The Infectious Disease Ontology in 2020

Shane Babcock a, John Beverley, b Lindsay G. Cowell c, Barry Smith d

a Niagara University, Lewiston NY, USA
b Northwestern University, IL, USA
c University of Texas Southwestern Medical Center TX, USA
d University at Buffalo, NY, USA

Abstract
The Infectious Disease Ontology (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 within various areas of infectious disease research, 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 the creation of IDO Virus, the Coronaviruses Infectious Disease Ontology (CIDO) and an extension of CIDO focused on COVID-19, IDO-CovID-19. We also 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.

1 INTRODUCTION

1.1 Data Silos and the OBO Foundry

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, 2], 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 is typically only locally accessible. The resultant silo-formation hinders both translational and comparative research and preventive and prognostic public health research.

As the experience of biologists and bioinformaticians has shown, ontologies are an effective data sharing tool [3]. 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,1 a collective of

1 http://obofoundry.org/
developer groups dedicated to creating, testing and maintaining a collection of ontologies based on an evolving set of design principles for ontology development [4], including:

- 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.

These principles were modelled initially on the practices of the Gene Ontology (GO) [3], 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 ontologies are created utilizing terms and relational expressions taken from existing Foundry ontologies and from the OBO Foundry Relation Ontology (RO) [5]. This practice - which applies also to the definitions in these ontologies - 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 a common ID-space means that the source of each term can be immediately identified by its prefix. Use of a common ID-space also ensures the ontologies retain backward compatibility with legacy annotations as those ontologies evolve. Moreover, 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’ by specialization. Following this model, the Infectious Disease Ontology (IDO) represents one step towards overcoming data silos [6].

1.2 The Infectious Disease Ontology

IDO Core is designed to be disease and pathogen neutral by covering just those entities that are relevant to infectious diseases generally, providing coverage for aspects of infectious disease across biological scale (e.g. gene, cell, organ, organism, population), disciplinary perspective (e.g. clinical, biological, epidemiological), relevant organism type (e.g. 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, with more narrowly focused ontologies extending from the core, providing coverage of all aspects of the infectious disease domain, including biomedical research, clinical care, and public health [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 from Foundry ontologies. 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]. Relevant data and information within multiple disparate sources is 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

---

See section 6.
team. The resultant annotated data thereby becomes 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 specific extension ontology [8], all of which are important for overcoming data silo problems in 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 existing IDO extensions in such a way as to yield greater conformance.

2 THE INFECTIOUS DISEASE ONTOLOGY

2.1 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 has as its coverage domain those types of entities relevant to clinical encounters between doctors and patients. 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], and is the official top-level ontology of the OBO (Open Biomedical Ontologies) Foundry. BFO has recently been published as international standard ISO/IEC 21838-2.3

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, for example Celiac disease shares many symptoms in common with, say, Chron’s disease, or perhaps none at all, for example Clostridioides difficile infection 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. Distinguishing manifestations of symptoms from dispositions to manifest symptoms provides the flexibility needed to represent, say, asymptomatic patients nevertheless disposed to manifest symptoms, and make sense of certain physician prescriptions such as, say, that patients on antibiotics continue treatment after symptoms have subsided. In addition to distinguishing between disease, signs and symptoms of disease, and pathological processes, OGMS further distinguishes between:

Disease diagnosis
A disease’s realization in a totality of processes, called a disease course
Underlying disorder(s) within the patient in which the disease is rooted.

3 https://www.iso.org/standard/74572.html, https://standards.iso.org/iso-iec/21838/-2/ed-1/en/
Each distinction provides needed flexibility in characterizing disease-related phenomena [9]. Different clinicians may diagnose the same disease differently; a disease may exist without having been yet diagnosed. Moreover, the same disease may have distinct disease courses and perhaps distinct underlying disorders in distinct patients. Indeed, the same disease may be manifested in a wide variety of different types of disease courses and clinical pictures depending upon the particular patient. And since disorders can exist before they are realized in overt pathological processes, the OGMS approach can capture the existence of pre-clinical manifestations of disease, and clinical risk factor combinations of disease and predispositions to disease which can exist within a single patient (as when an instance of disease of type A in a given patient is a risk factor for a second disease of type B). Conflating these distinctions – as seen in even widely-used disease vocabulary resources such as SNOMED-CT [11-13] - 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 record, and failures to accurately measure progress in tackling disease spread.

2.2 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 follows:

---

4'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 OGMs terms are defined as follows:

\[ \text{Disorder} = \text{def} \text{ A material entity which is clinically abnormal and part of an extended organism; disorders are the physical basis of disease} \]

\[ \text{Disease} = \text{def} \text{ A disposition to (i) undergo pathological processes that (ii) exists in an organism because of one or more disorders in that organism} \]

\[ \text{Disease course} = \text{def} \text{ A process that is the totality of all processes through which a given disease instance is realized} \]

The relevant IDO terms descending from these OGMs terms are defined as follows:

\[ \text{Infectious disorder} = \text{def} \text{ An infection that is clinically abnormal} \]

\[ \text{Infectious disease} = \text{def} \text{ A disease whose physical basis is an infectious disorder} \]

\[ \text{Infectious disease course} = \text{def} \text{ A disease course that is a realization of an infectious disease} \]

In addition, IDO Core defines:

\[ \text{Infection}^5 = \text{def} \text{ A part of an extended organism that itself has an infectious agent population as part and that:} \]

\[ \begin{align*}
(1) & \text{ Exists as a result of processes initiated by members of the infectious agent population and is,} \\
(2) & \text{ Clinically abnormal in virtue of the presence of this infectious agent population or,} \\
(3) & \text{ Has a disposition to bring clinical abnormality to immunocompetent organisms of the same Species as the host (the organism corresponding to the extended organism) through transmission of a member or offspring of a member of the infectious agent population.} \\
\end{align*} \]

To unpack this definition, we first introduce further needed definitions used by IDO Core. For example, IDO Core imports the class \textit{organism} from the Ontology for Biomedical Investigations (OBI) [14], as well as classes for a variety of organisms that can play the role of infectious agent, such as \textit{bacteria} and \textit{virus}, both imported from the NCBI organismal classification (NCBITaxon).\(^6\) Additionally, the class \textit{extended organism} is imported from OGMS, and is defined as an 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 [9]. To represent infectious agents and infectious agent populations, IDO Core introduces the following definitions:

\[ \text{infectious agent} = \text{def} \text{ An organism that has an infectious disposition} \]

\(^5\) For lack of suitable pathogen specific extensions, the following subclasses of infection are currently included in IDO Core (these terms will be ported out once appropriate extension ontologies are developed):

\[ \begin{align*}
\text{amebiasis} & = \text{def} \text{ An infection located in the colon and that has as part organisms of the Species Entamoeba histolytica} \\
\text{candidiasis} & = \text{def} \text{ An infection that has as part organisms of the Genus Candida} \\
\text{leptospiriasis} & = \text{def} \text{ An infection that has as part organisms of the Genus Leptospira} \\
\text{Shigellosis} & = \text{def} \text{ An infection located in the bowel and that has as part organisms of the Genus Shigella} \\
\text{trichomoniasis} & = \text{def} \text{ An infection that has as part organisms of the Species Trichomonas vaginalis} \\
\end{align*} \]

\(^6\) \url{http://www.obofoundry.org/ontology/ncbitaxon.html}
infectious agent population = _def_ An organism population whose members each have an infectious disposition.

As well as the following terms:

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

infectious disposition = _def_ A 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

Furthermore, IDO Core defines various processes involved in the establishment of infections:

colonization of host = _def_ An establishment of localization in host process in which an organism establishes itself in a host

establishment of a clinically abnormal colony = _def_ A colonization of host process that results in a clinically abnormal colony

process of establishing an infection = _def_ A process by which an infectious agent, established in a host, becomes part of an infection in the host

According to IDO Core, an infectious agent realizes its infectious disposition only when all of (1)-(4) occur. When an infectious agent colonizes and establishes a clinically abnormal colony in a host, it acquires an infectious agent role, and realizes aspects of its infectious disposition, such as becoming part of a disorder in the host (i.e. participates in a process of establishing an infection). But an infectious agent does not bear the infectious agent role when established within a host with which it is commensal. While the infectious agent realizes certain aspects of its infectious disposition (i.e. is transmitted to and colonizes the host), it doesn’t establish a clinically abnormal colony.

The definition of ‘infectious disposition’ makes reference to pathogenic dispositions. IDO Core defines these dispositions and related pathogen terms:

pathogen = _def_ A material entity\(^7\) bearing a pathogenic role

pathogen role = _def_ A 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\(^8\)

pathogenic disposition = _def_ A disposition to initiate processes that results in a disorder

With these definitions introduced, the complexity of the definition of ‘infection’ can be unpacked, in particular, clauses (2) and (3). Clause (2) ensures that infections are clinically abnormal owing to relevant infectious agents. Suppose extended organism S has two distinct parts themselves with infectious agent population parts, A and B. Suppose A is dormant while B is active in S, so that B

\(^7\)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.

\(^8\)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.
results in S manifesting symptoms while A does not. Clause (2) requires infection by B be tied to clinical abnormality stemming from B. Related, clause (3) allows infection by A to be represented since the relevant part of S which has infectious agent A as part at least has the disposition to initiate pathological processes in other members of S’s species, given transmission of A or A’s offspring to those members. Importantly, this definition of ‘infection’ allows for representation of commensal populations, infectious disorders that are caused by organisms that are typically commensal, and the fact that asymptomatic infected individuals are contagious (as we see now with SARS-CoV-2). There is currently no term for commensal populations in IDO Core, but the term should be added.

2.3 Transmission of Pathogens

IDO Core recognizes the importance of characterizing infectious disease transmission in its various forms, and relevant class definitions such as:

\[
\text{transmission process} = \text{def} \quad \text{A process that is the means during which the pathogen is transmitted directly or indirectly from its natural reservoir, a susceptible host or source to a new host}
\]

\[
\text{host} = \text{def} \quad \text{An organism bearing a host role}
\]

\[
\text{host role} = \text{def} \quad \text{A role borne by an organism in virtue of the fact that its extended organism contains a material entity other than the organism}
\]

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

\[
\text{infectious agent vector} = \text{def} \quad \text{An organism bearing an infectious agent vector role}
\]

\[
\text{infectious agent vector role} = \text{def} \quad \text{An infectious agent transporter role}\text{ that is borne by an organism active in the transfer of an infectious agent to an organism of another Species and in which the agent is infectious}
\]

\[
\text{indirect pathogen transmission process}^{10} = \text{def} \quad \text{A transmission process during which a pathogen is indirectly transferred from a reservoir, source or host to another host by intermediary vehicles, vectors, or as airborne dust particles}
\]

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:

\[
\text{intermediate host} = \text{def} \quad \text{An organism bearing an intermediate host role}
\]

\[
\text{intermediate host role} = \text{def} \quad \text{A symbiont host role borne by an organism in virtue of the fact that its partner in symbiosis utilizes the host to undergo development stage transition, and the host is required for continuation of the partner's life cycle}
\]

\[
\text{\textsuperscript{9}}\text{The superclass ‘infectious agent transporter role’ is defined as the role realized in a process that transfers an infectious organism from one location to another, which is borne by a material entity in virtue of the fact that an infectious agent is located in or on the entity and the entity has the capability to transfer (either actively or passively) the infectious agent from one location to another.}
\]

\[
\text{\textsuperscript{10}}\text{In the current version of IDO Core, this term has the label “indirect.” What is presented here is a suggested label.}
\]
symbiont host role $=_{def}$ A 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.

Moreover, ‘symbiont host role’ includes subclasses such as:

- definitive host role $=_{def}$ A 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 $=_{def}$ A 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.
- paratonic host role $=_{def}$ A 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.

2.4 Pathogen Inhibition and Control

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

- antibacterial $=_{def}$ A material entity bearing an antibacterial disposition.
- antibacterial disposition $=_{def}$ A disposition to kill or inhibit the reproduction of bacteria.
- antifungal $=_{def}$ A material entity bearing an antifungal disposition.
- antifungal disposition $=_{def}$ A disposition to kill or inhibit the development or reproduction of fungal organisms.
- antiparasitic $=_{def}$ A material entity bearing an antiparasitic disposition.
- antiparasitic disposition $=_{def}$ A disposition to kill or inhibit the development or reproduction of eukaryotic parasites.
- antiviral $=_{def}$ A material entity bearing an antiviral disposition.
- antiviral disposition $=_{def}$ A disposition to kill or inhibit the lifecycle of viruses.

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, a tumor cell’s resistance to chemotherapy, 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 [15]. 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 of resistance 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_ A 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.

This term is defined to systematically account for all different kinds of resistance, 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, 15], 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 prevent 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:

\[ \text{negatively_regulates} =_\text{def} \text{The progression of x reduces the frequency, rate or extent of y} \]

Thus, we can say that X’s immunity to Y is a blocking disposition for Y’s infectious disposition, so that X’s immunity is realized in certain processes that negatively_regulates 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.

2.5 Case Study: IDOSA and methicillin resistant *Staphylococcus aureus*

A particularly important use of protective resistance within IDO is that of modelling the resistance of certain bacteria to antibiotic drugs. For this purpose, the following subclasses of protective resistance are asserted in IDO Core:

- **drug resistance** =_def_ A protective resistance that mitigates the damaging effects of a drug
- **antibiotic resistance** =_def_ A drug resistance that mitigates the damaging effects of an antibiotic

Beta-lactum antibiotics, such as methicillin, are the most widely used antibiotics, and most work by preventing bacterial cell wall construction. Beta-lactum 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-lactum antibiotics, some bacteria have rapidly evolved novel-structured PBPs which lack an affinity for these antibiotics, thus rendering these antibiotics less effective in preventing cell wall construction. 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-lactum antibiotics. But negative characterizations of a phenomenon at one level of biological reality often
belie its positive aspects at another level [15]. An important desideratum in the construction of realist ontologies is to avoid as far as possible the use of definitions involving negative differentia (e.g. ‘non-contagious’, ‘not part of an infection’). In compliance with this “positivity design principle”, an ontologically correct representation of resistance will reveal those 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 [15].

We can see these commitments on display more clearly by considering an important case of beta-lactum 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 of the impetuses for the development of the *Staphylococcus aureus* Infectious Disease Ontology (IDOSA), an extension of IDO covering entities specific to Staph aureus (Sa) infection diseases [7, 15, 16].

IDOSA’s main hierarchy is built on BFO version 2.0, is well-organized hierarchy, 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, and imports terms from several relevant ontologies such as:

- Terms for proteins imported from the Protein Ontology (PRO)\(^{11}\)
- Terms *gene*, *gene group*, and *mobile genetic element* from the Sequence Ontology (SO)\(^{12}\)
- The term *mecA* – a term representing the gene responsible for the production of PBP2a in MRSa - from the Vaccine Ontology (VO) [17] (though it would be more appropriate to use for example the NCBI Gene name/identifier)
- Terms for biological processes from the GO Biological Process Ontology, such as *peptidoglycan biosynthesis process* and *cytolysis in other organism* (a subclass of *killing of cells of other organisms*)
- 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 Uber-anatomy ontology (UBERON) [18]
- Terms for several beta-lactum antibiotics, including *ciprofloxacin, methicillin, and penicillin* from Chemical Entities of Biological Interest (ChEBI) [19]
- The term *staphylococcus aureus* from NCBITaxon

IDOSA additionally introduces several domain-specific terms, such as *ccr gene complex* and *mec gene complex*, added as subclasses of *gene group*, antibacterial representing common treatments for *Staphylococcus* infections, for example *ciprofloxacin, methicillin, and penicillin* (all imported from CHEBI; all defined as having a relevant antibacterial disposition), *SCCMec* added as a subclass of *mobile genetic element*. *SCCMec* is the central determinant for broad-spectrum beta-lactum resistance encoded by *mecA*.

Of particular importance for us here, is that IDOSA adds the subclasses *methicillin-resistant Staphylococcus aureus* and *methicillin susceptible Staphylococcus aureus*, which are defined as follows:

\[ \text{methicillin-resistant Staphylococcus aureus} \equiv_{\text{def}} \text{An organism of type Staphylococcus aureus that has resistance to beta-lactam antibiotics} \]

\(^{11}\) https://proconsortium.org/

\(^{12}\) http://www.sequenceontology.org/
methicillin-susceptible Staphylococcus aureus = def An organism of type Staphylococcus aureus that lacks resistance to beta-lactam antibiotics

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

resistance to beta-lactam antibiotic = def An 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-lactum antibiotics in terms of protective resistance and blocking dispositions [7, 15].

Consider, MSSa is susceptible to the damaging effects of methicillin because it lacks protective resistance to that drug. A positive (active) characterization of its susceptibility to the drug can be in terms of the disposition of its PBPs. The PBPs within MSSa have an affinity for methicillin, a disposition to bind to methicillin which is realized in a methicillin PBP binding process. This process negatively_regulates the synthesis of peptidoglycan, thereby interfering with the formation of a stable cell wall. In the case of MSSa, PBP’s affinity for methicillin blocks the disposition the disposition to synthesize peptidoglycan, leaving MSSa susceptible to the damaging effects of the drug. Thus, PBP’s affinity for methicillin is the active mechanism which underlies MSSa’s lack of protective resistance to the drug. By contrast, MRSa bears a protective antibiotic resistance to the drug methicillin. PBP2a has the disposition to synthesize peptidoglycan, and thereby participate in the construction of a well-formed cell wall. Because the process of cell wall construction is incompatible with, and therefore negatively_regulates, the process of methicillin binding, PBP2a’s disposition to synthesize peptidoglycan is a blocking disposition for the disposition of methicillin to bind to PBPs. Thus, MRSa’s protective resistance can be seen as an active response in which one of its parts, PBP2a, manifests a disposition to mitigate the damaging effects of methicillin.

We can, moreover, provide formalized representation of MRSa’s resistance to methicillin, following that in [15], as a set of triples representing the relevant ontological relationships (e.g., PBP has_function_realized_as_process synthesis of peptidoglycan), and a series of inference rules, which can be used along with derived facts in order to show that the resistance of MRSa can be inferred from the triples in a manner that is both logical and provides an explanation of why MRSa bears such a resistance. This is important since in the future such inference rules could potentially be used by an automated reasoner to deduce from the triples that MRSa is resistant to methicillin. In this way, an ontology containing the relevant triples and inference rules may guide treatment decisions and facilitate automated drug discovery [14].

Importantly, the account of protective resistance can be generalized to other cases [15], such as the resistance against HIV conferred by CCR5-Δ32, and the resistance against malaria conferred by the sickle cell trait. 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 blocking disposition for the disposition of HIV to bind to a CCR5 molecule. Plasmodium falciparum, one of the infectious organisms that cause malaria, have 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.
As we have seen, IDO Core aligns with OBO Foundry principles, imports a range of terms from widely used ontologies, covers a wide range of phenomena in the domain of infectious diseases, and has been used effectively to characterize an important resistance phenomenon in IDOSA. In the next section, we examine several ontologies that purport to extend IDO Core in various ways.

3 EXTENDING IDO Core

An overview of the current state of each IDO extension ontology is provided in Table 1 (see also Appendix B). Those IDO extensions marked with an asterisk will require reengineering to address various issues – see the Supplementary Materials for details. In this section we highlight some of the positives. CIDO is discussed at the end of the paper.

| Coronavirus Infectious Disease Ontology (CIDO)* [20, 21]: Most recent version uploaded to BioPortal on April 9, 2020 | Influenza Ontology (FLU)* [22]: Most recent version uploaded to BioPortal on August 20, 2015 |
| HIV Ontology (HIV)*: Most recent version uploaded to BioPortal on April 4, 2017 | Brucellosis Ontology (IDOBRU)* [8, 23]: Most recent version uploaded to BioPortal on March 28, 2015 |
| Dengue Ontology (IDODEN)* [24]: Most recent version uploaded to BioPortal on July 12, 2013 | Malaria Ontology (IDOMAL)* [25, 26]: Most recent version uploaded to BioPortal on March 16, 2015 |
| Meningitis Ontology (IDOMEN)* [27]: Not yet formally released to BioPortal; most recent version uploaded on November 27, 2019 | Plant Disease Ontology (IDOPlant) [28]: A draft version released in 2012 |
| Staphylococcus aureus Infectious Disease Ontology (IDOSA) [7, 14, 15]: Not yet formally released to BioPortal; developmental version uploaded on June 22, 2012 | Schistosomiasis Ontology (IDOSCHISTO)* [29, 30]: Not yet formally released to BioPortal; most recent version uploaded on October 23, 2013 |

Table 1

3.1 IDOBRU

IDOBRU is maintained by Yongqun He’s laboratory 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 (BBP) [31]: a portal for the search and analysis of individual Brucella genes; linked to more than 20 other databases and programs.
- The PATRIC Pathosystems resource: visualization and analysis of Brucella genomes.
- The Vaccine Investigation and Online Information Network (VIOLIN, http://www.violinet.org/) a central repository for literature related to, and data resulting from, vaccine research; VIOLIN uses VO [17].

---

13[https://github.com/cedricbere/IDOMEN](https://github.com/cedricbere/IDOMEN)
14[http://purl.obolibrary.org/obo/ idoplant.owl](http://purl.obolibrary.org/obo/ idoplant.owl)
15[https://github.com/awqbi/ido-staph](https://github.com/awqbi/ido-staph)
16[https://github.com/gaoussoucamara/idoschisto](https://github.com/gaoussoucamara/idoschisto)
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 can be seen in Table 2, each of IDOBRU’s main axes of brucellosis representation descend from top-level terms imported from the IDO Core, or from other higher-level OBO Foundry ontologies:

| IDOBRU Axis                        | Top Level IDOBRU Classes                                                                 | Parent Class and the upper level ontology it derives from                                                                 |
|-----------------------------------|-----------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|
| host infection and zoonotic disease transmission | process of establishing *Brucella* infection in host, brucellosis disposition, *Brucella* host role | process of establishing an infection (IDO Core) zoonotic disposition (IDO Core) 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)                                                                                                        |

*Table 2*

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 [23] and as the basis for an online IDOBRU SPARQL query interface.\(^{17}\)

### 3.2 IDOPlant

IDOPlant is currently being developed as part of the Planteome Project\(^ {18}\), an international collaborative effort supported by primary funding\(^ {19}\) from the National Science Foundation of USA. The Planteome project maintains a large database of annotations from genomic and phenomic studies [32], and aims to “overcome the obstacles in annotating data for complex biological concepts

---

\(^{17}\)http://www.phidias.us/bbp/idobru/sparql/index.php

\(^{18}\)http://planteome.org/

\(^{19}\)IOS:1340112 award
that span multiple ontologies by developing both the ontology terms and the software tools needed to annotate data from all aspects of plant diseases” [28]. IDOPlant is currently being incorporated within the biotic plant stresses branch of the Plant Stress Ontology (PSO)\textsuperscript{20}, a Planteome plant reference ontology providing coverage for all abiotic plant stresses, such as drought, salinity, temperature, and nitrogen deficiencies, and for all biotic plant stresses, such as pests, pathogens, symbiotic organisms, and diseases.

The Planteome Project makes use of a variety of reference ontologies for plants, many of which are integrated with IDOPlant [28]. These include external ontologies:

- The Gene Ontology (GO) [3], from which IDOPlant imports terms for the molecular functions of host and pathogen genes and those biological processes in which either the host, pathogen, or both are involved.
- Chemical Entities of Biological Interest (ChEBI)
- Phenotypic Quality Ontology (PATO) [33]

As well as ontologies internally developed by the Planteome Project, including:

- The Plant Ontology (PO), which provides a standardized, species-neutral terminology for plant anatomy, morphology and development stages [34, 35]; imported PO terms are used to describe the plant structures at which infections happen and the development stages during which disease signs are observed.
- The Plant Trait Ontology (TO)\textsuperscript{21}; IDOPlant imports TO terms describing phenotypic plant traits relevant to plant disease, including qualities such as leaf color, processes such as sudden wilting, and independent continuants such as leaf lesion.

Additionally, IDOPlant imports:

- Terms for host plants, pathogens—such as *Oryza sativa* and *Xanthomonas oryzae*—from NCBI Taxon
- Terms relating to the ecology of host plants and diseases from the Environment Ontology (ENVO) [36, 37]
- And relevant terms from OBI

Particularly crucial to plant disease research is the study of the plant structures and development stages at which infections happen and symptoms of disease are observed. As such, plant scientists need to be mindful of the occurrence of infections and plant disease symptoms at different stages across different species. Unfortunately, comparison of development stages across species is hindered by the fact that biologists often describe development stages in species- or clade-specific terms. As indicated above, IDOPlant leverages species-neutral terms from PO to integrate data about plant structure development stages across different species. IDOPlant thus overcomes traditional barriers to classification. The integration of PO plant structure development stages to plant

\textsuperscript{20} https://github.com/Planteome/plant-stress-ontology PSO is still in development and has not yet been officially released; PSO’s developers are working with the National Program Staff at the USDA-ARS to seek support for the project and its applications which will extend towards analyses of gene expression, phenotype and genotype datasets from host plants and their interacting pests, pathogens and mutualists. For details see http://planteome.org/node/6

\textsuperscript{21} https://github.com/Planteome/plant-trait-ontology
pathology data within IDOPPlant makes it “possible to explore and learn the molecular and environmental basis of plant diseases using advanced semantic methods” [38]. Integrating plant disease research in IDOPPlant additionally leverages relevant terms from IDO Core. The following illustrate important relationships between these two ontologies:

- IDOPPlant:process of establishing a Xanthomonas oryzae infection is a subclass of IDOPPlant:process of establishing a plant bacterial infection, which itself is a subclass of IDO:process of establishing an infection
- IDOPPlant:plant bacterial disease is a subclass of IDOPPlant:plant infectious disease which is itself a subclass of IDO:infectious disease
- IDOPPlant:plant pathogen role is a subclass of IDO:pathogen role

Reflecting the fact that plant diseases can be caused by either bacteria, fungi or viruses, in IDOPPlant, infectious agent has the following subclasses each of which has a logical definition in OWL built from IDO:infectious disposition:

- bacterial infectious agent ≡ bacteria AND (has_disposition SOME infectious disposition)
- fungal infectious agent ≡ fungi AND (has_disposition SOME infectious disposition)
- viral infectious agent ≡ virus AND (has_disposition SOME infectious disposition)

(As these subclasses are clearly relevant to infectious diseases generally, these terms will eventually be ported over to IDO Core.) Similarly, IDO:infectious disposition can be used in the definition of classes for specific infectious agents. For instance, IDOPPlant might define

- xanthomonas oryzae infectious agent ≡ xanthomonas oryzae AND (has_disposition SOME infectious disposition)

3.3. IDOSCHISTO

IDOSCHISTO is an extension of IDO Core focusing on schistosomiasis, a waterborne infectious disease caused by Schistosoma helminth parasites. IDOSCHISTO is designed to support epidemiological monitoring systems and the running of qualitative simulation models. To facilitate this, the IDOSCHISTO team developed the Infectious Disease Spreading Domain Ontology (IDSDO Core) [29]. IDSDO provides the process class idsdo_spreading, and its children idsdo_endemic_spreading, idsdo_epidemic_spreading and idsdo_pandemic_spreading. The spreading of schistosomiasis within a given type of geographical location can be represented using the unfolds_in relation which links the class idsdo_spreading to the class geographical_location. Geographical_location 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))

While [28] refers to this term as an IDOPPlant subclass of NCBI:taxon:Xanthomonas oryzae, it doesn’t appear in the last publicly released version of the ontology: [http://purl.obolibrary.org/obo/idoplant.owl](http://purl.obolibrary.org/obo/idoplant.owl).
IDOSCHISTO contains classes for a variety of important snail species that serve as intermediary hosts for the disease, such as *bulin* and *biomphalaria*. The relation *has_intermediary_host* was created for IDOSCHISTO in order to link snail species to the *schistosoma* species for which they are an intermediary host. Given the information encoded in the IDOSCHISTO concerning the role that the *bulinus* and *biomphalaria* snail species play in the transmission of the *schistosoma* pathogen, these classes can be used in combination with 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 [30].

IDOSCHISTO has been used to annotate and query data from epidemiological investigations in Richard Toll, Senegal. IDOSCHISTO’s individual hierarchy 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 of Richard Toll 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* [30].

3.4 IDOMEN

IDOMEN is designed to assist in the analysis and filtering of data collected on social media platforms such as Twitter to help improve the early detection of meningitis epidemic risks in sub-Saharan Africa [27]. Increasingly, digital technology platforms such as search queries, social media posts, and web server access logs have been employed for disease surveillance. Digital platforms provide real time data streams from which information related to public health can be extracted at low to virtually no cost and in a timely manner. By mining these data sources for traces of health-related activities, they can be transformed into useful metrics for inclusion in statistical estimation models for disease incidence. Google, Wikipedia and Twitter have each been investigated as tools for quantifying disease incidence rates [39-41]. Furthermore, ontology-based approaches to internet-based disease surveillance have been developed and applied with good success [42, 43].

4 IDO NEXT STEPS

As with any ontology, IDO Core and its extensions require continual maintenance. In this section we detail some general issues to be addressed in future developments of the IDO suite.

4.1 Fungal Infections

While IDO Core includes the organism classes *bacteria*, *virus*, and *parasite*, a term for fungal organisms is currently not present. On the other hand, IDO Core does contain the terms *antifungal* and *antifungal disposition*, the definitions of which both refer to fungal organisms.

4.2 Pathogens and Pathogen Roles

IDO Core’s *pathogen* class has the child *infectious agent*. The latter class has the children, *emerging pathogen*, *infectious human pathogen*, *opportunist pathogen*, *primary pathogen* and *re-emerging pathogen*. Corresponding to the *pathogen* and *infectious agent* classes are the roles *pathogen role* and its single child *infectious agent role*. But *infectious agent role* itself has no children. Consideration might be given to including children of *infectious agent role* corresponding to the children of *infectious agent*.

4.3 The importation of disease terms from the Human Disease Ontology (DOID)
For both IDOMAL and IDOSA, new disease terms were created. But in line with OBO Foundry principles, they will be updated to import disease terms from DOID [44, 45]. Similarly, disease terms that were newly created for IDOBRU, IDOMEN, IDOSCHISTO, and IDODEN should be replaced with existing DOID terms. FLU already imports DOID:influenza and CIDO imports 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 isn’t consistency on this front. A better template is provided by IDOSA: *Staphylococcus aureus infectious disease*, defined as “An 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.

4.4. Treatment of symptoms in IDO Core and its extensions

Traditional disease classifications often place far too much emphasis on similarities in symptoms. Nevertheless, the study of symptoms is still an important factor in the infectious disease domain. IDO Core imports the term OGMS:symptom, defined as “A quality of a patient that is observed by a patient or a processual entity experienced by the patient, either of which is hypothesized by the patient to be a realization of a disease.” Accordingly, the majority of the IDO extensions follow OGMS and IDO Core in asserting symptom as a direct child of entity (as symptoms fail to form a natural kind; see [9]). Symptom, though, are represented in inconsistent ways across IDO extensions. In IDOBRU and CIDO, symptom is a child of quality. Whereas in IDOMAL and IDODEN symptom is a subclass of realizable entity.

Moving forward, closer collaboration is needed to ensure consistency here. We should emphasize that the OGMS community, including various IDO experts, is working closely to refine its treatment of symptoms. The following definition is currently being considered:

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

Given this proposal, we would reassign symptom as a subclass of BFO:process within OGMS, IDO Core and its extensions. This would ensure that all subclasses of symptom appearing within individual IDO extensions are consistently classified as processes.

The treatment of symptoms in IDOPlant is a special case [28]. To represent plant disease symptoms within IDOPlant, a new term was created instead of importing OGMS:symptom. This is because the OGMS definition requires a sentient host that can report its experiences. In IDOPlant, *plant disease symptom* is defined as “A feature of a plant is of the type that can be hypothesized to be involved in the realization of a plant disease”. This is in line with the usage of “symptom” in plant pathology for those phenotypes that are associated plant diseases, where the same phenotype may be linked to a multitude of diseases. The relation *has_plant_disease_symptom* was also created for the ontology in order to link plant diseases to those phenotypes, processes and independent continuants that are used in the diagnosis of plant disease. For instance, *rice bacterial leaf blight disease has_plant_disease_symptom leaf lesion*. The fact that IDOPlant:symptom encompasses qualities and

---

23 DOID doesn’t import OGMS:disease as it should, but rather includes a term ‘disease’ with a DOID prefix that uses the OGMS definition. We will work with the DOID team to make sure the OGMS term is imported.
independent continuants in addition to processes should not be an issue as the rationale behind the proposed change to OGMS: symptom would not extend to former.

4.5 Infectious Disease Epidemiology

One aspect in which IDO Core still has some limitations is in its representation of infectious disease epidemiology. While IDO Core contains qualities of disease affected populations, such as infectious disease incidence rate, infectious disease mortality rate, and infectious disease endemicity, previously it did not contain corresponding terms for sites at which these qualities are instantiated [8]. We are in the process of adding such terms to IDO Core, 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.

Similarly, IDOMAL contains the following classes none of which are present in IDO Core: holoendemicity, hypoendemicity, and mesoendemicity [26]. In IDOMAL these are subclasses of malaria endemicity. These terms are needed in IDOMAL to represent the varying degrees to which malaria is endemic within different populations. Holoendemicity is particularly relevant. 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 [46]).

IDO Core also currently lacks terms for infectious disease surveillance. Given the significant role surveillance plays within the study of infectious diseases generally, consideration should be given to adding relevant terms to IDO Core. Related terms already exist within the Vector Surveillance and Management Ontology (VSMO) [47]. These include the following classes, all of which are subclasses of OBI: planned process:

- **surveillance process** = def A 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 A 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 A 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

*Surveillance process* would not be appropriate to IDO (or indeed, really, to VSMO) since it has nothing specifically to do with infectious disease. It would be useful, for example, for an intelligence process ontology. Instead, we will add to OBI a more specific subclass of OBI: planned process covering any surveillance of entities related to biomedicine. This class, along with OBI: planned process, will be imported to IDO Core to serve as the parent of *pathogen surveillance* and *vector surveillance*, both of which are appropriate terms to add to IDO Core. Their definitions can then easily be revised by substituting the new biomedicine-related surveillance term for *surveillance process* in the definitions.

This is particularly important for the facilitation of interoperability between the IDO suite of ontology modules, several of which already contain coverage of infectious disease surveillance:

- IDOBRU contains the term *infectious disease surveillance* which serves as the parent for brucellosis surveillance.
• FLU is applied to influenza surveillance as part of the Centers for Excellence in Influenza Research and Surveillance program [22]. 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.

• IDOMAL contains a variety of processes relevant to malaria surveillance, including *surveillance in malaria eradication* and *process of malaria epidemiology*, the latter of which has many children including *collection of epidemiological data* and *investigation relating to the insect vector*.

• IDOSCHISTO contains a whole module devoted to the epidemiology of schistosomiasis. It contains classes such as *prevention strategy* and *control strategy* each of which has a variety of children classes pertaining to the prevention and control of schistosomiasis.

*Infectious disease surveillance* should also be added to IDO Core. It should be noted that presently within IDOBRU the term falls under a hierarchy of surveillance related terms (note: all of the terms below OBI:planned process are terms that were newly created for IDOBRU):

```
OBI:planned process
  Health Surveillance
    biosurveillance
disease surveillance
  animal disease surveillance
    human disease surveillance
      infectious disease surveillance
        brucellosis surveillance
        noninfectious disease surveillance
```

All of the terms in the hierarchy from human disease surveillance up should probably be ported out to OBI. Infectious disease surveillance should be ported out to IDO Core and then imported to IDOBRU. There doesn’t seem to be need in any ontology for the term noninfectious disease surveillance, so this term should be obsoleted.

Although IDO Core’s coverage of epidemiology terms is in some ways limited, these issues can also be addressed through collaboration with the developers of epidemiology focused ontologies such as the Apollo Structured Vocabulary (Apollo-SV) [48] or Genomic Epidemiology Ontology (GenEpio) [49]. GenEpio reuses a number of important IDO Core terms and IDO extensions could incorporate the appropriate terms from Apollo-SV or 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 provides a variety of pertinent terms including:

• Terms pertaining to disease control, including: *infectious disease control strategy* (a child of Information Artifact Ontology [24] (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*.

---

24 [http://www.obofoundry.org/ontology/iao.html](http://www.obofoundry.org/ontology/iao.html)
• 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

4.6 Partitioning of the IDO suite

While currently existing IDO extensions were designed as direct extensions of the IDO Core, in the future we will be taking advantage of the fact that IDO extensions can be partitioned into subgroups on the basis of pathogen-type. On this partitioning, the following reference ontologies will serve as direct extensions of IDO Core: IDO-Bacteria, IDO-Virus, IDO-Fungus and IDO-Parasite. IDOSA, IDOMEN, IDOTUB and IDOBRU would then be reengineered as extensions of IDO-Bacteria. FLU, HIV, IDODEN and CIDO would extend from IDO-Virus, with IDOSCHISTO and IDOMAL extending from IDO-Parasite:

![Figure 2](image)

But we also need flexibility to infer hierarchies that are relevant to various applications. In the future we will define IDO extension ontologies for vector-borne diseases on the basis of an extension of IDO Core consisting of just those IDO Core terms relevant to vector-borne diseases together with all additional terms needed to deal with vector-borne diseases in a pathogen neutral fashion, IDO VectorBorne. Extension ontologies for specific vector-borne diseases would then be developed as extensions from IDO-VectorBorne [16].

5 The Coronavirus Infectious Disease Ontology (CIDO) and IDO Virus

The Coronavirus Infectious Disease Ontology (CIDO) deals with coronavirus infectious diseases in general. Though only recently created, the ontology has already been added to the OBO Foundry library. CIDO imports terms from a wide range of ontologies, including ChEBI, ICDO, NDF-RT, UBERON, GO, VO, and the NCBITaxon, among others. In addition, CIDO introduces 8 terms specific to the coronavirus domain. Alignment with IDO Core, however, will require much curation (see the Supplementary Materials for details).

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, NDF-RT and the Drug Ontology (DrugOn) [50], each of which provide logical axioms linking drugs to their roles and mechanisms of action. This information was then extracted for analysis. To fill key gaps in this information 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 rearchitecture of CIDO as the extension of IDO Virus for coronavirus diseases. IDO Virus will facilitate the construction of ontology modules focused exclusively on SARS-coronavirus, which caused the 2002-2003 SARS outbreak, and SARS-coronavirus2, which caused the ongoing coronavirus disease 2019 (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. In line with these efforts, the Protein Ontology consortium has recently created new terms dealing with SARS-CoV-2 proteins.

To illustrate how such extension relationships operate with respect to IDO Core, recall 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:

- IDOPlant:rice bacterial leaf blight disease is a subclass of IDOPlant:plant bacterial disease
- IDOPlant:plant bacterial disease is a subclass of IDOPlant:plant infections disease
- IDOPlant:plant infections disease is a subclass of IDO:infections disease
- IDOMAL:malaria is a subclass of IDO:infectious disease
- IDOSA:Staphylococcus aureus infectious disease is a subclass of IDO:infectious disease

In some cases, a higher-level term that is needed within a particular extension ontology may not be present in IDO Core. If the relevant term is specific to the infectious disease domain and is truly pathogen neutral then it should be added to IDO Core. If the relevant term is included in an OBO Foundry ontology then it can be imported from there. Each of these considerations is being adhered to in the construction of IDO Virus.

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:

- viral disorder = def An infectious disorder involving virus infectious agents
- viral disease = def An infectious disease whose physical basis is a viral disorder

---

25 The uploaded files can be found here: [https://proconsortium.org/download/development/pro_sars2.obo](https://proconsortium.org/download/development/pro_sars2.obo) and here: [https://proconsortium.org/download/development/pro_sars2.gpi](https://proconsortium.org/download/development/pro_sars2.gpi).

26 Descriptions of various ontological extensions of IDO are included in Appendix C, along with disease ontologies employing many IDO terms, and planned extensions of IDO.
• **viral disease course** \(=_{\text{def}}\) An infectious disease course that realizes a viral disease

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

• **subclinical viral infection** \(=_{\text{def}}\) A subclinical infection involving virus infectious agents

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 3 provides a summary of several important links:

**Figure 3**

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* is a subclass of IDO Virus: *subclinical viral infection*
• IDO-CovID-19: *subclinical SARS-CoV-2 infection* is a subclass of CIDO: *subclinical coronavirus*

While distinguishing these instances from symptomatic cases:

• CIDO: *coronavirus disorder* is a subclass of IDO Virus: *viral disorder*
• IDO-CovID-19: *SARS-CoV-2 disorder* is a 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 is a subclass of IDO Virus:viral disease
- IDO-CoVID-19:COVID-19 is a subclass of CIDO:coronavirus disease
- CIDO:coronavirus disease course is a subclass of IDO Virus:viral disease course
- IDO-CoVID-19:COVID-19 disease course is a 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 provides terms for:

- IDO Virus:viral disease course has part IDO Virus:viral replication process
- IDO Virus:viral disease course has part IDO Virus:virus synthesis process

Allowing the following connections to CIDO:

- CIDO:coronavirus disease course has part CIDO:coronavirus synthesis process
- CIDO:coronavirus synthesis process is subclass of IDO Virus:virus synthesis process
- CIDO:coronavirus disease course has part CIDO:coronavirus replication process
- CIDO:coronavirus replication process is subclass of IDO Virus:viral replication process

And then IDO-CoVID-19:

- IDO-CoVID-19:SARS-CoV-2 is a subclass of CIDO:coronavirus
- IDO-CoVID-19:COVID-19 disease course has part IDO-CoVID-19:COVID-19 synthesis process
- CIDO:coronavirus synthesis process is a subclass of CIDO:coronavirus synthesis process
- IDO-CoVID-19:COVID-19 disease course has part IDO-CoVID-19:COVID-19 replication process
- CIDO:coronavirus replication process is a 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 $=$ def An antiviral bearing the viricidal disposition
- viricidal disposition $=$ def A disposition to kill viruses
- virostatic $=$ def An antiviral bearing the virostatic disposition
- virostatic disposition $=$ def A disposition to inhibit the lifecycle of viruses

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
• viral reproduction = def A 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 IDO: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 is a subclass of IDO Virus:negative regulation of viral replication
• CIDO:negative regulation of coronavirus replication IDO:negatively_regulates SOME CIDO:coronavirus replication process
• CIDO:negative regulation of coronavirus synthesis IDO:negatively_regulates SOME CIDO:coronavirus synthesis process

And in the case of IDO-CoVd-19:

• IDO-CovID-19:negative regulation of COVID-19 replication is a subclass of CIDO:negative regulation of coronavirus replication
• IDO-CovID-19:negative regulation of COVID-19 synthesis is a subclass of CIDO:negative regulation of coronavirus synthesis
• IDO-CovID-19:negative regulation of coronavirus synthesis IDO:negatively_regulates SOME IDO-CovID-19: COVID-19 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 An infectious disease epidemic largely involving viral infectious agents
• **viral disease pandemic** \(=_{\text{def}}\) An infectious disease pandemic consisting of viral epidemics

Themselves easily extended to CIDO:

• **coronavirus epidemic** \(=_{\text{def}}\) An infectious disease epidemic largely involving viral infectious agents
• **coronavirus pandemic** \(=_{\text{def}}\) An infectious disease pandemic consisting of viral epidemics

And from CIDO to IDO-CovID-19:

• **COVID-19 epidemic** \(=_{\text{def}}\) An infectious disease epidemic largely involving viral infectious agents
• **COVID-19 pandemic** \(=_{\text{def}}\) An infectious disease pandemic consisting of viral epidemics

Refining and expanding of 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:

• FLU:*influenza A virus infection* is a subclass of IDO Virus:*virus infection*
• HIV:*HIV virus infection* is a subclass of IDO Virus:*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.\(^{27}\)
• 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.\(^{28}\)
• 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 (At present this is little more than a list of datasets rather than a bona fide ontology.)\(^{29}\)

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

### 6. IDO Applications

\(^{27}\) https://bioportal.bioontology.org/ontologies/COVIDCRFRAPID
\(^{28}\) https://bioportal.bioontology.org/ontologies/COVID19
\(^{29}\) https://github.com/Research-Squirrel-Engineers/COVID-19
The following is an incomplete list of OBO Foundry related databases to which IDO ontology annotations can be, and in some cases are, applied:

- UniProt (https://www.uniprot.org/) representing millions of gene products annotated using terms from the Gene Ontology (GO) [3]
- The Eukaryotic Pathogen Databases (https://eupathdb.org/eupathdb/); genomic and other data for eukaryotic pathogens including Cryptosporidium, Giardia, Plasmodium, Theileria, Toxoplasma, and Trichomonas strains; uses EUPATH. The IDO Core terms human pathogenicity disposition, infection, infection prevalence, pathogen role, and primary infection are used in the annotation of its datasets.
- VectorBase (http://vectorbase.org); genomic and other data for a variety of invertebrate vectors of human pathogens; several terms from the Malaria Ontology (IDOMAL) are used in the annotation of its datasets [25, 51]. VectorBase includes a section on insecticide resistance section based on the Mosquito Insecticide Resistance Ontology (MIRO) [52]
- The Integrated Microbial Genomes System (http://img.jgi.doe.gov/)
- Microbes Online database (http://www.microbesonline.org/)
- Pathogen-Host Interaction Data Integration and Analysis System (http://phidias.us),
- Vicors virulence factors database (http://www.phidias.us/vicors) [53]
- BioHealthBase (http://www.biohealthbase.org/),
- National Microbial Pathogen Data Resource (http://www.nmpdr.org/).

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 [54-58]. A brief overview and further references are provided in [59]. 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 [60-62]. To facilitate interoperability and widespread circulation of future ADSSs, a Bacterial Clinical Infectious Disease Ontology (BCIDO) has been developed from IDO Core [63].
- In the case of methicillin-resistant *Staphylococcus aureus* MRSA, the *Staphylococcus aureus* Infectious Disease Ontology (IDOSA) has the potential to play a similar role by providing a classification of Sa that allows automated inference of resistance profiles [15, 16].
• IDDAP, a recently developed ontology-driven clinical decision support system for infectious disease diagnosis and antibiotic prescription makes heavy use of IDO Core [64].

• The Dengue Decision Support System (DDSS) is an ontology-driven computational application developed at Colorado State University [65, 66] 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 [67]. The DDSS makes use of IDOMAL [25], as well as the related ontologies MIRO [52] and the Vector Surveillance and Management Ontology (VSMO) [47].

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 and the Bioinformatics Resource for Invertebrate Vectors of Human Pathogens. These resources will be joined into one bioinformatics resource, the Eukaryotic Pathogen, Host & Vector Genomics Resource (VEuPathDB). The project will involve 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 [68]. IDO Core will play a role in this project, as the VEuPathDB ontology imports the following from IDO Core terms such as: human pathogenicity disposition, infection, infection prevalence, pathogen role, and primary infection.

7 CONCLUDING REMARKS

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. IDO ontologies such as IDO Virus, CIDO and IDO-CovID-19 fill this need by providing a standardized, computer-interpretable representation of heterogenous coronavirus knowledge to facilitate machine learning [88]. 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

30 https://www.cs.colostate.edu/ddss/index.html
31 https://eurekalert.org/pub_releases/2019-10/uop-nat101519.php
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 make a contribution also to addressing a further urgent problem faced not merely by CoVID 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 (OBI) [14], IDO offers a strategy to obtain comparable and integratable provenance metadata for the data generated in infectious disease research.

**APPENDIX A**

| Agronomy Ontology (Agro) [69] | Apollo Structured Vocabulary (Apollo-SV) [48] |
|-------------------------------|-----------------------------------------------|
| Beta Cell Genomics Ontology (BCGO) [70] | Cardiovascular Disease Ontology (CVDO) [71] |
| Coronavirus Infectious Disease Ontology (CIDO) [20, 21] | Human Disease Ontology (DOID) [44,45] |
| Drug Ontology (DrugON) [50] | VEuPathDB ontology (EUPATH) [68] |
| Influenza Ontology (FLU) [22] | Genomic Epidemiology Ontology (GenEpio) [49] |
| Hypertension Ontology (HTN) | Infectious Disease Ontology (IDO) [6] |
| Brucellosis Ontology (IDOBRU) [8, 23] | Meningitis Ontology (IDOMEN) [27] |
| The Plant Disease Ontology (IDOPlant) [28] | *Staph aureus* Infectious Disease Ontology (IDOSA) [7, 15, 16] |
| Schistosomiasis Ontology (IDOSCHISTO) [29, 30] | Mental Functioning Ontology (MFO) [72] |
| Mental Diseases Ontology (MFOMD) [73] | Monarch Disease Ontology (MonDO) [74] |
| Neurological Disease Ontology (ND) [75] | Ontology of Adverse Events (OAE) [76] |
| Ontology for Biomedical Investigations (OBI) [14] | OBI-NIAID-GSC-BRC-view |
| Ontology for Biobanking (OBIB) [77] | Ocular Disease Ontology (ODO) [78] |
| Oral Health and Disease Ontology (OHD) [79] | Ontology of Host Pathogen Interactions (OHPI) [64] |
| Ontology of Laboratory Animal Medicine (OLAM) | Ontologized MIABIS (OMIABIS) [80] |
| Ontology of Medically Relevant Social Entities (OMRSE) [81] | Obstetric and Neonatal Ontology (ONTONEO) [82] |
| Ontology of Precision Medicine and Investigation (OPMI) | Prescription of Drugs Ontology (PDRO) [83] |
APPENDIX B

Table 2 offers a brief overview of the extension ontologies that have been developed from the IDO Core. Some of these ontologies require reengineering – See the Supplementary Materials for details.

| Ontology Name | Description |
|---------------|-------------|
| Malaria Ontology (IDOMAL) [25, 26] | Created by the Christos Louis team at the Institute of Molecular Biology and Biochemistry (IMBB) in Crete, Greece, IDOMAL covers both clinical and epidemiological aspects of malaria, disease and vector biology, as well as intervention attempts to control the disease. IDOMAL was developed in the context of VectorBase, wherein several terms are used to annotate its datasets. IDOMAL was recently inherited by Christian Stoeckert and his team at the University of Pennsylvania where efforts are underway to reengineer the ontology’s content to bring about closer conformity to IDO Core. |
| Dengue Ontology (IDODEN) [24] | Also developed by the Louis research team at IMBB, IDODEN covers all aspects of dengue fever including disease biology, epidemiology, clinical features, and vector entomology. IDODEN was developed in the context of VectorBase and its structure was designed to mirror that of IDOMAL. Intended for use in Dengue decision support systems. |
| Brucellosis Ontology (IDOBRU) [8, 23] | Developed by Yongqun He and his team at the University of Michigan, IDOBRU focuses on Brucellosis, a highly contagious zoonotic infectious disease caused by the intracellular, Gram-negative bacteria Brucella. IDOBRU encompasses the domains of clinical care, public health, and biomedical research, along seven major axes: host infection and zoonotic disease transmission, symptoms, virulence factors and pathogenesis, diagnosis, intentional release, vaccine prevention, and treatment. |
| Influenza Ontology (FLU) [22] | Developed by Richard Scheuermann, Lynn Schriml, Joanne Luciano, and Burke Squires, FLU covers the natural, experimental and clinical realms related to influenza virus life cycle, infection and disease. FLU utilizes OBI classes for components of materials, qualities and processes to map influenza virus sequence and surveillance terms to their corresponding materials and qualities. FLU is applied to data collected by the Centers for Excellence in Influenza Research and Surveillance (CEIRS) project to help researchers more easily elucidate influenza virulence and pathogenesis etiology. |
| *Staphylococcus aureus* Infectious Disease Ontology (IDOSA) [7, 15, 16] | Developed by Lindsay Cowell, Barry Smith and Albert Goldfain, in collaboration with Dr. Vance Fowler at Duke University Medical Center, IDOSA focuses on Staph aureus (Sa) infection diseases. The ontology is used to analyze networks of functionally related gene products to identify host genes conferring susceptibility to *Staphylococcus aureus* bacteremia. |
HIV Ontology\(^{32}\) (HIV) \nDeveloped by Martin Schiller of UNLV, the HIV ontology is intended to cover all types of HIV data and information; intended for use in the HIVToolbox web application \([85, 86]\).

Schistosomiasis Ontology (IDOSCHISTO) \([29, 30]\) \nDeveloped by team of researchers based in Senegal led by Dr. Gaoussou Camara, IDOSCHISTO focuses on schistosomiasis, a waterborne infectious disease caused by \textit{Schistosoma} helminth parasites. The ontology is organized into 3 main sub-modules: i) Biology, e.g. pathogen/host interactions, host physiological reactions to the disease, pathogen taxonomy and life-cycle; ii) Epidemiology, for example risk factors, spread of disease, means to prevent and control; and iii) Clinical, including symptoms that influence differential diagnoses and treatment decisions.

Plant Disease Ontology (IDOPlant) \([28]\) \nUnder active development within the context of the Planteome Project, IDOPlant provides a comprehensive reference ontology for any infectious plant disease. The main aims of IDOPlant are to “provide plant scientists with the means identify genomic and genetic signatures host-pathogen interactions, resistance, or susceptibility, and to help agronomists and farmers by developing tools to identify disease phenotypes and gather epidemiological statistics” \([28]\).

Meningitis Ontology (IDOMEN) \([27]\) \nAlso developed by Dr. Camara’s team, IDOMEN focuses on meningitis, a disease caused by the gram-negative bacteria, \textit{neisseria meningitidis}. IDOMEN covers the meningitis domain along three main axes: i) biological (immunity, virulence factors, pathogen and host biology); ii) clinical (clinical manifestations, laboratory tests and findings, diagnosis and treatment); epidemiological (surveillance, prevention, epidemic emergence factors such as risk behaviors, climate and environment).

Coronavirus Infectious Disease Ontology (CIDO) \([20, 21]\) \nInitiated by Yongqun He and Hong Yu, CIDO was developed by a collaborative group of researchers in both the US and in China in response to the recent COVID-19 outbreaks. CIDO provides standardized human- and computer-interpretable annotation and representation of various coronavirus infectious diseases, including their etiology, transmission, epidemiology, pathogenesis, diagnosis, prevention, and treatment.

| Table 2: IDO Core extension ontologies |
|---|
| In addition IDO extension ontologies for Tuberculosis (IDOTUB) and Infective endocarditis are planned. Preliminary work on the latter is available online.\(^{33}\) Terms drawn from IDO Core are also used in a number of other important ontologies within the infectious disease domain, as illustrated in Table 3 below: |

Vaccine Ontology (VO) \([17]\) \nDeveloped in collaboration with the IDO Core team, VO has strong ties to the IDO initiative. Maintained by Yongqun He’s team at the University of Michigan, VO focuses on the classification of vaccines and vaccine components, vaccine quality and phenotypes, and host immune response to vaccines. VO is used in the Vaccine Investigation and Online Information System (VIONS) and the Vaccine Research Database (VREDB).

---

\(^{32}\) https://bioportal.bioontology.org/ontologies/HIV/?p=summary

\(^{33}\) https://sivsmusings.wordpress.com/
Network (VIOLIN, http://www.violinet.org/) a central repository for literature related to, and data resulting from, vaccine research.

Bacterial Clinical Infectious Disease Ontology (BCIDO) [63]

BCIDO provides a controlled terminology for clinical bacterial infectious diseases along with domain knowledge commonly used in the hospital in-patient setting. BCIDO was designed to augment the use of ADSSs, thus serving as a tool to guide differential bacterial diagnoses and to assist in the prescribing of appropriate antimicrobial treatments. As such, BCIDO “encompasses terms and knowledge about common clinical presentations of bacterial infections, patient specific factors that influence differential diagnoses and treatment options, the [bacteria] themselves, and the antimicrobial agents used to treat infections” [63]. Though not an IDO extension per se, BCIDO imports IDO Core in full.

Vector Surveillance and Management Ontology (VSMO) [47]

Developed by a team of researchers at Colorado State University led by Drs. Lars Eisen and Saul Lozano, in collaboration with Dr. Cowell, VSMO covers the domain of surveillance and management of vectors and vector-borne pathogens, with special emphasis on content to support operational activities through inclusion in databases, data management systems and decision support systems [47]. The ontology includes terms for i) arthropod species capable of being biological vectors and for pathogen species transmitted by arthropod vectors; ii) chemical compounds relevant to insecticide resistance and chemical pesticide active substances (originating from MIRO); iii) terms for equipment used to collect or control vectors or vertebrate pathogen hosts, and tools used to kill vectors or prevent contact with humans. In consultation with the IDO Core developers, the relation has_vector was created for VSMO in order to link pathogens to their biological vectors.

Genomic Epidemiology Ontology (GenEpio) [49]

GenEpio is a controlled vocabulary for infectious disease surveillance and outbreak investigations implementing whole genome sequencing (WGS) of microbial pathogens. GenEpio aims to enable the integration and promotion of all contextual information required to interpret pathogen genomics data, including critical knowledge about sequencing pipelines and sequence quality; lab results describing antimicrobial resistance and virulence phenotypes; epidemiological data concerning potential sources of risk and exposure; as well as data about susceptible populations and geographical distributions of pathogen strain [49]. GenEpio is currently being integrated into IRIDA (Integrated Rapid Infectious Disease Analysis), a user-friendly, decentralized, open-source bioinformatics and analytical web platform to support real-time infectious disease outbreak investigations using WGS data [87].

Table 2: Some other ontologies that make use of IDO Core

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 Chris Stoeckert and Yongqun (Oliver) He. Barry Smith’s

---

34 The most recent version of the ontology, dated March 4, 2013, can be found at: https://code.google.com/archive/p/vector-surveillance-and-management-ontology/downloads
contributions were supported by the NIH under NCATS 1UL1TR001412 (Buffalo Clinical and Translational Research Center).

Authors’ contributions

All authors read and extensively reviewed the manuscript. SB wrote the manuscript and conducted the research. JB’s contributions vastly improved the structured of the paper, as well as the sections on CIDO and VIDO. BS and LC 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. LC was a significant contributor to the development of BCIDO and VSMO.

REFERENCES

[1] J.D. Ferreira, D. Paolotti, F. M. Couto, M. J. Silva, (2013). “On the usefulness of ontologies in epidemiology research and practice”, Journal of Epidemiology and Community Health, 67(5): 385-388
[2] C. Pesquita, J. D. Ferreira, F. M. Couto, M. J. Silva, (2014). “The epidemiology ontology: an ontology for the semantic annotation of epidemiological resources”, The Journal of Biomedical Semantics, 5(4): 1-7
[3] The Gene Ontology Consortium, (2019) “The Gene Ontology Resource: 20 years and still GOing strong”, Nucleic Acids Research, 47 (D1): D330–D338.
[4] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall, OBI Consortium, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.-A. Sansone, R. H. Scheuermann, N. Shah, P. L. Whetzel, and S. Lewis, 2007. “The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration,” Nat. Biotechnol., 25(11): 1251–1255.
[5] B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A. L. Rector, and C. Rosse, 2005. “Relations in biomedical ontologies,” Genome Biol., 6(5) p. R46,
[6] Cowell L, Smith B. (2010). Infectious Diseases Ontology. In: Sintchenko V, ed. Infectious Disease Informatics. New York: Springer: 373-95.
[7] Goldfain A, Smith B, Cowell LG, (2010). “Dispositions and the Infectious Disease Ontology”, In: Proceedings of FOIS2010.
[8] Lin, Y., Xiang, Z. & He, Y., 2011. Brucellosis Ontology (IDOBRU) as an extension of the Infectious Disease Ontology. Journal of Biomedical Semantics, 2(1), p.9.
[9] Scheuermann, R.H., Ceusters, W. & Smith, B. (2009). Toward an Ontological Treatment of Disease and Diagnosis. Summit on Translational Bioinformatics, 2009, p.116.
[10] R. Arp, B. Smith and A. Spear, (2015). Building Ontologies with Basic Formal Ontology, Cambridge, MA: MIT Press.
[11] Ceusters W, Smith B, Kumar A, Dhaen C (2004a). “Mistakes in medical ontologies: where do they come from and how can they be detected? In: Pisanelli D (ed) Ontologies in medicine. IOS, Amsterdam, pp 145–164
[12] Ceusters W, Smith B, Kumar A, Dhaen C, (2004b). “Ontology-Based Error Detection in SNOMED-CT”, in M. Fieschi, et al. (eds.), Medinfo 2004, Amsterdam: IOS Press, 482–486.
[13] Bodenreider O, Smith B, Kumar A, Burgun A, (2007). “Investigating Subsumption in SNOMED CT: An Exploration into Large Description Logic-Based Biomedical Terminologies”, Artificial Intelligence in Medicine, 39: 183–195, PMC2442845
[14] Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, 748 Clancy K, Courtot M, Derom D, Dumontier M, et al. The ontology for biomedical investigations. PLoS One. 2016;11(4):e0154556.
[15] Goldfain A, Smith B, Cowell L.G. (2011). Toward an ontological representation of resistance: the case of MRSA. Journal of Biomedical Informatics, 44:35-41.
[16] Goldfain A, Smith B, Cowell L.G. (2012). Constructing a Lattice of Infectious Disease Ontologies from a Staphylococcus aureus Isolate Repository. In: Proceedings of the 3rd International Conference on Biomedical Ontology (ICBO 2012), KR-MED Series; 2012: 1-5.
[17] He Y, Cowell LG, Diehl AD, et al. (2009). VO: Vaccine Ontology The 1st International Conference on Biomedical Ontology (ICBO 2009); Buffalo, NY, USA. Nature Precedings.
[18] Haendel MA, Balhoff JP, Bastian FB, et al. (2014). Unification of multi-species vertebrate anatomy ontologies for comparative biology in Uberon. J Biomed Semantics. 5(21) doi:10.1186/2041-1480-5-21
[19] Degtyarenko et al. 2007. ChEBI: a database and ontology for chemical entities of biological interest. Nuc Acids Research 36, pp. D344-D350.
[20] Liu Y, Chan W, Wang Z, Hur J, Xie J, Yu H, He Y. Ontological and bioinformatic analysis of anti-coronavirus drugs and their Implication for drug repurposing against COVID-19. Preprints. 2020, 2020030413 (doi: https://doi.org/10.20944/preprints202003.0413.v1).
[21] Ong E, Wong M, Huffman A, He Y. COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. bioRxiv. Posted on March 23, 2020. doi: https://doi.org/10.1101/2020.03.20.000141
[22] Luciano J, Hirschman L, Colosimo M, (2008). Ontology Support for Influenza Research and Surveillance [Online]. Available: https://www.researchgate.net/publication/265222266_Ontology_Support_for_Influenza_Research_and_Surveillance
[23] Lin, Y., Xiang, Z., & He, Y. (2015). Ontology-based representation and analysis of host-Brucella interactions. Journal of biomedical semantics, 6, 37.
[24] E. Mitraka, P. Topalis, V. Dritsou, E. Dialynas, and C. Louis, 2015. “Describing the Breakbone Fever: IDODEN, an Ontology for Dengue Fever,” PLoS Negl. Trop. Dis., vol. 9, no. 2.
[25] Topalis P, Mitraka E, Bujila I, et al. (2010). IDOMAL: an ontology for malaria. Malaria Journal, 9:230.
[26] Topalis P, et. al, 2013. IDOMAL: the malaria ontology revisited. Journal of Biomedical Semantics, 4(16):
[27] Béré, C., Camara, G., Malo, S., & Lo M., Ouaro, S. IDOMEN: An Extension of Infectious Disease Ontology for MENingitis. MEDINFO 2019: Health and Wellbeing e-Networks for All L. Ohno-Machado and B. Séroussi (Eds.)
[28] Walls, R.L., Smith, B., Elser, J., Goldfain, A., Stevenson, D.W., and Jaiswal, P. 2012. “A plant disease extension of the infectious disease ontology,” in Proceedings of the 3rd International Conference on Biomedical Ontology, Medical University of Graz, Graz.
[29] G. Camara, S. Després, R. Djedidi, and M. Lo. (2013). “Building a Schistosomiasis Process Ontology for an Epidemiological Monitoring System,” in Recent Advances in Knowledge Engineering: Paradigms and Applications, C. Faucher and L. Jain, Eds. Springer-Verlag.
[30] G. Camara, S. Després, M., Lo. (2014). IDOSCHISTO: une extension de l’ontologie noyau des maladies infectieuses (IDO-Core) pour la schistosomiasis. In: Faron-Zucker, C., ed. IC - 25èmes Journées francophones d’Ingénierie des Connaissances, Clermont-Ferrand, France. Session 1: Construction, peuplement et exploitation d’ontologies: 39-50.
[31] Xiang Z, Zheng W, He Y, 2006. BBP: Brucella genome annotation with literature mining and curation. BMC Bioinformatics. 7(4):347.
Cooper, L, Meier, A, Laporte, MA, et al. (2018). The Planteome database: an integrated resource for reference ontologies, plant genomics and phenomics. *Nucleic acids research*, 46(D1), D1168–D1180.

Mungall, C. et al., 2010. Integrating phenotype ontologies across multiple species. Genome Biology, 11(1), p.R2.

Walls, RL, Athreya, B, Cooper, L, et al. (2012a). Ontologies as integrative tools for plant science. *American Journal of Botany*. 99, 1263–1275.

Cooper, L., Elser, J., Gandolfo, M. A., Stevenson, D.W., Smith, B., et al. (2013). The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant Cell Physiol*. 54:

P. L. Buttigieg, N. Morrison, B. Smith, C. J. Mungall, S. E. Lewis, (2013). “The environment ontology: contextualising biological and biomedical entities”, Journal of Biomedical Semantics, 4(1), 43: 1-9.

P. L. Buttigieg, E. Pafilis, S. E. Lewis, M. P. Schildhauer, R. L. Walls, C. J. Mungall, (2016). “The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperability”, Journal of Biomedical Semantics, 7(1), 57: 1-12.

Walls R. L., Cooper L, Elser J, Gandolfo M, Mungall C J., Smith B, Stevenson D, Jaiswal P. 2019, The Plant Ontology Facilitates Comparisons of Plant Development Stages Across Species. Frontiers in Plant Science. 10: 631

J. Ginsberg, M. Mohebbi, R. Patel, L. Brammer, M. Smolinski, and L. Brilliant (2009). “Detecting Influenza Epidemics using Search Engine Query Data”, *Nature*, 457 (7232): 2012-2014.

A. Signorini, A. M. Segre, and P. M. Polgreen, “The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic”, *PLoS ONE*, vol. 6, no. 5, p. e19467, 05 2011.

N. Generous, G. Fairchild, A. Deshpande, S. Del Valle, R. Priedhorsky. (2014). “Global disease monitoring and forecasting with Wikipedia”, *PLoS Computational Biology*, 10(11), e1003892.

N. Collier, S. Doan (2012). “Enhancing Twitter Data Analysis with Simple Semantic Filtering: Example in Tracking Influenza-Like Illness”, *Proc. Health Care Informatics, Imaging and Systems Biology*: 62-71.

M. Magumba, P. Nabende (2018). “Ontology boosted deep learning for disease name extraction from Twitter messages”, *Journal of Big Data*, 5(31)

Schrinl M. M., Arze C., Nandella S., et al. (2012), Disease Ontology: a backbone for disease semantic integration, *Nucleic Acids Research*, Volume 40, Issue D1, 1 Pages D940–D946, https://doi.org/10.1093/nar/gkr972

Kibbe W. A., Arze C., Felix V., et al., (2015), Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data, *Nucleic Acids Research*, Volume 43, Issue D1, Pages D1071–D1078, https://doi.org/10.1093/nar/gku1011

Stieglbauer, G; Coulibaly, B; Wickert, H; McLean, J; Kouyaté, BA; et al. (2006). "Malaria in a holoendemic area of Burkina Faso: a cross-sectional study". *Parasitology Research*. 98 (6): 596–599. doi:10.1007/s00436-005-0104-9. PMID 16416123

Lozano-Fuentes S, Bandyopadhyay A, Cowell LG, et al. Ontology for vector surveillance and management. Journal of Medical Entomology, 2013;50:1-14.

Hogan, W. R., Wagner, M. M., Brochhausen, M., Levander, J., Brown, S. T., Millett, N., ... Hanna, J. (2016). The Apollo Structured Vocabulary: an OWL2 ontology of phenomena in infectious disease epidemiology and population biology for use in epidemic simulation. *Journal of biomedical semantics*, 7, 50. doi:10.1186/s13326-016-0092-y

E. Griffiths, D. Dooley, M. Graham, G. Van Domselaar, F. S.L. Brinkman, W. W.L. Hsiao, (2017). “Context Is Everything: Harmonization of Critical Food Microbiology Descriptors and Metadata for Improved Food Safety and Surveillance”, Frontiers in Microbiology, 8(1068): 1-11.
[50] Hogan WR, Hanna J, Joseph E, et al. Towards a Consistent and Scientifically Accurate Drug Ontology. In: ICBO 2013 Conference Proceedings; 2013; 2013.
[51] Lawson D, Arensburger P, Atkinson P, et al. (2009) VectorBase: a data resource for invertebrate vector genomics. Nucl Acids Res, 37: D583-587.
[52] Dialynas E, Topalis P, Vontas J, Louis C (2009). MIRO and IRbase: IT Tools for the Epidemiological Monitoring of Insecticide Resistance in Mosquito Disease Vectors. PLoS Negl Trop Dis, 3:e465.
[53] Sayers S, Li L, Ong E, et al. (2019). Victors: a web-based knowledge base of virulence factors in human and animal pathogens. Nucleic Acid Research. 2019 Jan 8;47(D1):D693-D700. doi: 10.1093/nar/gky999. PMID: 30365026. PMCID: PMC6324020.
[54] Zhang Y-F, Gou L, Zhou T-S, Lin D-N, Zheng J, et al. 2017. An ontology-based approach to patient follow-up assessment for continuous and personalized chronic disease management. J. Biomed. Inform. 72:45–59
[55] Carletti G, Giuliodori P, DiPietri V, Peretti A, Amenta F. 2016. An ontology-based consultation system to support medical care on board seagoing vessels. Int. Marit. Health 67(1):14–20
[56] Maurice P, Dhombres F, Blondiaux E, Friszer S, Guibaud L, et al. 2017. Towards ontology-based decision support systems for complex ultrasound diagnosis in obstetrics and gynecology. J. Gynecol. Obstet. Hum. Reprod. 46(5):423–29
[57] Abidi S. 2017. A knowledge-modeling approach to integrate multiple clinical practice guidelines to provide evidence-based clinical decision support for managing comorbid conditions. J. Med. Syst. 41(12):193
[58] Lin Y, Staes CJ, Shields DE, Kandula V, Welch BM, Kawamoto K. 2015. Design, development, and initial evaluation of a terminology for clinical decision support and electronic clinical quality measurement. AMIA Annu. Symp. Proc. 2015:843–51.
[59] Melissa A. Haendel, Julie A. McMurry, Rose Relevo, Christopher J. Mungall, Peter N. Robinson, Christopher G. Chute. (2018). A Census of Disease Ontologies. Annual Review of Biomedical Data Science. 1(1): 305-331
[60] Evans RS, Pestotnik SL, Classen DC, et al. A computer-assisted management program for antibiotics and other antimicrobial agents. New England Journal of Medicine 1998;338:232-8.
[61] Thursky KA, Mahemoff M. User-centered design techniques for a computerised antibiotic decision support system in an intensive care unit. Int J Med Inform 2007; 76:760-8.
[62] Paterson DL. The role of antimicrobial management programs in optimizing antibiotic prescribing within hospitals. Clin Infect Dis 2006;42 Suppl 2: S90-5.
[63] C. L. Gordon, S. Pouch, L. G. Cowell, M. R. Boland, H. L. Platt, A. Goldfain, and C. Weng, “Design and evaluation of a bacterial clinical infectious diseases ontology,” AMIA Annu. Symp. Proc. AMIA Symp. AMIA Symp., vol. 2013, pp. 502– 511, 2013.
[64] Shen Y, Yuan K, Chen D, et al. An ontology-driven clinical decision support system (IDDAP) for infectious disease diagnosis and antibiotic prescription. Artif Intell Med 2018; 86: 20–32
[65] 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. [PubMed: 21468310]
[66] Lozano–Fuentes, S.; Barker, CM.; Coleman, M.; 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. InTech-Open Access Publisher; Rijeka, Croatia: 2011. p. 89-114.
[67] Lozano–Fuentes S, Wedyan F, Hernandez-Garcia E, Sadhu D, Ghosh S, Bieman JM, Tep-Chel D, García-Rejón JE, Eisen L. Cell phone-based system (Chaak) for surveillance of immatures of
dengue virus mosquito vectors. *J Med Entomol*. 2013 Jul;50(4):879-89. doi: 10.1603/me13008. PMID: 23926788; PMCID: PMC3929104.

[68] Zheng, J., Cade, J. S., Brunk, B., Roos, D. S., Stoeckert, C. J., et al (2016). Malaria study data integration and information retrieval based on OBO Foundry ontologies. CEUR Workshop Proceedings, 1747.

[69] Devare, M., Aubert, C., Laporte, M.-A., Valette, L., Arnaud, E., Buttigieg, P.L., 2016. Data-driven agricultural research for development - a need for data harmonization via semantics. In: Jaiswal, P., Hoehndorf, R. (Eds.), 7th International Conference on Biomedical Ontologies, ICBO’16, vol. 1747 of CEUR Workshop Proceedings, Corvallis, Oregon, USA, pp. 2, August.

[70] Zheng J, Manduchi E, and Stoeckert C. Development of an Application Ontology for Beta Cell Genomics Based on the Ontology for Biomedical Investigations. International Conference on Biomedical Ontologies (ICBO 2013), Montreal, Canada, July 2013.

[71] Barton A, Rosier A, Burgun A, Ethier J-F. 2014. The Cardiovascular disease ontology. In Formal Ontology in Information Systems, ed. P Garbacz, O Kutz, pp. 409-414. Front. Artif. Intell. Appl. 267. Amsterdam: IOS.

[72] Hastings J, Ceusters W, Jensen M, Mulligan K, and Smith B., (2012). Representing mental functioning: Ontologies for mental health and disease. Towards an Ontology of Mental Functioning, Third International Conference on Biomedical Ontology, Graz, July 22, 2012.

[73] Ceusters W. and Smith B. 2010. Foundations for a realist ontology of mental disease. Journal of Biomedical Semantics. 1(10).

[74] Mungall CJ, McMurry JA, Köhler S, Balhoff JP, Borrormeo C, et al. (2017). The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Res. 45(D1):D712–D722

[75] Jensen M, Cox AP, Chaudhry N, Ng M, Sule D, Duncan W, Ray P, Weinstock-Gutman B, Smith B, Ruttenberg A, Szigeti K, Diehl AD. (2013). The neurological disease ontology. *J Biomed Semantics*. Dec 6;4(1):42. doi: 10.1186/2041-1480-4-42. PMID: 24314207; PMCID: PMC4028878

[76] He Y, Sarntivijai S, Lin Y, et al. (2014). OAE: The Ontology of Adverse Events. Journal of Biomedical Semantics. 5:29 doi:10.1186/2041-1480-5-29. PMID: 25093068. PMCID: PMC4120740.

[77] Mathias Brochhausen, Jie Zheng, David Birtwell, Heather Williams, Anna Maria Maseci, Helena Judge Ellis and Christian J. Stoeckert Jr. (2016). OBIB-a novel ontology for biobanking” Journal of Biomedical Semantics. 7:23

[78] Ray, P. and Diehl, A. D. (2013). The Ocular Disease Ontology., in ICBO, p. 119

[79] Schleyer TK, Ruttenberg A, Duncan W, Haendel M, Torniai C, et al. (2013). An ontology-based method for secondary use of electronic dental record data. AMIA Jt. Summits Transl. Sci. Proc. 2013:234–238

[80] Brochhausen, M., Fransson, M.N., Kanaskar, N.V. et al. Developing a semantically rich ontology for the biobank-administration domain. *J Biomed Semant* 4, 23 (2013). https://doi.org/10.1186/2041-1480-4-23

[81] Hicks, A., Hanna, J., Welch, D. *et al*. The ontology of medically related social entities: recent developments. *J Biomed Semant* 7, 47 (2016). https://doi.org/10.1186/s13326-016-0087-8

[82] F. Farinelli, M.B. Almeida, P. L. Elkin, and B. Smith, (2016). OntONEo: The Obstetric and Neonatal Ontology, in: 7th International Conference on Biological Ontology (ICBO 2016), Oregon State University, Corvallis, OR.

[83] Ethier, J.F., Barton, A., and Taseen R. (2018) An ontological analysis of drug prescriptions, Applied Ontology, vol. 13, no. 4, pp. 273-294.

[84] Goldfain, A. Smith, B., Arabandi, S., Brochhausen, M., 2011. Hogan, W. Vital Sign Ontology. Proceedings of the Workshop on Bio-Ontologies, ISMB, Vienna, June 2011, 71-74.
[85] Sargeant, D., Deverasetty, S., Luo, Y., Villahoz Baleta, A., Zobrist, S., Rathnayake, V., … Schiller, M. R. (2011). HIVToolbox, an integrated web application for investigating HIV. *PloS one, 6*(5), e20122. doi:10.1371/journal.pone.0020122

[86] Sargeant DP, Deverasetty S, Strong CL, Alaniz IJ, Bartlett A, et al. (2014) The HIVToolbox 2 Web System Integrates Sequence, Structure, Function and Mutation Analysis. PLOS ONE 9(6): e98810. https://doi.org/10.1371/journal.pone.0098810

[87] Matthews TC, Bristow FR, Griffiths EJ, Petkau A, Adam J, Dooley D, Kruczkiewicz P, Curatcha J, Cabral J, Fornika D, Winsor GL, Courtot M, Bertelli C, Roudgar A, Feijao P, Mabon P, Enns E, Thiessen J, Keddy A, Isaac-Renton J, Gardy JL, Tang P, the IRIDA Consortium, Carrico JA, Chindelevitch L, Chauve C, Graham MR, McArthur AG, Taboada EN, Beiko RG, Brinkman FS, Hsiao WW, Van Domselaar G. (2018). The Integrated Rapid Infectious Disease Analysis (IRIDA) platform. bioRxiv. https://doi.org/10.1101/381830

[88] Yongqun He, Hong Yu, Edison Ong, et al. [unpublished: under review] CIDO: The community-based Coronavirus Infectious Disease Ontology for integrative knowledge and data representation, standardization, sharing, and analysis.