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

Ruba Al Khalaf[1][a], Anna Bernasconi[1][b], and Alberto García S.[2][c]
[1]*Politecnico di Milano, Milan, Italy*
[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 (WHO, 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).

[a] https://orcid.org/0000-0002-5645-5886
[b] https://orcid.org/0000-0001-8016-5750
[c] https://orcid.org/0000-0001-5910-4363

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.

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 Sampiling*, *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 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 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

To identify the effects of interest that are suited to be considered as entities in OntoEffect, we start by build-

ing a taxonomy (i.e., controlled vocabulary) of effects and transforming it into a structured ontology. The workflow for structuring OntoEffect is shown in figure 1.

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 is its high representativeness in the literature and connection with at least one non-synonymous mutation (or variant) of SARS-CoV-2.

To structure the taxonomy, we classified the selected terms into four categories according to the field of relevance of the impact: *Epidemiology, Immunology, Viral Kinetics and dynamics,* and *Diagnosis, prevention, and treatments*.

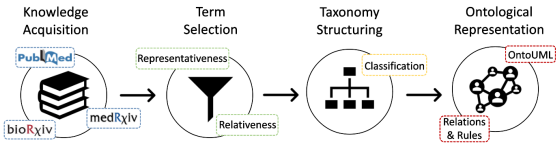

Figure 1: Schematic workflow to build OntoEffect.

Lastly, and after an informal exploration of the domain – supported by experts in molecular biology and bioinformatics, we started the ontological unpacking process 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 was performed manually and very carefully, finally obtaining an ontological representation that reveals the ontological semantics of the built taxonomy.

During the ontological unpacking process, instead of clustering according to the impacted field, we explored an alternative and more convenient way to cluster the effects related to what actor is impacted. As a consequence, we obtained four other different 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 et al., 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 facilitate the identification of correlations and intriguing relationships among data that correspond to their ontological counterparts (Amaral et al., 2021).

According to (Verdonck and Gailly, 2016), UFO and UFO-based conceptual modeling language (OntoUML) 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, 2004). OntoUML is based on Unified Modeling Language (UML) (Booch et al., 1997). It has been extended to support the modeling of ontologies, which involves modeling not only the structure of concepts but also the relationships between them and their properties.

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., 2022b). Therefore, we developed OntoEffect using OntoUML language (Guizzardi, 2005).

Table1 and Table2 schematically represent the ontological distinctions used in OntoEffect, both as class stereotypes and relationships.

Table 1: The types of OntoUML class stereotypes used in OntoEffect.

Stereotype	Definition
Type	are types whose instances are themselves types.
Category	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.
Collective	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.
Kind	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.
Subkind	is a construct used to represent rigid specializations of identity providers. By default, its usage does not require a relational dependency.
Phase	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).
Role	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.
Relator	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.
Mode	is a particular type of intrinsic property that has no structured value. Modes are individuals that existentially depend on their bearers.
Quality	is a particular type of intrinsic property that has a structured value. Qualities are things that are existentially dependent on the things they characterize.

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–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). 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 2. Since in OntoEffect we want to represent the impact of viral variation over the overall outcome of the virus behavior, the **SARSCOV2** entity represents the virus of our interest, which is already a mutated version of the virus. **SARSCOV2** 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 **IntramolecularInteractions**) or between molecules (known as **IntermolecularInteraction**). Instead, **IntraviralProteinProetinInteractions** is a bundle of relational dispositions connecting **SARSCOV2Protein** with itself

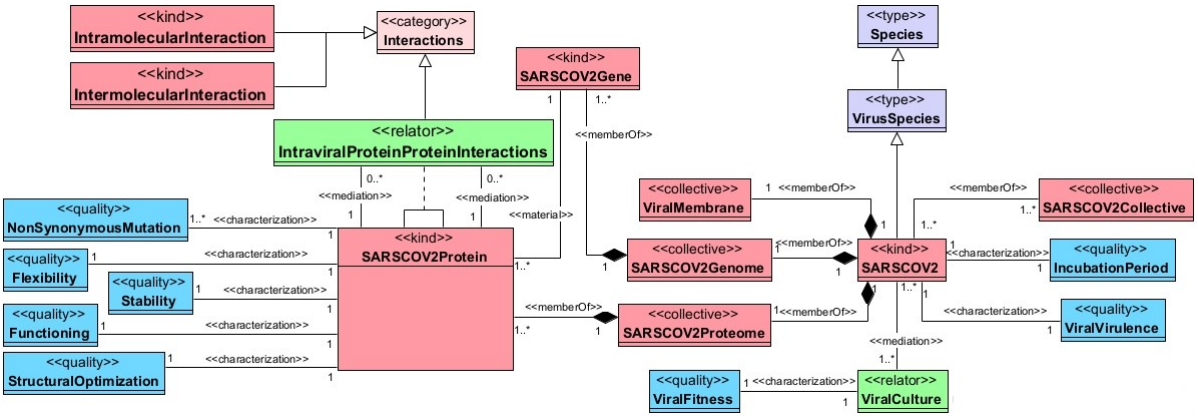

Figure 2: A partition of the viral related view.

Table 2: The types of relations used in OntoEffect.

Relation	Definition
Material	⟨Material⟩ relations have a material structure on their own. The relata of a material relation are mediated by individuals that are called relators.
Mediation	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.
Characterization	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.
ComponentOf	Is a parthood relation between two complexes.
MemberOf	Is a parthood relation between a functional complex or a ⟨Collective⟩ (as a part) and a ⟨Collective⟩ (as a whole).

through an association.

Other characteristics of the virus are the **IncubationPeriod**, which is the interval between the moment a host is infected and the onset of the disease, and the **ViralVirulence** which is the competence of the infectious agent to produce pathologic effects.

Finally, a virus could incubate in a specific **Culture**, which can be specified through **CultureConditions** (such as **Temperature**), **CultureMediums**, and the employed **CellType** and **CellLine**. We represent this incubation through a **ViralCulture** relator in which we can assess the **ViralFitness** of the virus in a specific environment.

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:

- **Livinig** 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

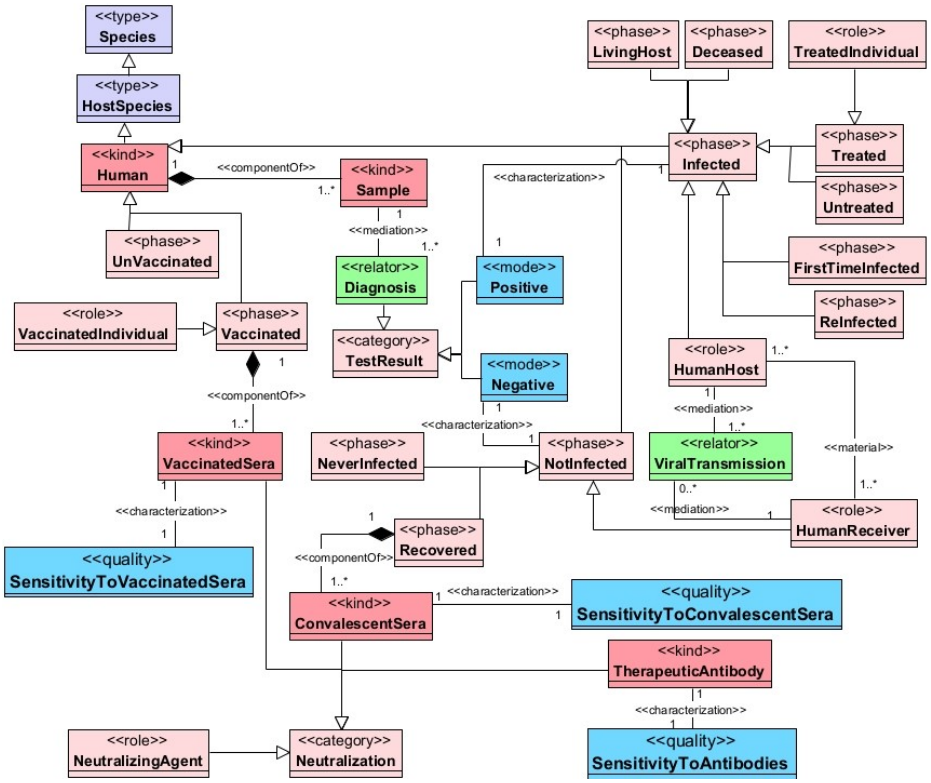

Figure 3: A partition of the host-related view.

the **Vaccinated** phase and plays a **VaccinatedIndividual** role, or could stay **UnVaccinated**.

Human is more complex species than a **Virus**; hence, it is composed of different systems, e.g. **ImmuneSystem** which involves **WhiteBloodCells**. A subkind of the **WhiteBloodCells** is **Lymphocytes** which play a role in the immune response for fighting infectious agent (Wheelock and Toy, 1973). The two main types of lymphocytes are **Tcells** and **Bcells**.

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 **ProProteinConvertase**), 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 **SARSCOV2** 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 **SensitivityToAntibody**, **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 **Interacrions**. **ViralInfectionFormation**, is a relator that mediates **SARSCOV2Collective** 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 **SARSCOV2genome** will be replicated and translated into **SARSCOV2Protein** through **SARSCOV2ProteinCodingProcess** 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 is cleaved by different **HumanProtease**; using **TMPRSS2** 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** subunits during their biosynthesis; therefore, this cleavage is also mediated through the **CleavingInteraction**.

Another class of **HostVirusInteraction** represents the **ImmuneReponse** 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 OntoEffect 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 **Human** in a **Hospital**.

COVID19 is the **Disease** caused by **SARSCOV2** 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**

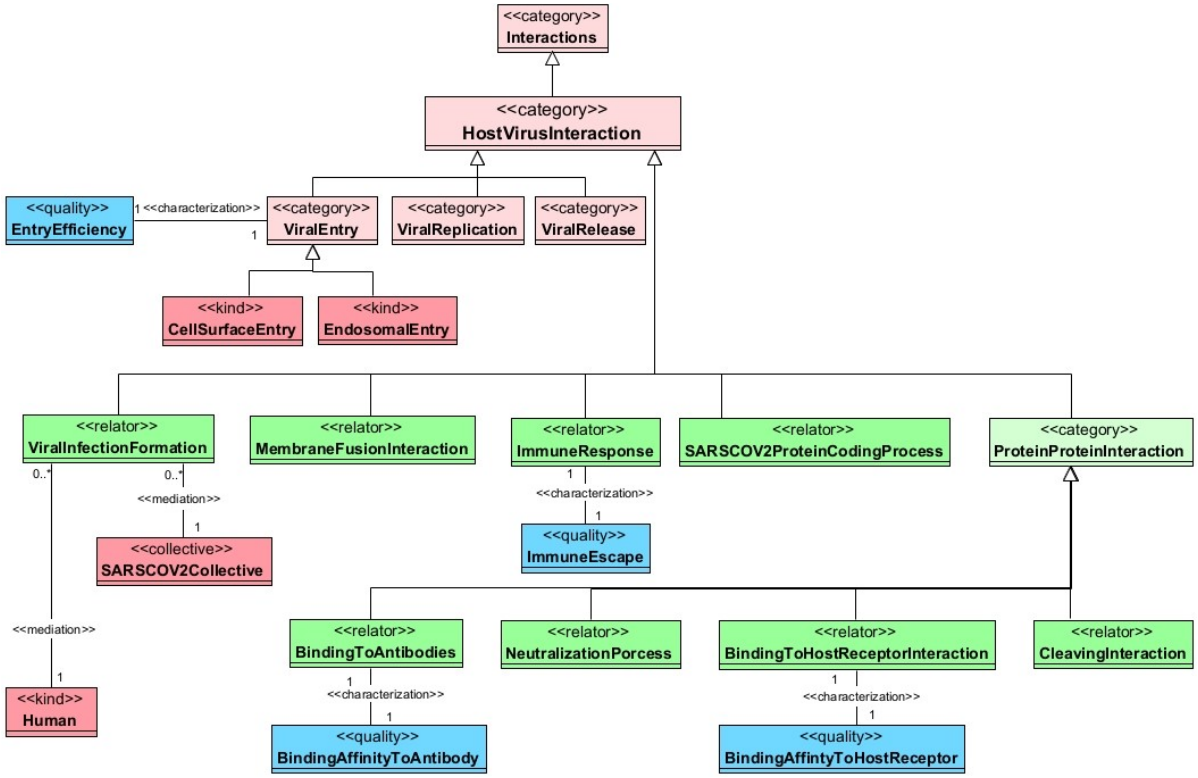

Figure 4: A partition of the Host-virus interaction view.

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 **SARSCOV2Protein** (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 **EffectivnessOfDiagnostics**.

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 **EffectivnessOfDrug** 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

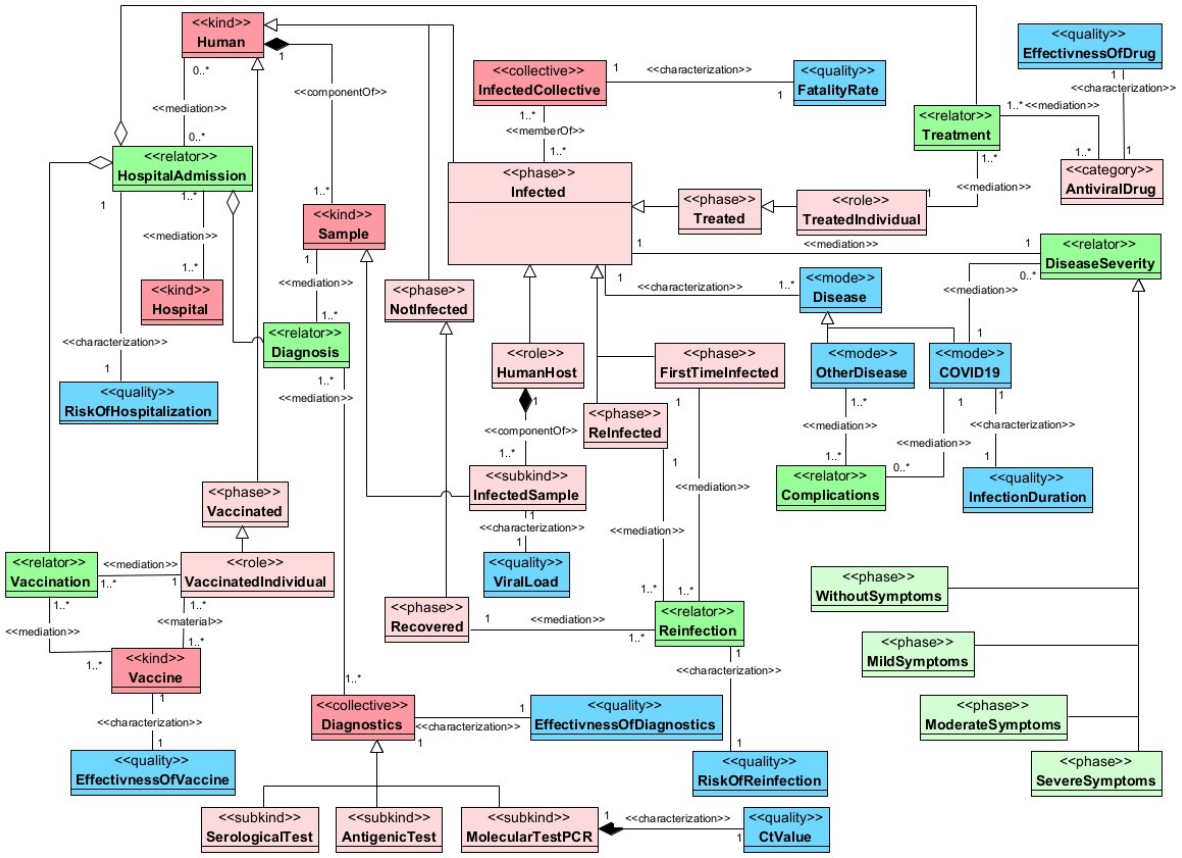

Figure 5: A partition of the disease and disease management view.

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 **EffectivnessOfVaccine** 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.

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. 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, on the lines of (García S et al., 2022a; Verdonck et al., 2019)–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 OntoUML presents a first step towards the full understanding of SARS-CoV-2 impacts to non-domain experts, powering intra-domains knowledge exchange and information interoperability. This schema will further enable knowledge-based applications for annotation and reasoning.

REFERENCES

Ahmad, A., Bandara, M., Fahmideh, M., Proper, H. A., Guizzardi, G., and Soar, J. (2021). An overview of ontologies and tool support for covid-19 analytics. In *2021 IEEE 25th International Enterprise Distributed Object Computing Workshop (EDOCW)*, pages 1–8. IEEE.

Amaral, G., Baiao, F., and Guizzardi, G. (2021). Foundational ontologies, ontology-driven conceptual modeling, and their multiple benefits to data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11(4):e1408.

Arp, R., Smith, B., and Spear, A. D. (2015). *Building ontologies with basic formal ontology*. Mit Press.

Barcelos, P. P. F., Sales, T. P., Fumagalli, M., Fonseca, C. M., Sousa, I. V., Romanenko, E., Kritz, J., and Guizzardi, G. (2022). A fair model catalog for ontology-driven conceptual modeling research. In *Conceptual Modeling: 41st International Conference, ER 2022, Hyderabad, India, October 17–20, 2022, Proceedings*, pages 3–17. Springer.

Bernasconi, A., Canakoglu, A., Pinoli, P., and Ceri, S. (2020). Empowering virus sequence research through conceptual modeling. In *Conceptual Modeling: 39th International Conference, ER 2020, Vienna, Austria, November 3–6, 2020, Proceedings 39*, pages 388–402. Springer.

Bernasconi, A., Guizzardi, G., Pastor, O., and Storey, V. C. (2022). Semantic interoperability: ontological unpacking of a viral conceptual model. *BMC bioinformatics*, 23(11):1–23.

Booch, G., Rumbaugh, J., and Jackobson, I. (1997). Uml: unified modeling language. *Versão*.

Borgo, S. and Masolo, C. (2009). Handbook on ontologies.

Fabio, I., Amaral, G. C., Griffo, C., Baião, F., and Guizzardi, G. (2021). " what exactly is a lockdown?": towards an ontology-based modeling of lockdown interventions during the covid-19 pandemic. In *ONTOBRAS*, pages 151–165.

García S, A., Bernasconi, A., Guizzardi, G., Pastor, O., Storey, V. C., and Costa, M. (2022a). An initial empirical assessment of an ontological model of the human genome. In *International Conference on Conceptual Modeling*, pages 55–65. Springer.

García S, A., Guizzardi, G., Pastor, O., Storey, V. C., and Bernasconi, A. (2022b). An ontological characterization of a conceptual model of the human genome. In *Intelligent Information Systems: CAiSE Forum 2022, Leuven, Belgium, June 6–10, 2022, Proceedings*, pages 27–35. Springer.

Guizzardi, G. (2005). *Ontological foundations for structural conceptual models*. CTIT, Centre for Telematics and Information Technology, Twente, Netherlands.

Guizzardi, G., Bernasconi, A., Pastor, O., and Storey, V. C. (2021). Ontological unpacking as explanation: the case of the viral conceptual model. In *Conceptual Modeling: 40th International Conference, ER 2021, Virtual Event, October 18–21, 2021, Proceedings 40*, pages 356–366. Springer.

Guizzardi, G. and Guarino, N. (2023). Semantics, ontology and explanation. *arXiv preprint arXiv:2304.11124*.

Guizzardi, G. and Wagner, G. (2004). A unified foundational ontology and some applications of it in business modeling. In *Proceedings of the Open InterOp Workshop on Enterprise Modelling and Ontologies for Interoperability, co-located with CAiSE'04 Conference*.

He, Y., Yu, H., Ong, E., Wang, Y., Liu, Y., Huffman, A., Huang, H.-h., Beverley, J., Hur, J., Yang, X., et al. (2020). Cido, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis. *Scientific data*, 7(1):181.

Lauring, A. S., Frydman, J., and Andino, R. (2013). The role of mutational robustness in rna virus evolution. *Nature Reviews Microbiology*, 11(5):327–336.

Mattenberger, F., Vila-Nistal, M., and Geller, R. (2021). Increased rna virus population diversity improves adaptability. *Scientific Reports*, 11(1):6824.

Niles, I., Pease, A., and Zúñiga, G. (2001). Fois'01: Proceedings of the international conference on formal ontology in information systems.

Poli, R., Healy, M., and Kameas, A. (2010). *Theory and applications of ontology: Computer applications*. Springer.

Romanenko, E., Calvanese, D., and Guizzardi, G. (2022). Towards pragmatic explanations for domain ontologies. In *Knowledge Engineering and Knowledge Management: 23rd International Conference, EKAW 2022, Bolzano, Italy, September 26–29, 2022, Proceedings*, pages 201–208. Springer.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungall, C. J., et al. (2007). The obo foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, 25(11):1251–1255.

Tanaka, H., Hirayama, A., Nagai, H., Shirai, C., Takahashi, Y., Shinomiya, H., Taniguchi, C., and Ogata, T. (2021). Increased transmissibility of the sars-cov-2 alpha variant in a japanese population. *International journal of environmental research and public health*, 18(15):7752.

Verdonck, M. and Gailly, F. (2016). Insights on the use and application of ontology and conceptual modeling languages in ontology-driven conceptual modeling. In *Conceptual Modeling: 35th International Conference, ER 2016, Gifu, Japan, November 14-17, 2016, Proceedings 35*, pages 83–97. Springer.

Verdonck, M., Gailly, F., Pergl, R., Guizzardi, G., Martins, B., and Pastor, O. (2019). Comparing traditional conceptual modeling with ontology-driven conceptual modeling: An empirical study. *Information Systems*, 81:92–103.

Wheelock, E. F. and Toy, S. T. (1973). Participation of lymphocytes in viral infections. *Advances in immunology*, 16:123–184.

WHO (2023). Statement on the fifteenth meeting of the ihr (2005) emergency committee on the covid-19 pandemic. In *2023-05-05)[2023-05-18]. https://www.who. int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations*.