The Infectious Disease Ontology in the Age of COVID-19

Shane Babcock[1*], John Beverley[2], Lindsay G. Cowell[3], Barry Smith[4]

[1]Department of Philosophy, National Center for Ontological Research, Niagara University, Lewiston, NY, USA
[2]Department of Philosophy, National Center for Ontological Research, Northwestern University, Evanston, IL, USA
[3]Cowell Lab, University of Texas Southwestern Medical Center, Dallas, TX, USA
[4]Department of Philosophy, National Center for Ontological Research, University at Buffalo, Buffalo, NY, USA

[*] **Correspondence:**
Shane Babcock
sbabcock@niagara.edu

This article is a preprint and has not been certified by peer review.

Abstract

Background

Efforts to respond effectively to public health emergencies, such as we are now experiencing with COVID-19, require data sharing across multiple disciplines, and this is hindered by the fact that relevant information is often collected using discipline-specific terminologies and coding systems and stored in heterogenous databases. Ontologies provide a powerful data sharing and integration tool. In practice, however, this method is often undermined by uncoordinated ontology development. Following the principles of the Open Biomedical Ontologies Foundry, the Infectious Disease Ontology (IDO) represents one step towards overcoming such silo problems.

Results

IDO is a suite of interoperable ontology modules that aims to provide coverage of all aspects of the infectious disease domain, including biomedical research, clinical care, and public health. IDO Core is designed to be a disease and pathogen neutral ontology, covering just those types of entities and relations that are relevant to infectious diseases generally. IDO Core is then extended by a collection of ontology modules focusing on specific diseases and pathogens. In this paper we present applications of IDO Core together with an overview of all IDO extension ontologies and the methodology on the basis of which they are built. We also survey recent developments involving IDO, including: IDO Virus (VIDO); the Coronavirus Infectious Disease Ontology (CIDO); and an extension of CIDO focusing on COVID-19 (IDO-COVID-19). We discuss how these ontologies might assist in information-driven efforts to deal with the ongoing COVID-19 pandemic, to accelerate data discovery in the early stages of future pandemics, and to promote reproducibility of infectious disease research.

Conclusions

As we face the continued threat of novel pathogens in the future, IDO provides a simple recipe for building new pathogen-specific ontologies in a way that allows data about novel diseases to be easily compared, along multiple dimensions, with already curated data from earlier diseases. IDO's tightly coordinated suite of ontologies modules provides a powerful method of data integration and sharing that will allow physicians, researchers, and public health organizations to respond rapidly and efficiently both to the current and future public health crises.

Keywords: coronavirus, covid-19, infectious disease, infectious disease ontology, ontology, data integration, data reproducibility

Background

Efforts by physicians, researchers, and public health organizations to respond to infectious diseases require the use of multiple, constantly changing data sources. Consider, for instance, a research team trying to build an effective, large-scale epidemiological system for modeling a given population's herd immunity to measles. This depends on the integration of data not merely from biology and medicine, but also from statistics, sociology, and geography. The system will need to incorporate society-wide data on measles occurrence rates, transmission mode, birth rates, vaccination rates, family structures, age distribution, and other relevant demographic factors [1], and also patient-specific data on clinical manifestations of disease, diagnoses, and treatments received. Because the relevant information is collected using discipline- and community-specific methodologies and is stored in heterogeneous and geographically distributed databases, the data are typically only locally accessible. The resultant silo-formation hinders both translational and comparative research and preventive and prognostic public health research. These problems can be solved by traditional means with the investment of sufficient time and effort. In circumstances of public health emergency such as we are now experiencing however, and where data must be shared across disciplinary borders from immunochemistry to behavioral population modeling, more powerful methods for data sharing and integration must be applied.

As the experience of biologists and bioinformaticians has shown, ontologies are an effective data sharing tool [2]. But in order for ontologies to be effective in this way, it is important that they are designed in a coordinated fashion – otherwise ontologies themselves will give rise to the creation of a new kind of silo. One of the most successful and widely adopted approaches to coordinated ontology development is that of the Open Biomedical Ontologies (OBO) Foundry [3], a collective of developer groups dedicated to creating, testing, and maintaining a collection of ontologies based on an evolving set of design principles for ontology development [4]. The principles include:

- Ontologies should use a well-specified syntax and share a common space of identifiers.
- Ontologies should be openly available in the public domain for reuse.
- Ontologies in neighboring domains should be developed in a collaborative effort.
- Ontologies should be developed in a modular fashion.
- Ontologies should have a clearly specified scope.
- Ontologies should use common unambiguously defined relations between their terms.
- Ontologies should conform to a common top-level architecture.

The OBO Foundry principles were modelled initially on the practices of the Gene Ontology (GO) [2], which is not only enormously successful in its own right but has also served as the model for a series of life science ontologies following in its wake.

Wherever possible, Foundry ontologies are created utilizing terms and relational expressions taken from existing Foundry ontologies, including the Relation Ontology (RO) [5]. This practice is applied also to the terms and relational expressions used in the definitions in these ontologies, and it thereby helps to ensure cross-linkage between ontologies in neighboring domains and also to prevent redundant efforts. It also puts ontologies in a position where they can support integration of data across a wide range while avoiding the generation of silos. Ontologies aligning with OBO

Foundry principles also require each class have a unique identifier with the bipartite form of *ID-space:Local-ID*, as in GO:0008150. Use of such unique identifiers means that the source of each term – and specifically the version of the ontology from which it is drawn – can be immediately identified by its prefix. Use of unique identifiers also ensures that ontologies retain backward compatibility with legacy annotations as those ontologies evolve.

Ontology construction and extension in accordance with these principles follows a 'hub and spoke' model, where a core or 'hub' ontology provides the basis for extension ontologies or 'spokes' which are created by a simple method of specialization. Following this model, the Infectious Disease Ontology (IDO) represents one step towards overcoming data silos [6], consisting of a central hub, formed by IDO Core, and series of spokes in which IDO Core is extended by terms specialized to specific pathogens.

Results

1 The Infectious Disease Ontology

IDO Core is designed to be disease and pathogen neutral. It covers just those entities that are relevant to infectious diseases generally, covering aspects of infectious disease across biological scale (gene, cell, organ, organism, population), disciplinary perspective (clinical, biological, epidemiological), and the chain of infection (host, pathogen, vector, reservoir) [7]. IDO Core thereby provides a 'hub' for a collection of ontology 'spokes' focusing on specific diseases and pathogens. Taken together, these form the IDO suite of ontologies [6].

Since IDO is built in accordance with the OBO Foundry principles described above, this puts IDO Core and its extensions in a position where they can promote interoperability with other ontologies built on the basis of the Foundry principles. This makes IDO Core and its extensions applicable to the annotation of a variety of databases relevant to infectious disease that already make use of Foundry ontologies in their annotations [6], as described in below.

Relevant data and information within multiple disparate sources are annotated using the same set of IDO terms, which all use the same ID-space, namely "IDO". Specific extension ontologies are demarcated via unique ID-blocks pre-assigned by the IDO Core team. The resultant annotated data thereby become available to computer processing as if it were a single body of linked data in virtue of the semantically controlled properties of these terms. These strategies have the benefits of preventing duplicate terms and efforts, enforcing the use of the same ontology development best practices, and encouraging tighter coordination between the IDO Core team and the teams responsible for each extension ontology [8], all of which are important for integration of data across the domain of infectious diseases.

In the ideal case, all IDO extension ontologies would be developed in the same way, and in conformance to all Foundry principles. Unfortunately, not all of these principles have been followed faithfully in the IDO extension ontologies developed thus far. We believe, however, that the work described below on coronavirus-related extensions will serve as a model for the re-engineering of these extensions in such a way as to yield greater conformance.

1.1 IDO Core Foundations

At the heart of the IDO ontology ecosystem is the term 'disease', which is imported from the Ontology for General Medical Science (OGMS) [9]. The latter covers those types of entities relevant to clinical encounters between doctor and patient. Thus, it includes representations of disease, causes and manifestations of disease, diagnosis, symptom, treatment, patient examination,

history taking, laboratory test, and so forth. While OGMS takes clinical encounters involving humans as its starting point, many of its terms can be applied to non-human organisms. OGMS is itself an extension of the Basic Formal Ontology (BFO), a top-level ontology comprised of highly general classes such as 'object', 'material entity', and 'process', and used by more than 300 ontology projects as their top-level architecture [10]. BFO is the official top-level ontology of the OBO Foundry, and it has recently been approved as international standard ISO/IEC 21838-2 [11, 12].

Developers of OGMS view the traditional practice of classifying diseases according to patterns of similarities in signs and symptoms (or, more generally, of phenotypes) as inadequate. A single disease may manifest a variety of symptoms making it difficult to distinguish from other diseases involving the same anatomical system. For example, Celiac disease shares many symptoms in common with Crohn's disease, and another disease of the gut, that caused by *Clostridioides difficile* infection, which is often asymptomatic. Moreover, the traditional practice fails to address the increasing importance played by genetic and environmental variables in disease taxonomy. Seeking to address these issues, OGMS characterizes diseases in BFO terms as *dispositions* of patients to undergo pathological processes of specific kinds. Distinguishing manifestations of symptoms from dispositions to manifest symptoms provides the flexibility needed to represent, say, asymptomatic patients, for the latter are *disposed* to manifest symptoms. It also allows us to make sense of certain physician treatment recommendations, for example that patients on antibiotics should continue treatment after symptoms have subsided. In addition to distinguishing between disease, signs and symptoms, and pathological processes, OGMS further distinguishes between:

- *Diagnosis*
- A disease's realization in a totality of processes forming what is called a *disease course*
- Underlying *disorders* in which the disease is rooted.

Where a *disorder* is characterized as a *clinically abnormal* feature of an organism, a feature that "(1) is not part of the life plan for an organism of the relevant type (unlike aging or pregnancy), (2) is causally linked to an elevated risk either of pain or other feelings of illness, or of death or dysfunction, and (3) is such that the elevated risk exceeds a certain threshold level" [9].

These distinctions provide both the flexibility needed when characterizing disease-related phenomena and also the clarity and consistency needed for collection of accurate data where, for example, different clinicians diagnose the same disease differently, or where a disease exists without having been yet diagnosed [9]. The same disease may be manifested in a wide variety of different types of disease courses depending on the particular patient and on the treatment regimen. And since disorders can exist before they are realized in overt pathological processes, the OGMS approach allows for the existence of pre-clinical manifestations of disease, and for clinical risk factor combinations of disease and predispositions to disease (as when an instance of AIDS in a given patient is a risk factor for a second disease such as tuberculosis). Conflating these distinctions – which has been a problem in even widely-used medical vocabulary resources such as SNOMED-CT [13-15] – makes it more difficult to coherently count disease instances, which in turn may lead to incoherent reasoning about diseases, inconsistent models of specific diseases, errors in patient records, and failures to accurately measure progress in tackling disease spread.

In addition to IDO, many ontologies make use of the OGMS approach to disease. See **Supplementary Table 1**, **Additional File 1** for details.

1.2 Extending IDO Core from OGMS

IDO Core extends OGMS, and in doing so distinguishes between:

- infectious disease
- sign and/or symptom of infectious disease
- infectious disease diagnosis
- infectious disease course
- infection

The OGMS representation of the relationships between disease, disorder and disease courses is inherited by IDO Core as illustrated in **Figure 1**.

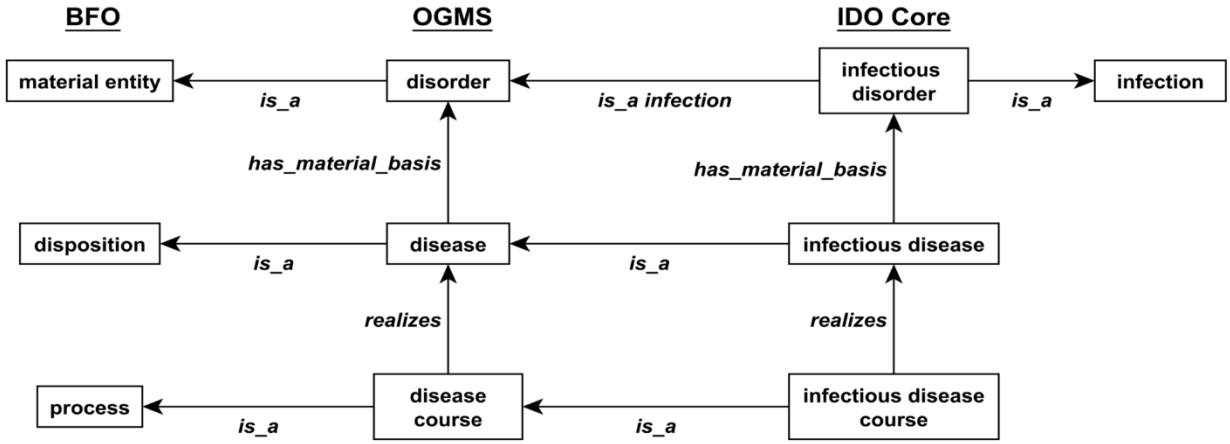

Figure 1 | Relationships between disease, disorder, and disease courses in IDO Core

The relevant OGMS terms are defined as follows:

Disorder =def. Material entity which is clinically abnormal and part of an extended organism; disorders are the physical basis of disease

Disease =def. Disposition to (i) undergo pathological processes that (ii) exists in an organism because of one or more disorders in that organism

Disease course =def. Process that is the totality of all processes through which a given disease instance is realized

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

Has_material_basis is a relational expression from BFO-ISO used to indicate the material basis of a disposition, in this case, a disease. The relevant IDO terms are defined as follows:

Infection =def. Part of an extended organism that itself has an infectious agent population as part, exists as a result of processes initiated by members of the infectious agent population, and is:

1) Clinically abnormal in virtue of the presence of this infectious agent population or,
2) Has a disposition to bring clinical abnormality to immunocompetent organisms of the same Species as the host (the organism corresponding

to the extended organism) through transmission of a member or offspring of a member of the infectious agent population.

Infectious disorder =def. Infection that is clinically abnormal

Infectious disease =def. Disease whose physical basis is an infectious disorder

Infectious disease course =def. Disease course that is a realization of an infectious disease

To elucidate the definition of *infection* and make clearer its relationship to the classes *disorder* and *infectious disorder*, we first introduce further needed terms used by IDO Core. IDO Core imports the class *organism* from the Ontology for Biomedical Investigations (OBI) [16]. It also imports from the NCBI organismal classification (NCBITaxon) [17] classes for a variety of pathogenic organisms such as *bacteria*, as well as host organisms, such as *homo sapiens*, while a adding a new term *virus* (replacing the previously imported NCBITaxon:*Viruses*).

IDO Core defines the following terms pertaining to hosts and pathogens:

pathogenic disposition =def. Disposition to initiate processes that results in a disorder

pathogen =def. Material entity with a pathogenic disposition

pathogen role =def. Role borne by a pathogen in virtue of the fact that it or one of its products is sufficiently close to an organism towards which it has the pathogenic disposition to allow realization of the pathogenic disposition

host role =def. Role borne by an organism in virtue of the fact that its extended organism contains a material entity other than the organism

host =def. Organism bearing a host role

establishment of localization in host =def. Establishment of localization process in which a material entity reaches a site in an organism in which it can survive, grow, multiply, or mature.

IDO Core treats pathogens not as organisms, but as a child class of material entity, since some pathogens are not organisms. Anything that causes or can cause a disorder is a pathogen, including non-living things, such as prions. We chose not to define *pathogenic disposition* in terms of OGMS: *etiological process*, which is defined as a process in an organism that leads to a subsequent disorder [9], because this definition is incorrect. Etiology is about causes of anything, not just disorders. The most common use in medicine is to talk about the etiology of disease, but the term itself is broader.

Building on its pathogen-related terms, IDO Core defines terms pertaining to infectious agents, their dispositions, and their roles (note the distinction between *pathogen role* and *infectious agent role*. *Clostridium botulinum* may bear the former but not the latter; various influenza virus strains, in contrast, may bear both):

infectious disposition =def. Pathogenic disposition that inheres in an organism and is a disposition for that organism to:
(1) be transmitted to a host,
(2) establish itself in the host,
(3) initiate processes that result in a disorder in the host, and
(4) become part of that disorder

infectious agent =def. Organism that has an infectious disposition

infectious agent population =def. Organism population whose members each have an infectious disposition.

infectious agent role =def. Role borne by an infectious agent when contained in a host in which its infectious disposition can be realized.

As well as defining various processes involved in the establishment of infections:

colonization of host =def. Establishment of localization in host process in which an organism establishes itself in a host (Comment: Notice that where this class is restricted to organisms, its parent class includes any case of a pathogen establishing itself in a host, including prions.)

process of establishing an infection =def. Process by which an infectious agent, established in a host, becomes part of an infection in the host

establishment of a clinically abnormal colony =def. Colonization of host process that results in a clinically abnormal colony

Aspects of *infectious disposition* are illustrated in **Figure 2**.

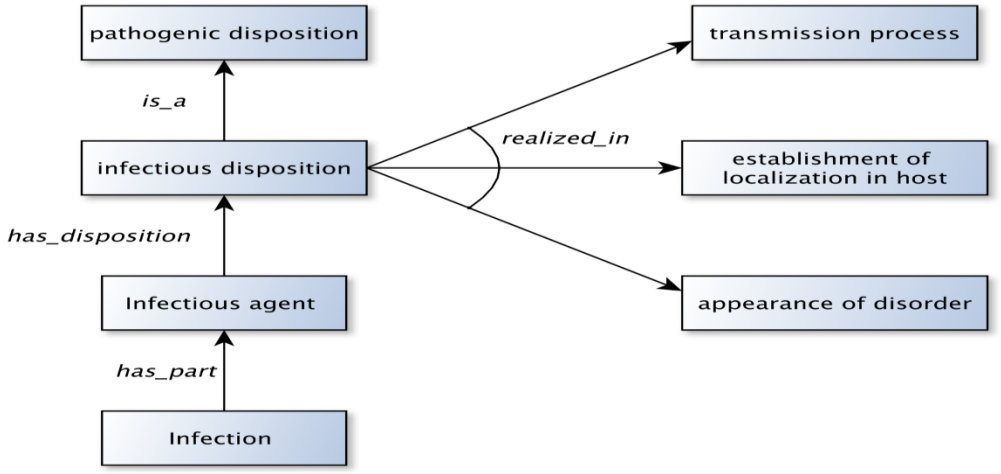

Figure 2 | Some aspects of IDO *infectious disposition*

Clauses (1)-(4) of *infectious disposition* can be thought of as sub-dispositions all of which are borne by the infectious agent. Some organisms may have the disposition to be transmitted to but not to establish itself in the host, in which case they would not have the infectious disposition. According to IDO Core, if an infectious agent realizes its infectious disposition, (1)-(4) are realized sequentially. In some cases, an infectious agent may only be able to realize its infectious disposition *partially*, as when it is transmitted to a host, but fails to establish itself in the host due to the host's immune response. When an infectious agent is transmitted to a host wherein all of its other subdispositions *can* be fully realized, it bears an *infectious agent role*. In this case, an infectious agent is able to colonize the host and then participate in a *process of establishing an infection*. If it participates in the *establishment of a clinically abnormal colony* and initiates processes that result in a disorder in the host, an *infectious disorder* is established. If an *infection* is established, but not a clinically abnormal colony, then the *infection* is not an *infectious disorder*. Within IDO Core, *infectious disorder* has the following logical definition:

infectious disorder ≡ *infection* **AND** *disorder*

Note that the presence of commensal microorganisms within a host, as in the human microbiome, is not clinically abnormal as it is not causally linked to an elevated risk of either pain or other feelings of illness, or of death or dysfunction. By contrast, a dormant infectious agent population of a species that is *not* part of the host's normal resident microbiota – as in the case of a latent infection – is

clinically abnormal and for that reason an *infectious disorder*. Of course, commensals can also initiate disorders in their hosts, *if* they end up in the wrong anatomical site, as in the case of *bacteremia* (defined in IDO Core as "Infection that has as part bacteria located in the blood"), or if the populations grow out of control, as in the case of yeast infections.

Some infections satisfy clause (2) of *infection*, but not clause (1), and so are not disorders. It is for this reason *infection* is not a subclass of *disorder*. An example would be an HIV infected human host that is resistant to the virus due to a mutation of the CCR5 gene. Such an infection is not clinically abnormal as it is not linked to an increased risk of dysfunction (and so on) in the resistant host. Instead, clause (2) is satisfied because the relevant infectious agent population is disposed to bring clinical abnormality to other potential immunocompetent human hosts without the mutation.

The IDO Core definition of *infection* thus allows for representation of commensal populations, infectious disorders that can be caused by organisms that are typically commensal, and the fact that asymptomatic infected individuals are contagious (as we see now with SARS-CoV-2).

1.2 Transmission of Pathogens

IDO Core recognizes the importance of characterizing infectious disease transmission in its various forms. IDO Core imports the following terms from the Pathogen Transmission Ontology [18]:

transmission process =def. Process that is the means during which the pathogen is transmitted directly or indirectly from its natural reservoir, a susceptible host, or source to a new host

indirect pathogen transmission process =def. Transmission process during which a pathogen is indirectly transferred from a reservoir, source or host to another host by intermediary vehicles, vectors, or as airborne dust particles

A variety of infectious diseases, including malaria and dengue fever, are vector borne. Thus, IDO Core contains terms such as:

infectious agent transporter role =def. Role borne by a material entity in virtue of the fact that an infectious agent is located in or on the entity and the entity has the capability to transfer (either actively or passively) the infectious agent from one location to another.

infectious agent vector role =def. Infectious agent transporter role that is borne by an organism active in the transfer of an infectious agent to an organism of another Species and in which the agent is infectious

infectious agent vector =def. Organism bearing an infectious agent vector role

In other cases, infectious agents, such as the *Schistosoma* parasites that cause *schistosomiasis*, spend part of their life cycle within intermediate hosts, after which the pathogen is transmitted into another medium, such as water, which then directly transmits the pathogen to susceptible organisms such as humans. Thus, IDO Core contains the following terms:

symbiont host role =def. Host role borne by an organism in virtue of the fact that its extended organism contains a second organism and provides an environment supportive for the survival, growth, maturation, or reproduction of that organism

intermediate host role =def. Symbiont host role borne by an organism in virtue of the fact that its partner in symbiosis utilizes the host to undergo development

stage transition, and the host is required for continuation of the partner's life cycle

intermediate host =def. Organism bearing an intermediate host role

Moreover, *symbiont host role* includes further subclasses such as:

definitive host role =def. Symbiont host role borne by an organism in virtue of the fact that its partner in symbiosis reaches developmental maturity or reproduces sexually in the host

parasite host role =def. Symbiont host role borne by an organism in virtue of the fact that its partner in symbiosis derives from the host a growth, survival, or fitness advantage while the host's growth, survival, or fitness is reduced

paratenic host role =def. Symbiont host role borne by an organism in virtue of the fact that its partner in symbiosis utilizes the host to undergo a developmental stage transition, but the host is not required for continuation of the partner's life cycle

The preceding selection does not exhaust those host roles included in IDO Core but does reflect the wide range of ways in which to characterize host-symbiont relationships.

1.3 Pathogen Inhibition and Control

IDO Core provides several terms relevant to the inhibition and killing of pathogens:

antibacterial disposition =def. Disposition to kill or inhibit the reproduction of bacteria

antibacterial =def. Material entity bearing an antibacterial disposition

antifungal disposition =def. Disposition to kill or inhibit the development or reproduction of fungal organisms

antifungal =def. Material entity bearing an antifungal disposition

antiparasitic disposition =def. Disposition to kill or inhibit the development or reproduction of eukaryotic parasites

antiparasitic =def. Material entity bearing an antiparasitic disposition

antiviral disposition =def. Disposition to kill or inhibit the lifecycle of viruses

antiviral =def. Material entity bearing an antiviral disposition

Relatedly, one of the most important applications of IDO is its treatment of the phenomenon of *resistance*. Examples of resistance include, a population's herd immunity to certain populations of infectious organisms and the resistance of certain pathogens to antimicrobial drugs. The correct identification of different types of resistance is essential to both treatment decisions and public health policies [19]. For instance, varying strains of a given bacterial pathogen type (e.g. *Staphylococcus aureus*) can differ in terms of their degree of resistance and in the types of drug to which they are resistant. In the examples described above, resistance is a feature of an organism, or population of organisms, that serves to protect it/them from being damaged by some other entity. To capture this aspect of resistance, IDO Core contains the term *protective resistance* which is defined as follows:

protective resistance =def. Disposition that inheres in a material entity in virtue of the fact that the entity has a part (e.g. a gene product), which itself has a disposition to mitigate damage to that entity

Protective resistance includes not just drug resistance on the part of infectious agents or the resistance of hosts to infectious agents, but also things like the resistance of vectors to insecticide.

Notice that what occurs in many cases where protective resistance is manifested is that another process is being prevented. Consider for instance the immunity of an individual X to a specific infectious organism Y that has the capability to cause damage to X. Given its infectious disposition, Y is disposed to be transmitted to and establish itself in X, initiate processes that result in a disorder in Y and become part of that disorder. X's immunity to Y is realized in certain processes that prevent certain of the aforementioned processes from occurring, thus mitigating the damage those process potentially may have caused to X. To capture this aspect of resistance, protective resistance has been characterized in terms of what can be called "blocking dispositions" [7, 19], a disposition the manifestation of which prevents, or at least mitigates, the realization of another disposition. The disposition whose realization is prevented (or mitigated) can be called a "blocked disposition." Thus, since X's immunity to infectious organism Y is realized in processes that prevents certain realizations of Y's infectious disposition, the former is a blocking disposition for the latter (the latter being a blocked disposition).

This characterization of resistance is further enhanced in IDO Core by importing from RO the relation *negatively_regulates*, holding between processes *x* and *y*, and defined as:

x negatively_regulates y =def. The progression of *x* reduces the frequency, rate or extent of *y*

Thus, we can say that X's immunity to Y is a blocking disposition for Y's infectious disposition insofar that X's immunity is realized in certain processes that *negatively_regulate* the manifestation of Y's infectious disposition. For instance, X's immunity may be realized in processes – such as, antibody secretion which would neutralize viral particles, preventing them from entering host cells – the progression of which reduces the rate at which, or the extent to which, *Y* establishes itself in X.

As we have seen, IDO Core aligns with OBO Foundry principles, imports a range of terms from widely used ontologies, and covers a wide range of phenomena in the domain of infectious diseases. In the next sections, we highlight some other ontologies that purport to extend IDO Core.

2 Extending IDO Core

An overview of the results of applying the IDO Core plus extensions method of ontology development is included in **Supplementary Table 2, Additional File 1**, which also documents other disease ontologies employing IDO terms (**Supplementary Table 3)**. The current state of each extension is summarized in **Table 1**. Some further issues concerning the IDO Extensions are detailed in **Additional File 2**.

Table 1 | IDO Extension Ontologies: *=subject to re-engineering, **obsoleted and will be replaced.

Coronavirus Infectious Disease Ontology (CIDO)* [20, 21]	Most recent version uploaded to Bioportal on May 22, 2020 [22]
Influenza Ontology (IDOFLU)* [23]	Most recent version uploaded to BioPortal on August 20, 2015 [24]

Brucellosis Ontology (IDOBRU)* [8, 25]	Most recent version uploaded to BioPortal on March 28, 2015 [26]
IDO Virus (VIDO) [XX]	Currently in development
IDO-COVID-19 Infectious Disease Ontology [XX]	Currently in development
Dengue Ontology (IDODEN)* [27]	Most recent version uploaded to BioPortal on February 17, 2014 [28]
Malaria Ontology (IDOMAL) [29, 30]**	Though obsoleted, IDOMAL is being hosted for legacy purposes [31].
Meningitis Ontology (IDOMEN)* [32]	Draft version uploaded on November 27, 2019 [33]
Plant Disease Ontology (IDOPlant) [34]	Draft version released in 2012 [35]
***Staphylococcus aureus* Infectious Disease Ontology (IDOSA) [7, 19, 36]**	Released on June 22, 2012 [37]
Schistosomiasis Ontology (IDOSCHISTO)* [38, 39]	Most recent version uploaded on October 23, 2013 [40]
HIV Ontology (IDOHIV)* [41, 42]	Most recent version uploaded to BioPortal on April 4, 2017 [43]
IDO Tuberculosis (IDOTUB) **IDO Infective Endocarditis**	Planned, not yet in development

2.1 Case Study: IDOSA and methicillin resistant Staphylococcus aureus

IDO:*protective resistance* is used to model also the resistance of certain bacteria to antibiotic drugs. For this purpose, the following subtypes of *protective resistance* are asserted in IDO Core:

drug resistance =def. Protective resistance that mitigates the damaging effects of a drug
antibiotic resistance =def. Drug resistance that mitigates the damaging effects of an antibiotic

Beta-lactam antibiotics such as methicillin are the most widely used antibiotics, and most work by preventing bacterial cell wall construction. Beta-lactam antibiotics act by binding to and inhibiting the penicillin-binding-proteins (PBPs) within bacteria that facilitate the synthesis of peptidoglycan molecules, thus compromising the structural integrity of the cell wall. In response to the widespread use of Beta-lactam antibiotics, some bacteria have rapidly evolved novel-structured PBPs which lack an affinity for these antibiotics, thus rendering them less effective. At a certain level, the description of antibacterial resistance seems to require a negative aspect; thus, the appeal to the fact that the PBPs of resistant bacteria lack an affinity for beta-lactam antibiotics. But negative characterizations of a phenomenon at one level of biological reality often belie its positive aspects at another level [19]. An important desideratum in the construction of realist ontologies is to avoid as far as possible the use of definitions involving negative differentia (such as 'non-mammal', 'non-contagious', 'not part of an infection'). In compliance with this design principle of positivity, an ontologically correct

representation of resistance will reveal the active mechanisms that produce resistance—in the case of resistant bacteria, the active dispositions inhering in novel-structured PBPs that inhibit antibiotics from manifesting their damaging effects [19].

Consider the case of beta-lactam antibiotic resistance: *methicillin resistant Staphylococcus aureus* (MRSa). MRSa's resistance to methicillin is conferred by PBP2a, a PBP that lacks affinity for methicillin and is the product of the gene *mecA*. The need to provide a coherent and consistent understanding of the mechanisms underlying MRSa antibiotic resistance is one impetus for the development of the *Staphylococcus aureus* Infectious Disease Ontology (IDOSA), an extension of IDO covering entities specific to Staph aureus (Sa) infectious diseases [7, 19, 36].

IDOSA's main hierarchy is built on BFO, and imports IDO Core in full. IDOSA provides terms covering all entities relevant to antibacterial resistance in Sa, including terms for proteins, genes, gene products, biological processes, and antibacterials. In addition, it imports

- terms for Sa relevant proteins from the Protein Ontology [44]
- the terms *gene, gene group,* and *mobile genetic element* from the Sequence Ontology [45]
- the term *mecA* – representing the gene responsible for the production of PBP2a in MRSa – from the Vaccine Ontology (VO) [46]
- terms for biological processes such as *peptidoglycan biosynthesis process* and *cytolysis in other organism* (a subclass of *killing of cells of other organisms*) from the GO Biological Process Ontology
- the term *gram-positive bacterium-type cell wall* from the GO Cellular Component Ontology
- terms representing common anatomical sites of Sa infections, such as *bone*, *lung*, and *endocardium*, from the UBERON anatomy ontology [47]
- terms for several beta-lactam antibiotics, including *ciprofloxacin*, *methicillin*, and *penicillin* from Chemical Entities of Biological Interest (ChEBI) [48]
- the term *staphylococcus aureus* from NCBITaxon

IDOSA also introduces several terms specific to the Sa domain, such as:

- *ccr gene complex* and *mec gene complex,* added as subclasses of *gene group*
- *SCCMec*, which is the central determinant for broad-spectrum beta-lactam resistance encoded by mecA, added as a subclass of *mobile genetic element*.

The IDOSA subclasses *methicillin-resistant Staphylococcus aureus* and *methicillin susceptible Staphylococcus aureus* are defined as follows:

methicillin-resistant Staphylococcus aureus =def. Organism of type Staphylococcus aureus that has resistance to beta-lactam antibiotics

methicillin-susceptible Staphylococcus aureus =def. Organism of type Staphylococcus aureus that lacks resistance to beta-lactam antibiotics

Both are defined in terms of IDOSA *resistance to beta-lactam antibiotic*, which itself is a subclass of IDO:*antibiotic resistance* and defined as follows:

resistance to beta-lactam antibiotic =def. Antibiotic resistance that mitigates the damaging effects of a beta-lactam antibiotic

With these terms and definitions, we can characterize both *Methicillin-susceptible Staphylococcus aureus*'s (MSSa) susceptibility, and MRSa's resistance, to beta-lactam antibiotics in terms of *protective resistance* and blocking dispositions [7, 19]).

MSSa is susceptible to the damaging effects of methicillin because it *lacks protective resistance* to that drug. Characterized positively, MSSa's PBPs have the disposition to undergo a methicillin PBP binding process that *negatively_ regulates* the synthesis of peptidoglycan, thereby interfering with the formation of a stable cell wall. Affinity for methicillin thus acts as a blocking disposition for the PBPs' disposition to synthesize peptidoglycan.

In the case of MRSa, in contrast, the disposition of its PBP2a parts to synthesize peptidoglycan, and thereby participate in the construction of a stable cell wall (which *negatively_regulates* methicillin binding), cannot be blocked. Thus, MRSa's protective *antibiotic resistance* to methicillin can be seen as an active response in which PBP2a manifests a disposition to mitigate the damaging effects of methicillin. [19] shows how the formal representation of these relations can be used in association with instance data to draw inferences that may facilitate automated drug discovery and guide treatment decisions in specific types of cases.

The IDO Core account of protective resistance can be applied also to other cases, such as the resistance against HIV conferred by CCR5-Δ32, and the resistance against malaria conferred by the sickle cell trait [19]. CCR5-Δ32 is a deletion mutation of the CCR5 gene resulting in cells which lack a functioning CCR5 receptor on their surfaces. In this case, the disposition of individuals with the CCR5-Δ32 mutation to develop cells that lack CCR5 on their surface acts as a blocking disposition for the disposition of HIV to bind to a CCR5 molecule. *Plasmodium falciparum*, one of the infectious organisms that causes malaria, has a disposition to spread through host red blood cells, a process that is reduced in dense, dehydrated red blood cells. In individuals with the sickle cell hemoglobin gene, red blood cells have a disposition to become dehydrated and thus increase in density. This disposition acts as a blocking disposition for the disposition of plasmodium to spread through red blood cells, a process requiring hydrated red blood cells.

2.2 Partitioning of the IDO suite and a lattice of infectious disease ontologies

While currently existing IDO extensions were designed as direct extensions of the Core, we can now take advantage of the fact that IDO extensions can be partitioned into subgroups on the basis of pathogen-type. Under this partitioning, the following reference ontologies serve as direct extensions of IDO Core: IDO Bacteria, IDO Virus, IDO Fungus and IDO Parasite. IDOSA, IDOMEN, IDOTUB and IDOBRU are reengineered as extensions of IDO Bacteria. IDOFLU, IDOHIV, IDODEN and CIDO extend from IDO Virus, while IDOSCHISTO and a new ontology for malaria (replacing IDOMAL) will extend from IDO Parasite.

In similar manner IDO extension ontologies for vector-borne diseases will from a new ontology – IDO VectorBorne – consisting of just those terms needed to deal with vector-borne diseases in a pathogen neutral fashion.

[36] shows how IDOSA annotations of genetic, phenotypic, and demographic data on Sa isolates maintained by the Network on Antimicrobial Resistance in *Staphylococcus aureus* (NARSA) [49] can be used to infer a lattice or network of application ontologies for specific subfamilies of Sa-related diseases, down to the level of specific strains. The method is generalizable to isolate repositories across the infectious disease domain. Leveraging the other extension ontologies within the IDO suite (**Table 1**), allows us to generate similar lattices for specific subfamilies of coronavirus-related diseases, influenza virus-related diseases, and so on. Together these automatically form a larger network of infectious disease ontologies under IDO Core (**Figure 3**). The resultant network

of IDO ontologies can be used to define a strategy for constructing a taxonomy of infectious diseases incorporating both high-throughput genetic and molecular data as well as clinical data. The network can also be used for rapidly creating new ontologies for novel pathogens or novel strains in a way that provides a pathway for automatic linking of emerging data to legacy data relating to existing pathogens and diseases. The IDO suite of ontologies can thereby contribute to the advance of what is called 'personalized' or 'precision' medicine, which depends upon effective classification and association of biological disease data with known clinical phenotypes and disease types at ever finer levels of detail.

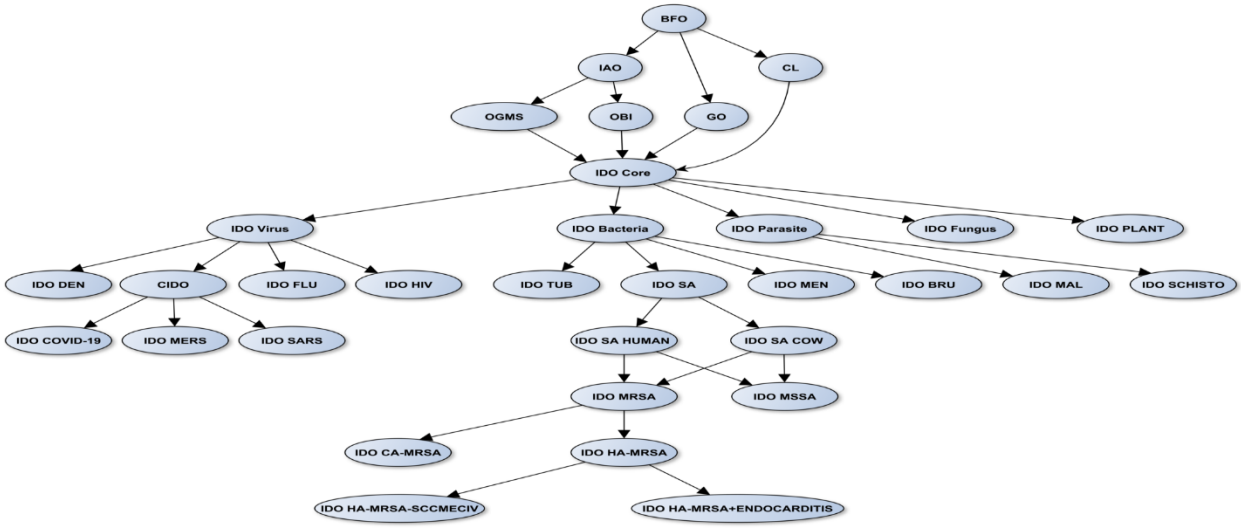

Figure 2 | A lattice of infectious disease ontologies

2.3 IDOBRU

IDOBRU, the Brucellosis Infectious Disease Ontology, is maintained by the He research team at the University of Michigan, and is used to facilitate the integration and exchange of brucellosis information stored in widely used databases, including:

- The *Brucella* Bioinformatics Portal [50]: a portal for the search and analysis of individual *Brucella* genes; linked to more than 20 other databases and programs.
- The Vaccine Investigation and Online Information Network (VIOLIN) [51] a central repository for literature related to, and data resulting from, vaccine research that uses VO [46].

IDOBRU exhibits a well-organized hierarchy with BFO, OGMS, and IDO Core imported in full, and is a good exemplar of the IDO Core hub and spokes model, as illustrated in **Table 2**.

Table 2 | IDOBRU Hierarchy

IDOBRU Axis	Top Level IDOBRU Classes	Imported OBO Ontology Class from which it descends
host infection and zoonotic disease transmission	*process of establishing Brucella infection in host*, *Brucella infectious disposition*,	*process of establishing an infection* (IDO Core) *zoonotic disposition* (IDO Core)

	Brucella host role	*infectious agent host role* (IDO Core)
virulence factors and pathogenesis	*Brucella virulence factor*, *Brucella virulence factor disposition* *brucellosis pathogen role*	*virulence factor* (IDO Core), *virulence factor disposition* (IDO Core), *pathogen role* (IDO Core)
symptoms	*brucellosis symptom*	*symptom* (OGMS)
diagnosis	*brucellosis diagnosis*	*diagnosis* (OGMS)
intentional release	*Brucella intentional release*	*planned process* (OBI)
vaccine prevention	*brucellosis vaccine*	*vaccine* (VO)
treatment	*brucellosis treatment*	*treatment* (OGMS)

In [8] IDOBRU classes are used to provide formalized representations of:

- The mechanisms by which *Brucella* successfully establishes an infection in the host, including the crucial role played by *Brucella* virulence factors *Brucella* VirB1 protein and *Brucella* lipopolysaccharide (LPS) in the bacterium's ability to survive and replicate within the vacuolar macrophage compartments of host cells
- Brucellosis diagnoses using a PCR assay to test a *Brucella* gene *omp-2* encoding for an outer membrane protein from patient's blood sample
- The intentional use of aerosolized *Brucella* as a weapon of bioterrorism, as well as the use of bleach for the purpose of disinfection in the case of a *Brucella* bioterrorist attack
- WHO's recommended standard treatment for uncomplicated brucellosis cases in adults and children that are 8 years of age or older (using *Brucella* vaccine terms imported from VO)

He's group has also used the ontology to provide a formal treatment of host-brucella interactions [25] and as the basis for an online IDOBRU SPARQL query interface [52].

2.4 IDOPlant

IDOPlant [34] is a plant infectious disease ontology being developed under the auspices of the Planteome Project, which maintains a large database of annotations from plant genomic and phenomic studies [53]. IDOPlant forms part of the biotic plant stresses branch of the Planteome Plant Stress Ontology. It leverages IDO Core in axioms such as the following:

- IDOPlant:*process of establishing a Xanthomonas oryzae infection* subclass-of IDOPlant:*process of establishing a plant bacterial infection*,
- IDOPlant:*process of establishing a plant bacterial infection* subclass-of IDO:*process of establishing an infection*
- IDOPlant:*rice bacterial leaf blight disease* subclass-of IDOPlant:*plant bacterial disease*
- IDOPlant:*plant bacterial disease* subclass-of IDOPlant:*plant infectious disease*
- IDOPlant:*plant infectious disease* subclass-of IDO:*infectious disease*

2.5 IDOSCHISTO

IDOSCHISTO is an extension of IDO Core focusing on schistosomiasis, a waterborne infectious disease caused by *Schistosoma* helminth parasites [38, 39]. It leverages IDO Core in axioms such as:

- NCIBTaxon:*Schistosoma* subclass-of IDO:*infectious agent*

IDOSCHISTO imports from NCBITaxon classes for a variety of snail species that serve as intermediary hosts for the disease, such as *bulin* and *biomphalaria*. The relation *has_intermediary_host* was created for IDOSCHISTO to link *Schistosoma* species to their intermediary hosts. Snail hosts are then linked to the class *geographical_location*, which has as subclasses the following defined classes:

- *intestinal schistosomiasis area* ≡ *geographical location* **AND** (*adjacent_to* **SOME** (*location_of* **SOME** *bulinus*))
- *urinary schistosomiasis area* ≡ *geographical location* **AND** (*adjacent_to* **SOME** (*location_of* **SOME** *biomphalaria*))

Given the information encoded in the IDOSCHISTO concerning the role that the *bulinus* and *biomphalaria* species play in the transmission of *Schistosoma*, these classes can be used in combination with IDOSCHISTO classes such as *human distribution*, *snail distribution*, and *parasite distribution* (which are subclasses of *population distribution*) to model risk factors for the spread of intestinal schistosomiasis and urinary schistosomiasis [39].

IDOSCHISTO is designed to support epidemiological monitoring systems and has been used to annotate and query data from epidemiological investigations in Senegal. IDOSCHISTO maintains an instance hierarchy that contains terms for districts, as well as nearby water bodies, with adjacency relations asserted between them. The *located_in* relation is asserted between water bodies and the snail species that inhabit them, while *Schistosoma* species are linked to snail species via the *has_ intermediary _host* relation. For example, Ndiaw is a district lying on the south bank of the Senegal River. Given the information encoded in IDOSCHISTO, a SPARQL query regarding which types of *schistosomiasis* the *ndiaw* population is exposed to will return the answers: *intestinal_schistosomiasis* and *urinary_schistosomiasis* [39].

3 IDO: New Developments

As with any ontology, IDO is continuously evolving. In this section we detail current developments of the IDO suite, including: IDO Virus (VIDO); the Coronaviruses Infectious Disease Ontology (CIDO); and an extension of CIDO focusing on COVID-19 (IDO-COVID-19).

3.1 The importation of disease terms from the Human Disease Ontology (DOID)

New disease terms were created for IDOMAL, IDOBRU, IDOMEN, IDODEN and IDOSA. But in line with OBO Foundry principles, they are being updated to import disease terms from DOID [54, 55]. IDOFLU, IDOSCHISTO, and CIDO import, respectively, the terms DOID:*influenza*, DOID: *schistosomiasis*, and DOID:*COVID-19*. Like IDO, DOID incorporates the OGMS definition of disease. But while some DOID disease terms are defined partially in terms of underlying disorders (such as DOID:*COVID-19*), there is no consistency on this front. A better template is provided by IDOSA: *Staphylococcus aureus infectious disease*, defined as "Infectious disease that has a staphylococcus aureus infectious disorder as its material basis." We will work with the DOID team to achieve consistency with the OGMS definition moving forward.

3.2 Treatment of symptoms in IDO Core and its extensions

While traditional disease classifications have been hampered by the fact that they placed too much emphasis on similarities in symptoms, the study of symptoms is still an important factor in any disease domain. The OGMS and IDO communities are working to refine the treatment of symptoms using the following definition:

- OGMS:*symptom* =def. Process experienced by the patient, which can only be experienced by the patient, that is hypothesized to be clinically relevant.

In line with this, all subclasses of *symptom* appearing within individual IDO extensions will henceforth be consistently classified as *processes*. (The treatment of symptoms in IDOPlant is a special case [34], since the OGMS definition requires a sentient host that is of a type that can report its experiences. In IDOPlant, *plant disease symptom* is defined as "Feature of a plant that is of a type that can be hypothesized to be involved in the realization of a plant disease".)

3.3 Infectious Disease Epidemiology and Surveillance

Following the methods outlined in sections 1 and 2, we have updated IDO Core to enhance its coverage of infectious disease epidemiology. IDO Core contains qualities of disease affected populations, such as *infectious disease incidence rate*, *infectious disease mortality rate*, and *infectious disease endemicity*, and we have recently added terms for the corresponding sites at which these qualities are instantiated. For example: *infectious disease endemic site*, *infectious disease free site*, and *infectious disease non-endemic site*. Adding these classes allows them to serve as parent classes for the brucellosis specific classes *brucellosis endemic site, brucellosis free site, and brucellosis non-endemic site* currently used in IDOBRU.

We have also introduced to IDO Core the terms *holoendemicity, hypoendemicity,* and *mesoendemicity* (previously from IDOMAL [30]) to represent the varying degrees to which diseases can be endemic within different populations. In several regions of sub-Saharan Africa, *holoendemicity* is frequently seen in malaria, in particular the strain caused by *Plasmodium falciparum* (with one study finding that traces of the pathogen were present in 98.6% of the population within a 4-month period [56]). *Holoendemicity* is also seen in ocular trachoma in certain areas of sub-Saharan Africa and in hepatitis B in areas of the Marquesas Islands.

Drawing on the Vector Surveillance and Management Ontology (VSMO) [57], IDO Core has been expanded with terms for infectious disease surveillance. VSMO includes the following classes, all of which are subclasses of OBI:*planned process*:

- *surveillance process* =def. Planned process with the objective to produce information about some evaluant with the purpose of, if justified by the information gathered, managing, directing, or protecting
- *pathogen surveillance* =def. Surveillance process aiming to produce information about one or several objects, in the form of microorganisms, which have the role of pathogen
- *vector surveillance* =def. Surveillance process aiming to produce information about one or several objects, in the form of arthropods, which have the role of serving as biological pathogen vectors

Both *pathogen surveillance* and *vector surveillance* have been added IDO Core. The term *surveillance process* was not appropriate for inclusion in IDO, since it is not a term that has specifically to do with infectious disease. Its children in other ontologies might include, for example, *intelligence surveillance*, *engineering surveillance*, *aircraft health surveillance*, *drug adverse event surveillance*, and so forth. It will instead be moved up to OBI. Several IDO extension ontologies already contain coverage of infectious

disease surveillance for the corresponding pathogen, and these ontologies will now be re-engineered as appropriate. IDOFLU, in particular, has an extensive treatment of influenza surveillance as part of the Centers for Excellence in Influenza Research and Surveillance program [23]. The ontology is meant to be applicable to any virus sequence and surveillance collection project, consolidating sequence and surveillance terms from a variety of online databases. IDOSCHISTO contains a module devoted to the epidemiology of schistosomiasis. It contains classes such as *prevention strategy* and *control strategy*, each of which has a variety of sub-classes pertaining to the prevention and control of schistosomiasis.

The following is an example of a hierarchy of disease surveillance terms which were created for IDOBRU (further superclasses appearing above infectious disease surveillance are omitted here):

infectious disease surveillance
    human infectious disease surveillance
        human brucellosis surveillance
    veterinary infectious disease surveillance
        swine brucellosis surveillance
        cat brucellosis surveillance
        goat brucellosis surveillance
        camel brucellosis surveillance
        horse brucellosis surveillance
        sheep brucellosis surveillance

Infectious disease surveillance will be promoted to IDO Core and then imported from there to IDOBRU.

IDO Core's expanded coverage of epidemiology will be developed through collaboration with the developers of the epidemiology focused ontologies Apollo Structured Vocabulary (Apollo-SV) [58] or Genomic Epidemiology Ontology (GenEpio) [59]. GenEpio itself reuses a number of important IDO Core terms and IDO extensions will for its part incorporate appropriate terms from Apollo-SV and GenEpio. As in all biomedical ontology, each of the mentioned ontologies evolves in response to term requests from users in order to accommodate the community's needs.

Apollo-SV provides a standardized vocabulary for terms and relations required for the interoperation between epidemic simulator models and public health application software that interface with these models. Apollo-SV, which draws heavily on the Information Artifact Ontology (IOA) [60], provides a variety of pertinent terms including:

- Terms pertaining to disease control, including: *infectious disease control strategy* (a child of IAO:*plan specification*), which has subclasses such as *vector control strategy*, *place closure control strategy*, *travel-related infectious disease control strategy*, and *quarantine control strategy*, the first three of which have more specific subclasses of their own; processes such as *infectious disease control strategy execution*
- Terms pertaining to disease surveillance, including: *disease surveillance objective specification*; subclasses of IAO:*algorithm* such as *disease transmission model* and *epidemic model*; *simulation software* (a child of IAO:*software*) and its subclasses, including *disease transmission model software*, *infectious disease forecasting software*, and *pathogen evolution model*; subclasses of OBI:*planned process* such as *disease surveillance, infectious disease forecast*, and *epidemic simulating*
- Terms relating to the spread of disease, including: processes such as *infection in ecosystem* and its subclasses *endemic*, *epidemic* and *infection in population*

4 The Coronavirus Infectious Disease Ontology (CIDO), IDO Virus (VIDO) and IDO COVID-19

The Coronavirus Infectious Disease Ontology (CIDO) deals with coronavirus infectious diseases in general. CIDO imports terms from a wide range of ontologies, including IDO Core ChEBI, the National Drug File - Reference Terminology (NDF-RT), UBERON, GO, VO, and the NCBITaxon. In addition, CIDO introduces 8 terms specific to the coronavirus domain. Alignment with IDO Core, however, will require much curation.

One application of CIDO is to the analysis and integration of information on anti-coronavirus drugs to facilitate drug repurposing against COVID-19. In a recent study [20], members of the CIDO team used text mining to identify chemical drugs and antibodies effective against at least one human coronavirus infection *in vitro* or *in vivo* and then mapped these drugs to ChEBI, the Drug Ontology (DrugOn) [61], and NDF-RT, each of which provide logical axioms linking drugs to their roles and mechanisms of action. This information was then extracted for analysis. Further relations will be built into CIDO linking drugs, coronaviruses, and the conditions under which the drugs work against the coronaviruses. CIDO is also being used in ongoing work to represent vaccines against coronavirus. In [21] reverse vaccinology and machine learning is used to predict potential vaccine targets for safe and effective COVID-19 vaccine development. The CIDO team will systematically annotate these vaccine candidates, along with their formulations and host responses, while working with the VO team to ontologically model and analyze these vaccines. To facilitate vaccine design, CIDO will be used in a further study to investigate host-pathogen interactions in order to better understand protective immune mechanisms.

IDO Virus is in the later stages of its development under John Beverley and collaborators as a preparation for the re-architecting of CIDO as the extension of IDO Virus for coronavirus diseases. IDO Virus will then facilitate the construction of ontology modules focused exclusively on SARS-coronavirus, which caused the 2002-2003 SARS outbreak, and SARS-CoV-2, which is causing the current coronavirus (COVID-19) pandemic. Our immediate priority is to fast track the creation of an ontology concerning COVID-19 - IDO-COVID-19 - as an extension of CIDO. (The Protein Ontology consortium has recently created new terms dealing with SARS-CoV-2 proteins [62].)

Recall that each IDO extension ontology should be developed in a modular fashion, providing terms specific to their own domains which build upon the disease and pathogen-neutral terms imported from IDO Core. In the typical case, terms from extension ontologies are created via downward population of the upper level terms provided by IDO Core. For example:

- IDOSA:*Staphylococcus aureus infectious disease* subclass-of IDO:*infectious disease*

Higher-level terms needed within a particular extension ontology will be added to IDO Core where the term is specific to the infectious disease domain and is truly pathogen neutral. If the relevant term is included in an OBO Foundry ontology then it can be imported from there.

To create IDO Virus, relevant IDO Core terms were introduced by adding the term 'virus' to generate child terms pertaining to viral disease. Thus, to give a few examples, IDO Virus extends IDO Core by introducing terms such as:

- *virus disorder* =def. Infectious disorder that exists as a result of processes initiated by members of a virus population.
- *viral disease* =def. Infectious disease whose physical basis is a virus disorder that is clinically abnormal in virtue of the presence of the relevant virus population

- *viral disease course* =def. Infectious disease course that realizes a viral disease

paralleling the IDO Core extension of OGMS introduced earlier. For example, IDO Core:*subclinical infection* – an infection that is part of an asymptomatic host - provides resources needed to represent asymptomatic viral infections:

- *subclinical virus infection* =def. Subclinical infection for which the infectious agent population is a virus population

As viral infections need not result in the manifestation of symptoms, yet nevertheless satisfy all other criteria for inclusion as infections.

Leveraging IDO Virus, relevant terms will be added to CIDO and IDO-COVID-19. **Figure 4** provides a summary of several important links:

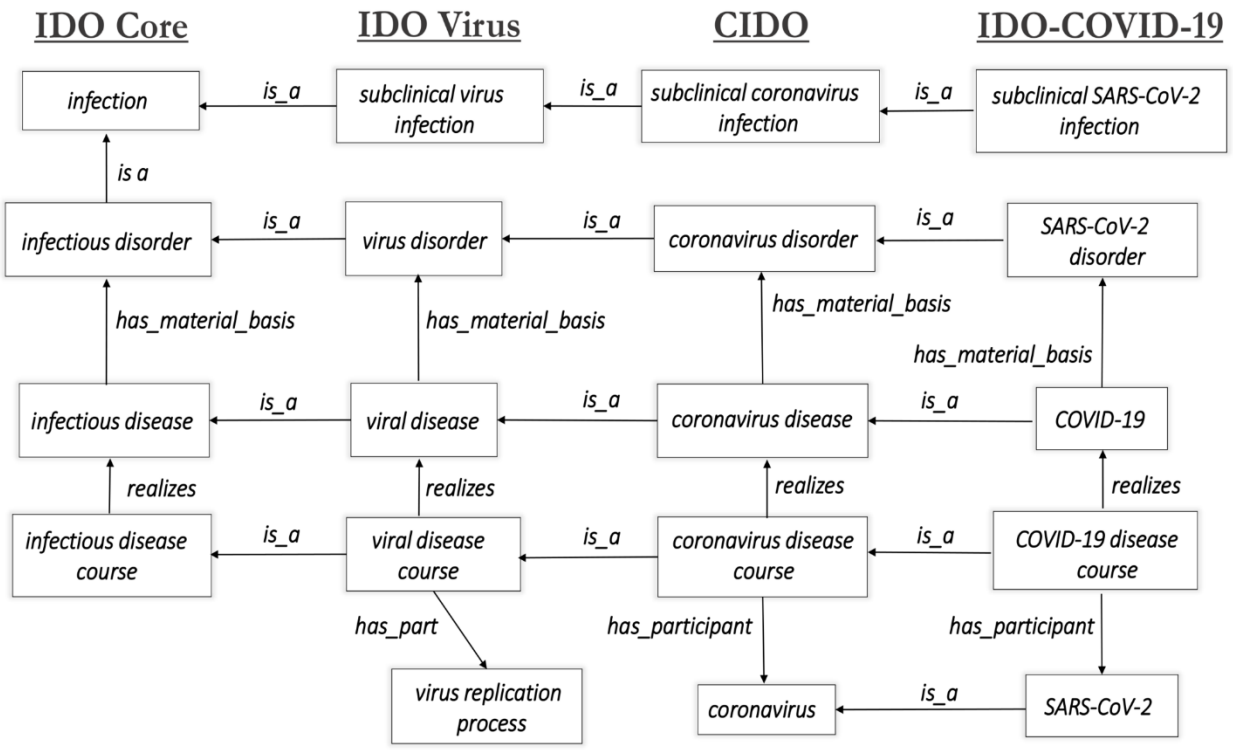

Figure 4 | Links between IDO Virus, CIDO and IDO-COVID-19

Representing asymptomatic infection is of crucial importance, as the spread of SARS-CoV-2 makes clear. IDO Virus provides CIDO and IDO-COVID-19 needed resources in that respect:

- CIDO:*subclinical coronavirus infection* subclass-of VIDO:*subclinical virus infection*
- IDO-COVID-19:*subclinical SARS-CoV-2 infection* subclass-of CIDO:*subclinical coronavirus infection*

While distinguishing these instances from symptomatic cases:

- CIDO:*coronavirus disorder* subclass-of VIDO:*virus disorder*

- IDO-COVID-19:*SARS-CoV-2 disorder* subclass-of CIDO:*coronavirus disorder*

These disorders are, moreover, the material bases for relevant diseases and disease courses, themselves illustrating links among IDO Virus, CIDO, and IDO-COVID-19:

- CIDO:*coronavirus disease* subclass-of VIDO:*viral disease*
- IDO-COVID-19:*COVID-19* subclass-of CIDO:*coronavirus disease*
- CIDO:*coronavirus disease course* subclass-of VIDO:*viral disease course*
- IDO-COVID-19:*COVID-19 disease course* subclass-of CIDO:*coronavirus disease course*

Where the relevant disease courses have the expected participants:

- CIDO:*coronavirus disease course* has participant CIDO:*coronavirus*
- IDO-COVID-19:*COVID-19 disease course* has participant IDO-COVID-19:*SARS-CoV-2*

Viral disease courses involve process parts which are useful for characterizing the development of a given virus in a host. This in mind, IDO Virus imports terms from GO [2] to reflect such processes:

- VIDO:*viral disease course* has part GO:*virus replication process*
- VIDO:*viral disease course* has part GO:*virus synthesis process*

Allowing the following connections to CIDO:

- CIDO:*coronavirus disease course* has part CIDO:*coronavirus synthesis process*
- CIDO:*coronavirus synthesis process* subclass-of GO:*virus synthesis process*
- CIDO:*coronavirus disease course* has part CIDO:*coronavirus replication process*
- CIDO:*coronavirus replication process* subclass-of GO:*virus replication process*

And then IDO-COVID-19:

- IDO-COVID-19:*SARS-CoV-2* subclass-of CIDO:*coronavirus*
- IDO-COVID-19:*COVID-19 disease course* has part IDO-COVID-19:*COVID-19 synthesis process*
- IDO-COVID-19:*SARS-CoV-2 synthesis process* subclass-of CIDO:*coronavirus synthesis process*
- IDO-COVID-19:*COVID-19 disease course* has part IDO-COVID-19: *SARS-CoV-2 replication process*
- IDO-COVID-19:*SARS-CoV-2 replication process* subclass-of CIDO:*coronavirus replication process*

The introduction in IDO Virus of the many stages of viral reproduction will assist researchers in classifying specific mechanisms of efficacy for given antivirals.

To see how, consider IDO Core includes a class populated by antiviral instances, and this class is extended in IDO Virus:

- *viricidal disposition* =def. Antiviral disposition to kill viruses
- *viricidal* =def. Antiviral bearing the viricidal disposition
- *virostatic disposition* =def. Antiviral disposition to inhibit the lifecycle of viruses

- *virostatic* =def. Antiviral bearing the virostatic disposition

Logical definitions can also be constructed for these terms, such as:

- *viricidal* ≡ *antiviral* **AND** *has_disposition* **SOME** *viricidal disposition*
- *viricidal disposition* ≡ *antiviral disposition* **AND** *inheres_in* **SOME** (*antiviral*) **AND** *realized_in* **ONLY** (*process* **AND** *results_in* **SOME** (*virus death temporal boundary*))
- *virostatic* ≡ *antiviral* **AND** *has_disposition* **SOME** *virostatic disposition*
- *virostatic disposition* ≡ *antiviral disposition* **AND** *inheres_in* **SOME** (*antiviral*) **AND** *realized_in* **ONLY** (*process* **AND** *negatively_regulates* **SOME** *viral reproduction*))

By introducing terms such as:

- *virus death temporal boundary* =def. Organism death temporal boundary that marks the end of the life cycle of a virus
- *virus birth temporal boundary* =def. Organism birth temporal boundary that marks the beginning of the life of a virus
- *virus reproduction* =def. Reproduction process involving the production of a virus containing some portion of genetic material inherited from a parent virus

Themselves relying on IDO Core terms such as IDO:*organism death temporal boundary*, IDO:*organism birth temporal boundary*, and GO:*reproduction*, the latter defined as the production of individuals containing some portion of genetic material inherited from one or more parent organisms. These terms allow for characterization of negative regulation of viruses via drugs targeting specific parts of the virus reproduction process in these ontologies, for example in CIDO:

- CIDO:*negative regulation of coronavirus replication* subclass-of IDO Virus:*negative regulation of virus replication*
- CIDO:*negative regulation of coronavirus replication* negatively regulates **SOME** CIDO:*coronavirus replication process*
- CIDO:*negative regulation of coronavirus synthesis* negatively regulates **SOME** CIDO:*coronavirus synthesis process*

And in the case of IDO-COVID-19:

- IDO-COVID-19:*negative regulation of SARS-CoV-2 replication* subclass-of CIDO:*negative regulation of coronavirus replication*
- IDO-COVID-19:*negative regulation of SARS-CoV-2 synthesis* subclass-of CIDO:*negative regulation of coronavirus synthesis*
- IDO-COVID-19:*negative regulation of coronavirus synthesis* negatively regulates **SOME** IDO-COVID-19: *SARS-CoV-2 synthesis process*

Such classes provide resources needed to annotate and unify existing data concerning coronavirus antivirals in general, and COVID-19 antivirals in particular. Given the pressing need for progress in combatting the spread of these viruses in humans, consolidating and interpreting such data is of paramount importance. As is having terms needed for viral disease monitoring, such as the

following terms in IDO Virus extending IDO:*infectious disease epidemic* and IDO:*infectious disease pandemic*, respectively:

- *viral disease epidemic* =def. Infectious disease epidemic largely involving viruses
- *viral disease pandemic* =def. Infectious disease pandemic consisting of viral disease epidemics

Themselves easily extended to CIDO:

- *coronavirus epidemic* =def. Viral disease epidemic largely involving members of the species *Coronaviridae*.
- *coronavirus pandemic* =def. Viral disease pandemic consisting of coronavirus epidemics

And from CIDO to IDO-COVID-19:

- *COVID-19 epidemic* =def. Coronavirus epidemic largely involving members of the virus strain *SARS-CoV-2*
- *COVID-19 pandemic* =def. Coronavirus pandemic consisting of COVID-19 epidemics

Refining and expanding these terms in IDO Virus, CIDO, and IDO-COVID-19 is still in progress, but made much easier by importing IDO Core as a starting point.

A similar strategy to that illustrated above will be used to link IDO Virus to existing virus ontologies, for example:

- IDOFLU:*influenza A virus infection* is a subclass of VIDO:*virus infection*
- IDOHIV:*HIV virus infection* is a subclass of VIDO:*virus infection*

Additionally, as should be clear, IDO Virus provides terms needed to distinguish bacterial, fungal, and viral infectious agents, infectious agent dispositions realized in relevant infectious processes. This makes possible deploying the preceding strategy to the construction of IDO-Bacteria, IDO-Fungus, and IDO-Parasite, and the linking of these reference ontologies to IDO Core. We can then use these reference ontologies to distinguish between viral, bacterial, fungal, and parasitic infections, between viral, bacterial, fungal, and parasitic infectious diseases, and so on, linking existing ontologies covering more specific bacteria, fungi, and parasites to IDO Core.

Other ontology initiatives being developed to support curation of COVID-19 data outside the IDO framework 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 [63]
- 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 [64]
- 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 [65]. (At present this is little more than a list of datasets rather than a bona fide ontology.)
- The NASA Jet Propulsion Laboratory's COVID-19 Research Knowledge Graph builds a knowledge graph from the COVID-19 Open Research Dataset (CORD-19) [66]

These are all stand-alone initiatives, and are thus subject to the silo problems documented in the foregoing.

5 IDO Applications

Ontology metadata can be used to combine heterogeneous bodies of research data to enable structured querying and analysis [67]. [68] applies these same methods to immunology data using IDO related ontologies such as OBI [16] and VO [42]. The following is an incomplete list of databases to which IDO ontology annotations are applied:

- The Eukaryotic Pathogen Genomics Database [69]; genomic and other data for eukaryotic pathogens including *Cryptosporidium*, *Giardia*, *Plasmodium*, *Theileria*, *Toxoplasma*, and *Trichomonas* strains.
- VectorBase: Bioinformatics Resource for Invertebrate Vectors of Human Pathogens [70]; genomic and other data for a variety of invertabrae vectors of human pathogens; until recently, several terms from IDOMAL (29, 71), as well as terms from the Mosquito Insecticide Resistance Ontology (MIRO) [72], were used in the annotation of its datasets. Both of these ontologies have since been obsoleted. A team of researchers at the University of Pennsylvania led by Chris Stoeckert is in the process of replacing the obsoleted IDOMAL and MIRO terms used in VectorBase with suitable terms from OBO Foundry ontologies.
- Recently the National Institute of Allergy and Infectious Diseases awarded a new 5-year contract, worth up to $7.2 million in 2019-2020, to support the integration of the Eukaryotic Pathogen Genomics Database database and VectorBase [73]. The Stoeckert team has since joined these resources into one bioinformatics resource, the Eukaryotic Pathogen, Host & Vector Genomics Resource (VEuPathDB). The project involves the leveraging of ontologies to expand the harmonization of semantic terms across all sites. Understanding of what the data stored in VEuPathDB is about is encoded through the use of the VEuPathDB application ontology [74]. IDO Core will play a role in this project, as the VEuPathDB ontology imports IDO Core terms such as: *human pathogenicity disposition*, *infection*, *infection prevalence*, *pathogen role*, and *primary infection*, each of which are used in the annotation of its datasets.
- Influenza Research Database [75]; makes use of IDO Core and IDOFLU.
- Virus Pathogen Resource [76]
- Pathogen-Host Interaction Data Integration and Analysis System (77), and the related:
- Victors virulence factors database (77)

In addition to providing a basis for database interoperability, IDO annotations can also serve a variety of other purposes, including [6]:

- Enhanced interpretation of data from genome-wide and high-throughput experiments.
- Use in software tools for the analysis and interpretation of microarray data, and in the development of new bioinformatic approaches to analysis of such data.
- The integration of text-mining approaches with microarray data to facilitate disease gene identification.

Reflecting their use in supporting knowledge re-use and automated reasoning, ontologies have been implemented in a variety of applications for the enhancement of patient diagnosis, care management and clinical decision support [78-82]. A brief overview and further references are provided in [83]. In the fields of infectious disease Clinical Decision Support Systems (CDSSs) are commonly used in diagnostic assistance, guidance in the prescription of anti-infectives, biosurveillance, and vector control. The use of ontologies in CDSSs is increasingly on the rise:

- Use of antibiotic decision support systems (ADSSs) has been shown to be effective in mitigating inappropriate antibiotic prescribing and lowering local antimicrobial resistance [84-86]. To facilitate interoperability and widespread circulation of future ADSSs, a Bacterial Clinical Infectious Disease Ontology (BCIDO) has been developed from IDO Core [87].
- IDDAP, a recently developed ontology-driven clinical decision support system for infectious disease diagnosis and antibiotic prescription makes heavy use of IDO Core [88].
- The Dengue Decision Support System (DDSS) is an ontology driven computational application developed at Colorado State University [89, 90] with the aim of guiding the implementation of locally appropriate Dengue and Dengue Vector control programs. The DDSS is used in conjunction with Chaak, a cell phone-based system for i) the field capture of data relating to Dengue vector surveillance; and ii) the rapid transfer of the data to the central DDSS database [91]. The DDSS makes use of IDOMAL (21), as well as the related ontologies MIRO [72] and VSMO [57]. Given its use of obsoleted ontologies the DDSS requires reengineering.

DISCUSSION

Successful information-driven research on COVID-19 needs to be able to integrate the already massive and exponentially growing body of research and data concerning coronavirus diseases. Ontologies such as IDO Virus, CIDO and IDO-COVID-19 fill this need by providing a standardized, computer-interpretable representation of heterogenous coronavirus knowledge. Another potential application of this work is to ontology-based deep learning, for instance as illustrated in [92], which describes a novel method for learning features of entities such as proteins and viruses from their associations to ontology classes, and describes how this method can be employed for fast identification of virus–host interactions that can shed light on potential treatments and drug discoveries. Because these ontologies are built as part of an interoperable suite of IDO ontology modules, it becomes easier, for example, to compare COVID-19 to other respiratory diseases such as SARS, MERS, and influenza – along multiple dimensions including underlying disorders, pathogen features (such as strain, virulence factors, and drug resistance), host-pathogen interactions, routes of transmission, anatomical sites of infection, genetic and environmental variables, symptoms, diagnostic criteria, disease courses, prevention measures and so on.

The utility of the IDO framework turns not least on the fact that we will continue to face the threat of novel viruses (as well as bacteria and parasites) in the future, and the IDO suite provides an easy to follow recipe for building new pathogen-specific ontologies in a way that allows easy comparison, along multiple dimensions, of novel pathogens and diseases with pathogens and diseases about which data have already been assembled.

The IDO strategy also brings about a situation in which there is a community trained in how to build, use, understand and correct both single ontologies and groups of ontologies that fit together. This promotes coordination of data curation as it relates to data crossing disciplinary boundaries, for example between vaccine research and pathogen genomics. Moreover, the IDO corpus provides a set of rigorously curated definitions of terms used in infectious disease research

that can be employed to provide a useful vehicle for cross-disciplinary collaboration, allowing specialists in one sub-domain to rapidly gain an understanding of the meanings of the technical terms used in neighboring domains.

Finally, the IDO ontologies can contribution to addressing a further urgent problem faced not merely by COVID-19 research but by contemporary biomedical research in general. This is the problem of *reproducibility*. This problem applies not only to scientific findings which are the results of experimental studies, but also to findings deriving from the application of different types of diagnostic tests. For an experiment, or a test, to be reproducible, it is crucial that we have a clear understanding of how the experimental or test results were obtained. For this to be possible, however, it is crucial that the constituent processes are described in a terminology that is widely used and whose terms are well defined. We believe that, when used in combination with the Ontology for Biomedical Investigations [16], IDO offers a promising strategy for the creation of comparable, integratable and discoverable provenance metadata for the data generated in infectious disease research.

Conclusions

As we face the continued threat of novel pathogens in the future, IDO provides a simple recipe for building new pathogen-specific ontologies in a way that allows data about novel diseases to be easily compared, along multiple dimensions, with already curated data from earlier diseases. IDO's tightly coordinated suite of ontologies modules thus provides a powerful method of data integration and sharing that will allow physicians, researchers, and public health organizations to respond rapidly and efficiently both to the current and future public health crises.

Methods

IDO Core is developed using the Protégé ontology development tool [93], leveraging the enhanced expressivity of the Web Ontology Language (OWL).

While IDO Core was designed to provide coverage of the infectious disease domain generally, we never considered IDO Core to be complete and intended to have its expansion be driven by term requests. As with most OBO ontologies, IDO Core is an open project with its own GitHub repository [94], where the most recent published and developmental versions of the ontology are available for download. We encourage members of the ontology community, as well as infectious disease researchers, to submit term requests to our GitHub Issues tracker. The Issues tracker can also be used to report any errors or concerns related to the ontology. Before requesting a new term, please search online ontology repositories such as Ontobee and BioPortal to see if the needed term already exists. Further advice is available in the OBO tutorial for term requests [95]. Once a term request is received, it will be reviewed by the main IDO Core team to determine whether the term is most appropriate for IDO Core, one of its extensions, or another OBO ontology. If the term is within IDO Core's scope, then it will be added with a formal definition, written in conjunction with the term requestor to ensure biological accuracy as well as adherence to OBO Foundry best practices and consistency with IDO logical structure. We can assign a unique ID for the term so that it can be used for immediate annotation prior to the definition being finalized.

Abbreviations

Apollo-SV: Apollo Structured Vocabulary; BCIDO: Bacterial Infectious Disease Ontology; BFO: Basic Formal Ontology; CIDO: Coronavirus Infectious Disease Ontology; ChEBI: Chemical Entities of Biological

Interest; CL: Cell Ontology; DDSS: Dengue Decision Support System; DrugOn: Drug Ontology; GenEpio: Genomic Epidemiology Ontology; GO: Gene Ontology; IDOBRU: Brucellosis Ontology; IDO Core; Infectious Disease Ontology Core; IDODEN: Dengue Ontology; IDOFLU; IDOHIV: HIV Ontology; Influenza Ontology; IDOMAL: Malaria Ontology; IDOMEN: Meningitis Ontology; IDOPlant: Plant Disease Ontology; IDOSCHISTO: Schistosomiasis Ontology; IDOSA; *Staphylococcus aureus* Infectious Disease Ontology; IOA: Information Artifact Ontology; MIRO: Mosquito Insecticide Resistance Ontology; NARSA: Network on Antimicrobial Resistance in *Staphylococcus aureus*; NDF-RT: National Drug File - Reference Terminology; NCBITaxon: NCBI organismal classification; OBI: Ontology for Biomedical Investigations; OBO: Open Biomedical Ontologies; OGMS: Ontology for General Medical Science; OWL: Web Ontology Language; RO: Relations Ontology; VEuPathDB: Eukaryotic Pathogen, Host & Vector Genomics Resource; VIDO: IDO Virus; VO: Vaccine Ontology; VSMO: Vector Surveillance and Management Ontology.

Declarations

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Availability of Data and Materials
The datasets generated and/or analysed during the current study are freely publicly available in the IDO Core GitHub repository [https://github.com/infectious-disease-ontology/infectious-disease-ontology] as well as online ontology repositories such as Ontobee [http://www.ontobee.org/ontology/IDO] and BioPortal [http://www.ontobee.org/ontology/IDO]. IDO extensions are also freely publicly available on Github, Ontobee and BioPortal.

Competing interests
The authors declare that they have no competing interests.

Funding
BS's contributions were supported by the NIH under NCATS 1UL1TR001412 (Buffalo Clinical and Translational Research Center).

Author's contributions
All authors read and extensively reviewed the manuscript. SB wrote the manuscript and conducted the research. JB vastly improved the structure of the paper, as well as the sections on CIDO and IDO Virus. BS and LGC were principal developers of IDO Core and IDOSA. BS is a principal developer of BFO, OGMS, and IDOPlant. JB is principal developer of IDO Virus and SB is a contributor. LGC was a significant contributor to the development of BCIDO and VSMO.

Acknowledgements
We would like to acknowledge the following for their contributions to the development of IDO Core: Alexander Diehl, Alan Ruttenberg, Albert Goldfain, Bjoern Peters, and Jie Zhang. This paper has benefitted from feedback from Alexander Diehl, Chris Stoeckert, Yongqun (Oliver) He, Asiyah Yu Lin, and Werner Ceusters.

References

1. Pesquita C, Ferreirra JD, Couto FM, Silva MJ. The epidemiology ontology: an ontology for the semantic annotation of epidemiological resources. *J Biomed Semant*. 2014; doi:10.1186/ 2041-1480-5-4.
2. 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.
3. 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.
4. The Open Biomedical Ontologies Foundry. http://obofoundry.org/. Accessed 27 Apr 2020.
5. Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, et al. Relations in biomedical ontologies. *Genome Biol*. 2005; 6:R46. doi:10.1186/gb-2005-6-5-r46.
6. Cowell LG, Smith B. Infectious Diseases Ontology. In: Sintchenko V, editor. *Infectious Disease Informatics*. New York, NY: Springer; 2010. p. 373-95.
7. 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.
8. 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.
9. Scheuermann RH, Ceusters W, Smith B. Toward an ontological treatment of disease and diagnosis. *AMIA Summit on Translat Bioinform*. 2009; p. 116-120.
10. Arp R, Smith B, Spear A. *Building Ontologies with Basic Formal Ontology*. Cambridge, MA: MIT Press; 2015.
11. ISO/IEC 21838-2. https://www.iso.org/standard/74572.html. Accessed 27 Apr 2020.
12. ISO Standards Maintenance Portal. https://standards.iso.org/iso-iec/21838/-2/ed-1/en/. Accessed 27 Apr 2020.
13. Ceusters W, Smith B, Kumar A, Dhaen C. Mistakes in medical ontologies: where do they come from and how can they be detected? In: Pisanelli D, editor. *Ontologies in medicine*. Amsterdam: IOS Press; 2004a. p. 145–164.
14. Ceusters W, Smith B, Kumar A, Dhaen C. Ontology-based error detection in SNOMED-CT. In Fieschi M, et al., editors. *Medinfo 2004*. Amsterdam: IOS Press; 2004b. p. 482–486.
15. Bodenreider O, Smith B, Kumar A, Burgun A. Investigating subsumption in SNOMED-CT: an exploration into large description logic-based biomedical terminologies. *Artif Intell in Med*. 2007; 39:183–195. doi: 10.1016/j.artmed.2006.12.003.
16. 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.
17. Federhen S. The NCBI Taxonomy Database. *Nucleic Acids Res*. 2012; 40:D136-D143. doi: 10.1093/nar/gkr1178.
18. Pathogen Transmission Ontology. https://bioportal.bioontology.org/ontologies/PTRANS. Accessed 27 Apr 2020.
19. 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.
20. Liu Y, Chan W, Wang Z, Hur J, Xie J, Yu H, et al. Ontological and bioinformatic analysis of anti-coronavirus drugs and their implication for drug repurposing against COVID-19. *Preprints*. 2020; Available at: https://doi.org/10.20944/preprints202003.0413.v1. Accessed 27 Apr 2020.

21. Ong E, Wong M, Huffman A, He Y. COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. *bioRxiv*. 2020; Available at: https://doi.org/10.1101/2020.03.20.000141. Accessed 27 April 2020.
22. Coronavirus Infectious Disease Ontology. https://bioportal.bioontology.org/ontologies/CIDO. Accessed 27 Apr 2020.
23. 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.
24. Influenza Ontology. https://bioportal.bioontology.org/ontologies/FLU. Accessed 27 Apr 2020.
25. 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.
26. Brucellosis Ontology. https:// bioportal.bioontology.org/ontologies/IDOBRU. Accessed 27 Apr 2020.
27. Mitraka E, Topalis P, Dritsou V, Dialynas E, Louis C. Describing the breakbone fever: IDODEN, an ontology for dengue fever. *PLOS Negl Trop Dis*. 2015; 9(2): e0003479. doi: 10.1371/journal.pntd.0003479.
28. Dengue Ontology. https://bioportal.bioontology.org/ontologies/IDODEN. Accessed 27 Apr 2020.
29. Topalis P, Mitraka E, Bujila I, Deligianni E, Dialynas E, Siden-Kiamos I, et al. IDOMAL: an ontology for malaria. *Malar J*. 2010; 9:230. doi: 10.1186/1475-2875-9-230.
30. Topalis P, Mitraka E, Dritsou V, Dialynas E, Louis C. IDOMAL: the malaria ontology revisited. *J Biomed Semant*. 2013; doi: 10.1186/2041-1480-4-16.
31. Malaria Ontology. https:// github.com/VeuPathDB-ontology/IDOMAL. Accessed 27 Apr 2020.
32. Béré C, Camara G, Malo S, Lo M, Ouaro S. IDOMEN: an extension of infectious disease ontology for MENingitis. In: Ohno-Machado L, Séroussi B, editors. *MEDINFO 2019: Health and Wellbeing e-Networks for All*. Amsterdam: IOS Press; 2019. p. 313-317.
33. Meningitis Ontology. https://github.com/cedricbere/IDOMEN. Accessed 27 Apr 2020.
34. Walls RL, Smith B, Elser J, Goldfain A, Stevenson DW, Jaiswal P. A plant disease extension of the infectious disease ontology. In: Cornet R, Stevens R, editors. *Proceedings of the 3rd International Conference on Biomedical Ontology*. CEURS-WS.org; 2012. p. 1-5.
35. Plant Disease Ontology. http://purl.obolibrary.org/obo/idoplant.owl. Accessed 27 Apr 2020.
36. Goldfain A, Smith B, Cowell LG. Constructing a lattice of infectious disease ontologies from a staphylococcus aureus isolate repository. In: Cornet R, Stevens R, editors. *Proceedings of the 3rd International Conference on Biomedical Ontology (ICBO 2012)*. CEURS-WS.org; 2012. p. 1-5.
37. *Staphylococcus aureus* Infectious Disease Ontology. https://github.com/awqbi/ido-staph. Accessed 27 Apr 2020.
38. Camara G, Desprès S, Djedidi R, Lo M. Building a schistosomiasis process ontology for an epidemiological monitoring system. In: Faucher C, Jain L, editors. *Innovations in Intelligent Machines-4, Studies in Computational Intelligence*. Springer International Publishing; 2014. p. 75–99.
39. Camara G, Desprès S, Lo M. IDOSCHISTO: une extension de l'ontologie noyau des maladies infectieuses (IDO-Core) pour la schistosomiases. In: Faron-Zucker C, editor. *IC – 25èmes Journées francophones d'Ingénierie des Connaissances*, *Clermont-Ferrand, France. Session 1: Construction, peuplement et exploitation d'ontologies*. 2014. p. 39-50.
40. Schistosomiasis Ontology. https://github.com/gaoussoucamara/idoschisto. Accessed 27 Apr 2020.
41. Sargeant D, Deverasetty S, Luo Y, Villahoz Baleta A, Zobrist S, Rathnayake V, et al. HIVToolbox, an integrated web application for investigating HIV. *PLOS ONE*. 2011; 6:e20122. doi: 10.1371/journal.pone.0020122.

42. Sargeant D, Deverasetty S, Strong CL, Alaniz IJ, Bartlett A, Brandon NR, et al. The HIVToolbox 2 Web System Integrates Sequence, Structure, Function and Mutation Analysis. *PLOS ONE*. 2014; 9:e98810. doi: 10.1371/journal.pone.0098810.
43. HIV Ontology. https:// bioportal.bioontology.org/ontologies/HIV. Accessed 27 Apr 2020.
44. Natale DA, Arighi CN, Blake JA, Bona J, Chen C, Chen S, et al. Protein Ontology (PRO): enhancing and scaling up the representation of protein entities. *Nucleic Acids Res*. 2017; 45: D339-D346. doi: 10.1093/nar/gkw1075.
45. Eilbeck K, Lewis SE, Mungall CJ, Yandell M, Stein L, Durbin R, et al. The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol*. 2005; 6(5):R44. doi: 10.1186/gb-2005-6-5-r44.
46. He Y, Cowell LG, Diehl AD, Mobley H, Peters B, Ruttenberg A, et al. VO: Vaccine Ontology. In: Smith B, editor. *Proceedings of the 1st International Conference on Biomedical Ontology (ICBO 2009)*. Buffalo: NCOR; 2009. p. 172.
47. Haendel MA, Balhoff JP, Bastian FB, et al. Unification of multi-species vertebrate anatomy ontologies for comparative biology in Uberon. *J Biomed Semant*. 2014; doi:10.1186/2041-1480-5-21.
48. 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.
49. Network on Antimicrobial Resistance in *Staphylococcus aureus*. http://www.narsa.net/. Accessed 27 Apr 2020.
50. Xiang Z, Zheng W, He Y. BBP: Brucella genome annotation with literature mining and curation. *BMC Bioinform*. 2006; doi: 10.1186/1471-2105-7-347.
51. Vaccine Investigation and Online Information Network. http://www.violinet.org/. Accessed 27 Apr 2020.
52. IDOBRU SPARQL. http://www.phidias.us/bbp/idobru/sparql/index.php. Accessed 27 Apr 2020.
53. Cooper L, Meier A, Laporte MA, Elser J, Mungall CJ, Sinn BT, et al. The Planteome database: an integrated resource for reference ontologies, plant genomics and phenomics. *Nucleic Acids Res*. 2018; 46:D1168–D1180. doi: 10.1093/nar/gkx1152.
54. Schriml LM, Arze C, Nadendla S, Chang YW, Mazaitis M, Felix V, et al. Disease ontology: a backbone for disease semantic integration. *Nucleic Acids Res*. 2012; 40:D940–D946. doi: 10.1093/nar/gkr972
55. Kibbe WA, Arze C, Felix V, Mitraka E, Bolton E, Fu G, et al. Disease ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data, *Nucleic Acids Res*. 2015; 43:D1071–D1078. doi: 10.1093/nar/gku1011.
56. Stich A, Oster N, Abdel-Aziz IZ, Stieglbauer G, Coulibaly B, Wickert H, et al. Malaria in a holoendemic area of Burkina Faso: a cross-sectional study. *Parasitol Res*. 2006; 98:596–599. doi: 10.1007/s00436-005-0104-9.
57. Lozano-Fuentes S, Bandyopadhyay A, Cowell LG, Goldfain A, Eisen L. Ontology for vector surveillance and management. *J Med Entom*. 2013; 50:1-14. doi: 10.1603/me12169.
58. Hogan WR, Wagner MM, Brochhausen M, Levander J, Brown ST, Millet N. The Apollo Structured Vocabulary: an OWL2 ontology of phenomena in infectious disease epidemiology and population biology for use in epidemic simulation. *J Biomed Semant*. 2016; 7:50. doi: 10.1186/s13326-016-0092-y.
59. Griffiths E, Dooley D, Graham M, Van Domselaar G, Brinkman FSL, Hsiao WWL. Context Is Everything: Harmonization of Critical Food Microbiology Descriptors and Metadata for Improved Food Safety and Surveillance. *Front Microbiol*. 2017; 8:1068. doi: 10.3389/fmicb.2017.01068.

60. Ceusters W, Smith B. About: towards foundations for the information artifact ontology. In: Couto FM, Hasting J, editors. *Proceedings of the 6th International Conference on Biomedical Ontology (ICBO 2015)*. CEUR-WS.org; 2015. p. 1-5.
61. Hogan WR, Hanna J, Joseph E, Brochhausen M. Towards a Consistent and Scientifically Accurate Drug Ontology. In: Dumontier M, Hoehndorf R, Baker CJO, editors. *Proceedings of the 4th International Conference on Biomedical Ontology (ICBO 2013)*. CEUR-WS.org; 2013. p. 68-73.
62. https://proconsortium.org/download/development/pro_sars2.obo; https:// proconsortium.org/download/ development/pro_sars2.gpi. Accessed 27 Apr 2020.
63. WHO COVID-19 Rapid Version CRF. https:// bioportal.bioontology.org/ ontologies/ COVIDCRFRAPID. Accessed 27 Apr 2020.
64. COVID-19 Surveillance Ontology. https:// bioportal.bioontology.org/ontologies/COVID19. Accessed 27 Apr 2020.
65. Linked COVID-19 Data Ontology. https://github.com/Research-Squirrel-Engineers/COVID-19. Accessed 27 Apr 2020.
66. COVID-19 Research Knowledge Graph. https://github.com/nasa-jpl-cord-19/covid19-knowledge-graph. Accessed 27 Apr 2020.
67. Scheuermann R, Kong M, Dahlke C, Cai J, Lee J, Qian Y, et al. Ontology-based knowledge representation of experiment metadata in biological data mining. In: Chen J, Lonardi S, editors. *Biological Data Mining*. Boca Raton, FL: Chapman & Hall; 2009. p. 529-559.
68. Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen Jieming, et al. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Sci Data*. 2018; 5:180015. doi: 10.1038/sdata.2018.15.
69. Eukaryotic Pathogen Genomics Database. https://eupathdb.org/eupathdb/. Accessed 27 Apr 2020.
70. VectorBase: Bioinformatics Resource for Invertebrate Vectors of Human Pathogens. http://vectorbase.org. Accessed 27 Apr 2020.
71. Lawson D, Arensburger P, Atkinson P, Pesanky NJ, Bruggner RV, Butler R, et al. VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Res*. 2009; 37: D583-D587. doi: 10.1093/nar/gkn857.
72. Dialynas E, Topalis P, Vontas J, Louis C. MIRO and Irbase: IT tools for the epidemiological monitoring of insecticide resistance in mosquito disease vectors. *PLOS Negl Trop Dis*. 2009; 3(6):e465. doi: 10.1371/journal.pntd.0000465.
73. https://eurekalert.org/pub_releases/2019-10/uop-nat101519.php. Accessed 27 Apr 2020.
74. Zheng J, Cade JS, Brunk B, Roos DS, Stoeckert CJ, Sullivan SA, et al. Malaria study data integration and information retrieval based on OBO Foundry ontologies. In: Jaiswal P, Hoehndorf R, editors. *Proceedings of the Joint International Conference on Biological Ontology and BioCreative (ICBO-BioCreative 2016)*. CEUR-WS.org; 2016. p. 38.
75. Zhang Y, Aevermann BD, Anderson TK, Burke DF, Gwenaelle D, Gu Z, et al. Influenza Research Database: An integrated bioinformatics resource for influenza virus research. *Nucleic Acids Res*. 2017; 45:D466-D474. doi: 10.1093/nar/gkw857.
76. Virus Pathogen Resource. http://www.viprbrc.org.
77. 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.
78. Zhang YF, Gou L, Zhou TS, Lin DN, Zheng J, Li Y, et al. An ontology-based approach to patient follow-up assessment for continuous and personalized chronic disease management. *J Biomed Inform*. 2017; 72:45–59. doi: 10.1016/j.jbi.2017.06.021.

79. Carletti G, Giuliodori P, DiPietri V, Peretti A, Amenta F. An ontology-based consultation system to support medical care on board seagoing vessels. *Int Marit Health*. 2016; 67(1):14–20. doi: 10.5603/IMH.2016.0004.
80. Maurice P, Dhombres F, Blondiaux E, Friszer S, Guilbaud L, Lelong N, et al. Towards ontology-based decision support systems for complex ultrasound diagnosis in obstetrics and gynecology. *J Gynecol Obstet Hum Reprod*. 2017; 46(5):423–29. doi: 10.1016/j.jogoh.2017.03.004
81. Abidi S. A knowledge-modeling approach to integrate multiple clinical practice guidelines to provide evidence-based clinical decision support for managing comorbid conditions. *J Med Syst*. 2017; 41(12):193. doi: 10.1007/s10916-017-0841-1.
82. Lin Y, Staes CJ, Shields DE, Kandula V, Welch BM, Kawamoto K. Design, development, and initial evaluation of a terminology for clinical decision support and electronic clinical quality measurement. *AMIA Annu Symp Proc*. 2015; p. 843–51.
83. Haendel MA, McMurry JA, Relevo R, Mungall CJ, Robinson PN, Chute CG. A Census of Disease Ontologies. *Annu Rev of Biomed Data Sci*. 2018; 1(1):305-331.
84. Evans RS, Pestotnik SL, Classen DC, Clemmer TP, Weaver LK, Orme JF Jr, et al. A computer-assisted management program for antibiotics and other antiinfective agents. *N Engl J Med*. 1998; 338:232-8.
85. Thursky KA, Mahemoff M. User-centered design techniques for a computerized antibiotic decision support system in an intensive care unit. *Int J Med Inform*. 2007; 76:760-8.
86. Paterson DL. The role of antimicrobial management programs in optimizing antibiotic prescribing within hospitals. *Clin Infect Dis*. 2006; 42 Suppl 2: S90-5. doi: 10.1086/499407.
87. Gordon CL, Pouch S, Cowell LG, Boland MR, Platt HL, Goldfain A, et al. Design and evaluation of a bacterial clinical infectious diseases ontology. *AMIA Annu Symp Proc*. 2013; p. 502– 511.
88. Shen Y, Yuan K, Chen D, Colloc J, Yang M, Li Y, et al. An ontology-driven clinical decision support system (IDDAP) for infectious disease diagnosis and antibiotic prescription. *Artif Intell Med*. 2018; 86:20–32.
89. Eisen L, Coleman M, Lozano–Fuentes S, McEachen N, Orlans M, Coleman M. Multi-disease data management system platform for vector-borne diseases. *PLOS Negl Trop Dis*. 2011; 5:e1016. doi: 10.1371/journal.pntd.0001016.
90. Lozano-Fuentes S, Barker CM, Coleman M, Park BB, Reisen WK, Eisen L. Emerging information technologies to provide improved decision support for surveillance, prevention, and control of vector-borne diseases. In: Jao C, editor. *Efficient Decision Support Systems: Practice and Challenges in Biomedical Related Domain*. Rijeka, Croatia: InTech-Open Access Publisher; 2011. p. 89-114.
91. Lozano-Fuentes S, Wedyan F, Hernandez-Garcia E, Devadatta S, Ghosh S, Bieman JM, et al. Cell phone-based system (Chaak) for surveillance of immatures of dengue virus mosquito vectors. *J Med Entomol*. 2013; 50(4):879-89. doi: 10.1603/me13008.
92. Liu-Wei W, Kafkas Ş, Chen J, Tegnér J, Hoehndorf R. Prediction of novel virus–host interactions by integrating clinical symptoms and protein sequences. *bioRxiv*. 2020; Available at: https://doi.org/10.1101/2020.04.22.055095. Accessed 27 Apr 2020.
93. Protégé. http://protege.stanford.edu. Accessed 27 Apr 2020.
94. https://github.com/infectious-disease-ontology/infectious-disease-ontology. Accessed 27 Apr 2020.
95. https://github.com/jamesaoverton/obo-tutorial/blob/master/docs/ontology-development.md#filing-effective-term-requests-and-bug-reports. Accessed 27 Apr 2020.

Additional Files

Additional File 1: Supplementary Tables (.docx)
Table S1 | Ontologies building on the OGMS treatment of disease and diagnosis; Table S2 | Overview of IDO extension ontologies that have been developed or planned; Table S3 | Some other ontologies within the infectious disease domain that make use of IDO Core.

Additional File 2: The Infectious Disease Ontology Extensions: Some Issues. (.docx)
Several of the IDO ontologies require significant reengineering if they are to be considered bona fide extensions of IDO Core. This supplementary document provides an overview of some issues concerning specific IDO extensions, while providing some suggestions for how they can be addressed.