apr 2020.
118. *COVID-19 Research Knowledge Graph.* https://github.com/nasa-jpl-cord-19/covid19-knowledge-graph. Accessed 27 Apr 2020.
119. Guo, YR; Cao, QD; Hong, ZS; Tan, YY; Chen, SD; Jin, HJ; Tan, KS; Wang, DY; Yan, Y (13 March 2020). "The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status". *Military Medical Research.* **7** (1): 11. doi:10.1186/s40779-020-00240-0
120. Solaimanzadeh, I (20 March 2020). "Acetazolamide, Nifedipine and Phosphodiesterase Inhibitors: Rationale for Their Utilization as Adjunctive Countermeasures in the Treatment of Coronavirus Disease 2019 (COVID-19)". *Cureus.* **12** (3): e7343. doi:10.7759/cureus.7343
121. Ananthanarayan, R. & Paniker, J. (2005). *Textbook of Microbiology.* 7th Edition.