OntoEffect: an OntoUML-based ontology to explain SARS-CoV-2 variants’ effects

Ruba Al Khalaf\textsuperscript{1}, Anna Bernasconi\textsuperscript{1}, and Alberto García S.\textsuperscript{2}

\textsuperscript{1}Politecnico di Milano, Milan, Italy
\textsuperscript{2}Universitat Politècnica de València, Valencia, Spain
{ruba.al,anna.bernasconi}@polimi.it, algarsi3@inf.upv.es

Keywords: Ontological Analysis, OntoUML, SARS-CoV-2, Variants impacts.

Abstract: The SARS-CoV-2 virus continuously accumulates genetic variation through mutations; mutations are the virus’ way to achieve viral adaptation. Although the huge amount of information accumulated on the virus during the COVID-19 pandemic, the knowledge that contributes to explaining and supporting the research related to SARS-CoV-2 characteristics and evolution is not currently organized, nor systematized. Here, we present OntoEffect, an ontology that captures and represents such information systematically. Specifically, we aim to represent the dimensions of the virus and its mutations, discussing their impacts on the virus itself, as well as on public health, prevention, and treatment protocols. Aiming to obtain ontological clarity in such a complex domain, OntoEffect was built using OntoUML, an ontology-driven conceptual modeling language, grounded on the Unified Foundational Ontology (UFO). In the highly specialized context of virology, we show the powerful ability of ontological models to provide clear and precise explanations of a domain and allow its shared understanding among stakeholders.

1 INTRODUCTION

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was initially detected in China in December 2019, causing the coronavirus disease 2019 (COVID-19). Consequently, after the continuous spreading of the virus across the globe, the World Health Organization (WHO) declared a Public Health Emergency of International Concern (PHEIC) on 30 January 2020. Subsequently, on 11 March 2020, the WHO officially classified the outbreak as a pandemic.

After more than three years into the pandemic, specifically on 5 May 2023, the WHO announced that the disease no longer met the criteria for a PHEIC due to its established and ongoing nature. According to the same announcement, this decision does not imply that the pandemic itself has ended, but rather that the global emergency it caused has subsided, at least for the time being (World Health Organization, 2023).

The impact of the global COVID-19 pandemic has affected all human lives worldwide on multiple levels. Since its emergence, the SARS-CoV-2 virus has continued to evolve genetically with an expanding range of variants. As it happens in other RNA viruses, achieving genetic diversity is an essential aspect for the survival and continuation of SARS-CoV-2; indeed, it brings viral survival, fitness, and pathogenesis (Lauring et al., 2013; Mattenberger et al., 2021). The observed impacts are continuously present, and they vary depending on the specific variant that spreads at given times of the outbreak. For example, it was widely known that the viral transmission increased with the Alpha variant of SARS-CoV-2 virus (Tanaka et al., 2021). Therefore, collecting information about the impacts of genetic variations in SARS-CoV-2 in a systematic way, structuring it, and drawing the relations between these impacts is of great importance.

In this research, we aim to fill this gap; specifically, we draw the relationships between the virus and the impacts of its mutations and variants formation over the progression of the pandemic.

In the context of COVID-19, it has been recognized that many ontologies are being used to address the challenges related to the pandemic analytics – with different intents like structural representation and semantics of data, visualizing the status of infections, querying, and ontology editing and extraction (Ahmad et al., 2021).
Mapping knowledge into standardized models, such as ontologies, allows for a precise understanding of a particular domain (Romanenko et al., 2022; Guizzardi and Guarino, 2023), which can be used to support data representation, integration, sharing, and analysis.

Along this direction, we present OntoEffect to formally represent the set of relevant concepts in this domain. More specifically, we capture the concepts related to the impact of viral variation over the overall outcome of the virus behavior. OntoEffect also captures the relationships between these concepts in a structured format that can be used to integrate data and facilitate knowledge sharing among researchers.

As methodological support for this process, we employ OntoUML, a well-founded ontology-driven conceptual modeling language whose meta-model complies with the theoretically well-grounded Unified Foundational Ontology (UFO) (Guizzardi, 2005). OntoUML has previously been employed for ontological analysis of specialized life sciences domains (Guizzardi et al., 2021; García S et al., 2022b) and proven effective in improving the explanation of complex domains with respect to traditional conceptual models (García S et al., 2022a; Verdonck et al., 2019).

The scope of this research is broad, including concepts that encompass - but are not restricted to - epidemiology, immunology, and structural biology. For achieving the highest possible clarification, we follow the principles of the ontological unpacking method (Bernasconi et al., 2022), proposed to analyze and break down a complex concept or phenomenon into its constituent parts based on an ontology that defines the relationships between different types of entities, finally resulting in a more granular understanding of that complex system. In other words, the ontological unpacking method is a procedure in which Ontology – as a field of study – is utilized to apply the theory of ontological analysis, thereby uncovering the ontological conceptual model that depicts an information structure.

The remainder of the paper is structured as follows. Section 2 reviews related works, including well-known ontologies concerning the SARS-CoV-2 domain, and motivates our effort. Section 3 overviews the methodology used to conduct the ontological analysis, listing the most relevant OntoUML class and relationship stereotypes. Section 4 presents the OntoEffect ontology and its advantages. Finally, Section 5 briefly describes the validation and concludes the work.

## 2 RELATED WORK

Ontologies have gained attention during the pandemic of COVID-19, and several researchers have built ontologies to solve diverse problems they were facing.

Within the list of UFO and OntoUML ontology models presented in the OntoUML/UFO-catalog (Barcelos et al., 2022) – a catalog for ontology-driven conceptual modeling research (https://github.com/unibz-core/ontouml-models), sharing all the models that employ OntoUML for advancing ontology-driven research – a pair of works related to what we are presenting can be found:

- (Bernasconi et al., 2020) presented the Viral Conceptual Model (VCM), which is a conceptual model that focuses on viral genomic data structuring to characterize genomic sequences.
- The same model went through an ontological unpacking process (Guizzardi et al., 2021) achieved with the use of the OntoUML language; the obtained ontology represents the semantic areas of Virus Infection, Tissue Sampling, Virus Sequencing, and Virus Sequence Annotation. While the VCM only served as a base for building a genomic data structure, its ontological unpacking provided a detailed analysis of the complex notions hidden within the original VCM model.

A domain ontology for modeling lockdown interventions during the COVID-19 pandemic was presented in (Fabio et al., 2021). It aims to provide a comprehensive understanding of what a lockdown entails by proposing an ontology-based modeling approach. The authors highlight the need for a well-grounded domain ontology that incorporates real-world semantics to capture the complexities and variations of lockdown measures.

Other related ontologies were developed with a different upper-level ontology, the Basic Formal Ontology (BFO) (Arp et al., 2015). For instance, He et al. presented the Coronavirus Infectious Disease Ontology (CIDO, He et al., 2020) at the beginning of the COVID-19 pandemic. CIDO is a community-based open-source biomedical ontology that supports coronavirus disease knowledge, along with data standardization, integration, sharing, and analysis. In addition to being based on BFO, CIDO follows the Open Biomedical and Biomedical Ontologies (OBO) Foundry Principles (Smith et al., 2007).

To our knowledge and after studying the domain, we believe OntoEffect is unique in its scope and design approach as it systematically and comprehensively describes the impact of viral variations on different levels and aspects.
3 METHODOLOGY

The workflow for building OntoEffect is shown in Figure 1. To identify the effects of interest that are suited to be considered as entities in OntoEffect, we first built a taxonomy (i.e., controlled vocabulary) of effects. In this article, we propose the transformation of the taxonomy into a structured ontology.

We started from the beginning of the COVID-19 pandemic, following a knowledge acquisition protocol to search the literature from different sources considering both peer-reviewed articles and preprints, with more focus on the peer-reviewed ones. Such sources are Google Scholar, PubMed, GISAID/COG-UK reports, MedRxiv, and BioRxiv. For keywords search and to achieve relativeness to our field of interest, we used keywords composed of a virus-related term (e.g., SARS-CoV-2), a known clinical impact (e.g., infectivity), and a known mutation and/or variant (e.g., Alpha). The criteria for inclusion of an effect in the taxonomy of effect are its high representativeness in the literature and connection with at least one non-synonymous mutation (or variant) of SARS-CoV-2.

Along with the evolution of the virus, we continued curating the newest related publications in the field and adding new terms to the taxonomy. To structure the final version of the taxonomy, we classified the selected terms into four categories according to the field of impact. They include Epidemiology (clustering the impacts of a variant on epidemiological features); Immunology (impacts on the host immune system); Viral kinetics and dynamics (of viral proteins and the viral cell cycle); and Diagnosis, prevention, and treatments (impacts on these measures).

Note that different versions of the taxonomy were proposed in past works: a preliminary idea is in (Al Khalaf et al., 2021); consolidated in (Alfonsi et al., 2022) (employed within a providing a complete schema representing COVID-19 knowledge and data interplay); finalized in (Serna García et al., 2023b), used within the CoVEffect system, which mines the effects of SARS-CoV-2 mutations and variants based on deep learning (Serna García et al., 2023a).

![Figure 1: Schematic workflow to build OntoEffect.](image)

Lastly, after an informal exploration of the domain supported by experts in molecular biology and bioinformatics, we started the ontological unpacking process (Guizzardi et al., 2021), leading to identifying units of our model (classes), starting from the previously developed terms’ definitions. Then, we also defined the relations among them. The transformation process from each term in the taxonomy into OntoEffect entities was performed manually and very carefully following the previously listed theories of OntoUML, finally obtaining an ontological representation that reveals the ontological semantics of the built taxonomy.

Here, we explored an alternative way to cluster the effects of the taxonomy during the application of the ontological unpacking process. We considered what actor is impacted by each effect; this conceptual classification allowed a clearer representation of the relations between the different entities of the ontology. As a consequence, we obtained four clusters: i) Viral-related, which contains the impacts related to the virus characterization regardless of the host; ii) Host-related, which contains the impacts related to the host characterization regardless of the virus; iii) Virus-host interaction, which contains the impacts related to the interactions between the virus and the host of interest; and iv) Disease and disease management, which contains the impacts involved after the disease is established in the host and those related to the management of the disease (diagnosis, prevention, and treatment protocols).

3.1 UFO and OntoUML Language

Foundational ontologies are comprehensive systems of domain-independent categories and their connections based on solid philosophical principles. These systems consist of axioms that define various elements such as objects, events, causality, spatial-temporal relationships, dependencies, and more. They play a crucial role in representing phenomena across different material domains. Some examples of foundational ontologies include Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) (Borgo and Masolo, 2009), General Formal Ontology (GFO) (Poli et al., 2010), Suggested Upper Merged Ontology (SUMO) (Niles and Pease, 2001), and Unified Foundational Ontology (UFO) (Guizzardi, 2005). By providing a conceptual framework, they enable us to grasp the fundamental nature of specific domains and make it easier to identify connections and interesting associations and/or relationships among data that align with their ontological counterparts (Amaral et al., 2021).

According to (Verdonck and Gailly, 2016), UFO and UFO-based conceptual modeling language (On-
toUML) have emerged among the most used approaches in the field, and since the choice of an upper-level ontology depends on the specific needs and requirements of the application or domain being modeled. OntoUML is an ontology-driven conceptual modeling language whose meta-model complies with the ontological distinctions and axiomatization of UFO (Guizzardi and Wagner, 2003). We believe that UFO is the most suitable top-level ontology to fulfill our needs of explanation of a complex domain – as UFO informs OntoUML, which has been used recently for this purpose in the life sciences domains (Guizzardi et al., 2021; Bernasconi et al., 2022; García S et al., 2022). Therefore, we developed OntoEffect using OntoUML language (Guizzardi, 2005).

OntoUML is an extension of the Unified Modeling Language (UML) (Booch et al., 1997) based on UFO. It incorporates ontological concepts and focuses on modeling not only software systems but also the underlying conceptual structure and semantics of the domains. OntoUML aims to represent entities, relationships, and categories in a way that aligns with formal ontology principles, enabling a more comprehensive understanding of complex systems by modeling them into ontologies. There are three theories in OntoUML specification listed as follows:

- **Types and Individuals**, OntoUML is founded upon the essential distinction between types and individuals. Types are abstract things that are formed to assist in the perception and categorization of the world around us. These things serve as bundles of characteristics that can be anticipated to be encountered in other specific entities - individuals which it can be endurants (e.g., objects) or perdurants (e.g., events).

- **Identity principle**, is a sort of function that is being used to differentiate two individuals, and every individual must have exactly one identity principle (e.g., a fingerprint is an identity principle of a person).

- **Rigidity principle**, endurants can be either rigid when an individual instantiates the same type in all possible scenarios; or anti-rigid when it instantiates a type only in a specific scenario.

Tables[1] and Table[2] schematically represent the ontological distinctions used in OntoEffect, both as class stereotypes and relationships.

## 4 ONTOEFFECT ARCHITECTURE

Here, we present the four views of OntoEffect, corresponding to the listed clusters. The master view that connects all these views (along with the complete views) is available at [https://tinyurl.com/5n8c7k8t](https://tinyurl.com/5n8c7k8t) - not shown in the paper due to space limitations. The schemata were drawn with Visual Paradigm Community Edition (v17.0) and the OntoUML plugin, release 0.5.3 ([https://github.com/OnToUML/ontouml-vp-plugin](https://github.com/OnToUML/ontouml-vp-plugin)). In the following subsections, we wrote in **bold**-type the classes (i.e., entities) of the ontology and *italic*-type the relations between the entities.

### 4.1 Viral-Related View

We present a partition of the viral-related view in Figure[3] OntoEffect aims to represent the impact of viral variation over the overall outcome of the virus behavior; thus, the SARSCOV2 entity represents the virus of our interest, which is already a mutated version of the virus. We note that OntoEffect in its current state does not include entities related to the formation of a SARS-CoV-2 variant nor the composition of its sequence. This is already modeled in the Virus Sequence Annotation view of the ontological unpacking of the Viral Conceptual Model (VCM) (Bernasconi et al., 2022, Guizzardi et al., 2021), which can be easily integrated with OntoEffect.

SARS-CoV2 is an instance of VirusSpecies, which itself is an instance of the Species. Several SARSCOV2 can group forming SARSCOV2Collective. A virus has a viral membrane, a genome, and a proteome represented as ViralMembrane, SARSCOV2Genome, and SARSCOV2Proteome respectively.

A SARSCOV2Gene is a member of SARSCOV2Genome, and it is linked through a material relation with SARSCOV2Protein which is a member of SARSCOV2Proteome. As an example of SARSCOV2Protein, we present Spike protein that is composed of two subunits S1 (which contains the receptor binding domain, RBD) and S2. Moreover, a viral protein could have in its structure a region called Epitope that is a region responsible for provoking an immune response in the host. Each protein has its own characteristics, such as Flexibility, Stability, Functioning, and StructuralOptimization. Moreover, a protein might include multiple NonSynonymousMutation.

Different types of Interactions could occur either inside the same molecule (known as Intramolecular-
Table 1: The types of OntoUML class stereotypes used in OntoEffect.

<table>
<thead>
<tr>
<th>Stereotype</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>are types whose instances are themselves types.</td>
</tr>
<tr>
<td>Category</td>
<td>is a rigid mixin that does not require a dependency to be specified. It is used to aggregate essential properties of individuals who follow different identity principles.</td>
</tr>
<tr>
<td>Collective</td>
<td>is used to represent rigid concepts that provide an identity principle for their instances. The main characteristic of a ⟨Collective⟩ is that it has a homogenous internal structure.</td>
</tr>
<tr>
<td>Kind</td>
<td>is used to represent rigid concepts that provide an identity principle for their instances and do not require a relational dependency. A ⟨Kind⟩ represents a Functional Complex; its parts contribute in different ways to its functionality.</td>
</tr>
<tr>
<td>Subkind</td>
<td>is a construct used to represent rigid specializations of identity providers. By default, its usage does not require a relational dependency.</td>
</tr>
<tr>
<td>Phase</td>
<td>is used to represent anti-rigid subtypes of identity providers that are instantiated by changes in intrinsic properties (e.g., the age of a person).</td>
</tr>
<tr>
<td>Role</td>
<td>is a construct used to represent anti-rigid specializations of identity providers that are instantiated in relational contexts. Like ⟨Phase⟩, all instances of a ⟨Role⟩ must follow the same identity principle.</td>
</tr>
<tr>
<td>Relator</td>
<td>is used to represent truth-makers of material relations, i.e., the “things” that must exist in order for two or more individuals to be connected by material relations.</td>
</tr>
<tr>
<td>Mode</td>
<td>is a particular type of intrinsic property that has no structured value. Modes are individuals that existentially depend on their bearers.</td>
</tr>
<tr>
<td>Quality</td>
<td>is a particular type of intrinsic property that has a structured value. Qualities are things that are existentially dependent on the things they characterize.</td>
</tr>
</tbody>
</table>

4.2 Host-Related View

In Figure 3, a partition of the host-related view is presented. A Human is an instance of the HostSpecies, which itself is an instance of the Species. This schema comprises two sections; the right one describes the different possible phases a human can go through, while the left describes the human’s molecules, cells, and systems.

A Sample is taken from a Human and is mediated through a Diagnosis relator that results in a TestResult which could be a Positive or Negative which characterize if the Human is in Infected phase or NotInfected phase, respectively. A Human is in the Infected phase and plays a HumanHost role. This phase could have different generalizations:

- Living or Deceased host.
- Treated (i.e., playing a treatedIndividual role) or Untreated host.
- According to the host’s history, the infected host could be FirstTimeInfected or ReInfected with the virus.

The NotInfected host plays a HumanReceiver role, that is the human who has the ability to get the viral infection from another host. Therefore, a material relation links the HumanReceiver to the HumanHost, and both are mediated through a ViralTransmission relator. The NotInfected host is either NeverInfected with the virus or got the infection earlier and already Recovered from it.

Also, a Human could receive a vaccine and enter
Table 2: The types of relations used in OntoEffect.

<table>
<thead>
<tr>
<th>Relation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material</td>
<td>⟨Material⟩ relations have a material structure on their own. The relata of a material relation are mediated by individuals that are called relators.</td>
</tr>
<tr>
<td>Mediation</td>
<td>A relation between a ⟨Relator⟩ and the entities it connects. It is a type of existential dependence relation. It can be derived from the relation between the relata and the individuals that compose the relator and that inhere in the relata.</td>
</tr>
<tr>
<td>Characterization</td>
<td>A relation between a bearer type and its feature. A feature is the intrinsic (inherent) moment of its bearer type, and thus existentially dependent on the bearer.</td>
</tr>
<tr>
<td>ComponentOf</td>
<td>Is a parthood relation between two complexes.</td>
</tr>
<tr>
<td>MemberOf</td>
<td>Is a parthood relation between a functional complex or a ⟨Collective⟩ (as a part) and a ⟨Collective⟩ (as a whole).</td>
</tr>
</tbody>
</table>

A HumanGene is a member of HumanGenome, which is found exclusively in the HumanCells. The HumanGene is linked through a material relation with HumanProtein which is a member of HumanProteome. Also, HumanGene and HumanProtein are participants of the HUMANProteinCodingProcess.

The HostInfectedCells are specifically the group of HumanCells that are infected with the virus (or have the ability to be infected). Each of these cells has a membrane (represented as HostCellMembrane) that encloses different organelles, apparatus, and proteins (represented as InfectedCellProtein).

Since the HumanProtein is a big class of molecules, we only represent:

- InfectedCellProteins.
- ImmuneSystemProteins, which are the proteins of the immune system, e.g. Immunoglobulin and Antibody.
- HumanEnzyme like the Angiotensin-converting enzyme 2 (ACE2) which plays a role as a Receptor of the virus. Other enzymes are the HumanProtease, like TMPRSS2 (i.e., Transmembrane protease serine 2, playing a role as TransmembraneProtease), Furin (playing a role as Pro-ProteinConvertase), and Cathepsin L.

Finally, we have the Neutralization category, which involves different types of NeutralizingAgents that can neutralize the virus and prevent it from entering the host cell. Such agents (besides the Antibodies that are naturally part of the ImmuneSystem) are TherapeuticAntibody as the laboratory manufactured antibodies, ConvalescentSera as the blood serum that is obtained from an individual who has Recovered from a previous SARS-CoV2 infection, and VaccinatedSera as the blood serum that is obtained from an individual who has been Vaccinated with one
of the SARS-CoV-2 vaccines. Finally, measurements of the sensitivity of the virus to each of these agents can be characterized as qualities of SensitivityToAntibodies, SensitivityToConvalescentSera, and SensitivityToVaccinatedSera.

4.3 Virus-Host Interaction View

In Figure 4, a partition of the virus-host interaction view is presented. To establish a viral infection, there are plenty of interactions between the virus and its host that are needed. We collected these interactions in one category that is called HostVirusInteraction, which is an instance of Interactions. ViralInfectionFormation, is a relator that mediates SARS-CoV2Collective and Human. To say that a cell is infected with a virus, three main steps must have happened (represented as instances of the HostVirusInteraction):

1. **ViralEntry**, a category that represents the process of a virus entering a host cell through one of two entry pathways; CellSurfaceEntry and EndosomalEntry. The efficiency of this process can be measured, and it is represented with a quality called EntryEfficiency.

2. **ViralReplication**, after that the virus enters the cell or its genome enters the host cell (through EndosomalEntry or CellSurfaceEntry, respectively). The SARS-CoV2genome will be replicated and translated into SARS-CoV2Protein through SARS-CoV2ProteinCodingProcess with the help of InfectedCellProtein.

3. **ViralRelease**, after that a host cell becomes infected with the virus and starts to replicate the virus and biosynthesis it, this cell becomes a virus-producer cell which releases the mature virus to infect other host cells in a process called ViralRelease.

We are focusing on the viral entry process, as it is the initial step of the ViralInfectionFormation. Preventing this process, either by a vaccine or a treatment, means preventing the disease. Both entry pathways start with the virus binding to the host cell receptor in which the RBD of the Spike protein binds to the ACE2 receptor. This process is presented through the BindingToHostReceptorInteraction relator (has BindingAffinityToHostReceptor as a quality). This binding induces conformational changes in the Spike, which is followed by the S2 cleavage through CleavingInteraction. Depending on the entry route, the S2...
Figure 4: A partition of the Host-virus interaction view.

is cleaved by different HumanProtease; using TM-PRSS2 in the CellSurfaceEntry or using Cathepsin L in the other pathway. Regardless of which entry pathway the virus follows, a membrane fusion must occur that is represented through MembraneFusionInteraction relator that mediates ViralMembrane and HostCellMembrane since the fusion of these two membranes is crucial to uncoating the genome of SARS-CoV-2 to start the ViralReplication.

Just like MERS-CoV, the spike protein of SARS-CoV-2 virus is cleaved by Furin into S1 and S2 sub-units during their biosynthesis; therefore, this cleavage is also mediated through the CleavingInteraction.

Another class of HostVirusInteraction represents the ImmuneResponse of the host toward the presence of the virus inside it. To have such a response, Epitope from a viral protein must be recognized by an Antibody or Immunoglobulin. The immune response is a complex cascade of immune mechanisms recruiting various immune cells and chemical mediators. Therefore, we simplified it into one relator called ImmuneResponse. A virus could obtain the power to escape the host immune system and minimize or completely vanish the ImmuneResponse; this phenomenon can be measured and is represented as ImmuneEscape. The binding between Epitope and Antibody is mediated through BindingToAntibody relator (has BindingAffinityToAntibody as a quality). Moreover, there is NeutralizationProcess relator that mediates the neutralization of SARSCOV2 inside a HumanHost by the previously mentioned NeutralizingAgent.

Since BindingToHostReceptorInteraction, CleavingInteraction, NeutralizationProcess, and BindingToAntibody are interactions between a HumanProtein and a SARSCOV2Protein, we generalized them in one category; ProteinProteinInteraction.

4.4 Disease and Disease Management View

In Figure 5 we present a partition of the view of On-toEffect that characterizes the disease and its management through diagnosis, treatment, and prevention. We started this view form Human that could admit into a Hospital (for many reasons, e.g., suspicion of infection, receiving treatment or vaccine, or having severe symptoms that require hospitalization) through HospitalAdmission relator; also, many factors could increase or decrease the RiskOfHospitalization of a
COVID19 is the Disease caused by SARS-CoV2 infection and beside it an Infected host could suffer from (or have) OtherDisease. Such a scenario might lead to complications between COVID19 and that OtherDisease and worsen the overall health status of the Infected host; this was mediated through a Complications relator. The FatalityRate is a quality that is calculated from the InfectedCollective and is the proportion of persons who Deceased after the viral infection over the number of confirmed Infected people.

COVID19, as any other viral disease, is characterized by InfectionDuration, which is the duration of the host being infected with the virus. This disease has different ranges of symptoms, which can be represented as phases which are: WithoutSymptoms, MildSymptoms, ModerateSymptoms, and SeverSymptoms. The symptoms of the disease determine its severity; therefore, the different phases of symptoms are associated with the DiseaseSeverity relator. Furthermore, an InfectedSample can be obtained from the HumanHost to measure the ViralLoad inside it; ViralLoad is the number of copies of RNA of a given virus per milliliter.

Lastly, there is a risk of a second infection, represented as Reinfection relator, by the same virus after being Recovered from (or during the course of) a primary infection; such risk can be measured, and it is represented with the RiskOfReinfection quality.

For disease diagnosis, as previously mentioned, a Sample is taken from a Human and mediated with Diagnostics collective through a Diagnosis relator. A sample is used in a diagnostic test to detect either the presence of SARS-CoV2Protein (with AntigenicTest, or MolecularTestPCR which has a CtValue) or the presence of ImmuneSystemProtein (with SerologicalTest) through DetectionProcess. Such Diagnostics have measurements for their effectiveness in detecting the presence of its target molecules; we characterized it with a quality called EffectivenessOfDiagnostics.

Considering the treatment of the disease, a category of AntiviralDrug is mediated with the Treatment relator and associated with different kinds of drugs such as EntryInhibitors, AttachmentInhibitors, and PolymeraseInhibitors. Another mediation relation with the Treatment relator is with TreatedIndividual that is the role played by an Treated host, and it represents an Infected host who is Treated with any kind of AntiviralDrugs. The efficacy of a drug to treat the infected host can be measured, and it is represented as EffectivenessOfDrug quality.

Finally, considering the prevention protocols of the disease, we are addressing the Vaccination rather than other protocols (like self-isolation) since we are focusing on the effects of the viral variation on the outcome of the virus and its resulting disease. A Human enters the Vaccinated phase and plays a VaccinatedIndividual role if a Vaccine was received, and the Vaccination relator mediates this process. A Vaccine is a NeutralizingAgent, and there are multiple kinds of vaccines that were developed to prevent the disease, e.g., Pfizer, Moderna, and AstraZeneca. The efficacy of a vaccine to neutralize the virus and prevent viral infection formation can be measured, and it is represented as EffectivenessOfVaccine quality.

5 VALIDATION AND CONCLUSIONS

This paper presents OntoEffect, an OntoUML-based ontology that explains SARS-CoV-2 variants’ effects. As a core part of the research, we presented the OntoUML model that represents classes and relationships describing the domain. OntoEffect aims to provide a detailed analysis of the complex notions of the impacts of SARS-CoV-2 variations. The main challenges of this effort are related to the complexity of describing the biological domain with knowledge representation artifacts. Many concepts are entangled or nested, and their semantics are non-homogenous; thus, a deep understanding is needed and can be achieved through an ontological representation.

Our current version of the ontology is undergoing continuous updating of information. We designed semi-supervised methods for extracting content from the CORD-19 literature corpus to continuously collect instances of knowledge-related entities (Serna García et al., 2023a). Understanding and mastering the complex domain of SARS-CoV-2 mutation effects has proven extremely useful, in our data-driven analysis of SARS-CoV-2 phenomena, where data science methods were integrated with knowledge-driven hypothesis making (Bernasconi et al., 2021b, Bernasconi et al., 2021a).

As a preliminary assessment of this effort, we presented the ontology to three domain experts, who have a background in molecular biology and previously worked on several projects with viral genomic sequences; they were asked to evaluate the ontology for accuracy, relevance, and completeness. We had two focus groups with the experts, resulting in a series of observations that will inspire our next refinement of the views. Generally, our modeling choices.
were confirmed (after we properly guided their understanding of OntoUML stereotypes). This work requires broader validation by means of empirical studies, along the lines of (Verdonck et al., 2019; García S et al., 2022a; García S. et al., 2023)—showing the benefits of this representation when the explanation of such a complex domain is required to non-expert stakeholders. All in all, we here defend that OntoEffect presents a first step towards the full understanding of SARS-CoV-2 impacts to non-domain experts, powering intra-domain knowledge exchange and information interoperability. This schema will further enable knowledge-based applications for annotation and reasoning.

REFERENCES


IEEE.


