

OntoCovid: Aplicando SABiO para a modelagem conceitual bem fundamentada no domínio da COVID-19

Lucas Maddalena, Fernanda Baião

Departamento de Engenharia Industrial – Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rio de Janeiro – RJ – Brazil

lucasmadda@aluno.puc-rio.br, fbaiao@puc-rio.br

Abstract. *The relevance of foundational ontologies and well-founded conceptual models is acknowledged in several contexts for making the real-world semantics of data explicit, and for reducing the impact of semantic ambiguities during the integration and manipulation of different data sources. Few domains pose such demands as urgently as the analysis and knowledge extraction from COVID-19 data. Since COVID-19 was declared a pandemic in early 2020, huge efforts from around the world provided an avalanche of data for research and analysis at an unprecedented rate. However, the coexistence of semantically divergent and non-explicit definitions for data from distinct countries and time periods that are being integrated and analyzed makes the conclusions of such analysis and the extracted knowledge potentially questionable. This work contributes to the development of a preliminary version of OntoCOVID, an ontology for the domain of COVID-19 well-founded in UFO and built using the SABiO methodology.*

Resumo. *A relevância das ontologias de fundamentação e de modelos conceituais bem fundamentados para tornar explícita a semântica de mundo real dos dados, dirimindo ambiguidades na integração e manipulação de diferentes fontes, é amplamente reconhecida em diversos contextos. No entanto, poucos domínios de aplicação apresentam tais demandas de forma tão urgente quanto a epidemia de COVID-19, em que esforços de todo o mundo propiciaram uma avalanche de dados para análises e pesquisas científicas sem precedentes. Porém, a coexistência de definições semanticamente divergentes e não explícitas para conjuntos de dados manipulados de forma conjunta torna potencialmente questionável boa parte das conclusões destas análises. Esse trabalho contribui com o desenvolvimento, via metodologia SABiO, de uma versão preliminar da OntoCOVID, uma ontologia bem fundamentada em UFO para o domínio da COVID-19.*

1. Introdução

Desde seu aparecimento em Dezembro de 2019, SARS-CoV-2, o vírus que causa a doença *coronavirus disease 2019* (COVID-19), tomou proporções elevadíssimas; em Agosto de 2021, já haviam sido contabilizados mais de 200 milhões de casos e mais de 4 milhões mortes por COVID-19 pelo mundo (OMS, 2021).

Apesar da extensa quantidade de informação cientificamente reconhecida que vem sendo tornada disponível para análise, existem discussões sobre como enfrentar a doença e o vírus (Guimarães et. al, 2020). Um ponto crucial é a definição dos conceitos que dão base

a estes dados, nem sempre estabelecidos de forma consensual e estável entre diferentes regiões e ao longo do tempo. Tsang et. al (2020) identificaram sete mudanças de definição de Caso COVID-19 da Comissão Nacional de Saúde Chinesa, e a contagem de casos entre uma definição e outra chega a aumentar em aproximadamente sete vezes. Ou seja, é imperioso criar mecanismos para definir formalmente e de forma explícita os conceitos que estão sendo considerados na disponibilização dos dados.

Um dos principais objetivos das ontologias, e em particular das ontologias de fundamentação, é deixar explícita a semântica de mundo real dos dados (Guizzardi, 2020). Segundo Guarino et (2009), ontologias podem ser classificadas em diferentes tipos. Ontologias de Fundamentação fornecem meta-propriedades para construtos independentes de domínio, e ontologias de domínio instanciam Ontologias de Fundamentação a fim de prover uma conceituação em um universo de discurso particular.

Assim, este trabalho descreve o desenvolvimento preliminar da OntoCOVID, uma ontologia bem fundamentada na UFO para o domínio da COVID-19, com o objetivo de esclarecer dúvidas presentes nas discussões sobre o assunto e constituir um artefato relevante aos que trabalham com os dados epidemiológicos para aplicação de técnicas e algoritmos de extração de conhecimento acerca da doença (Amaral et. al, 2021). Foi aplicada metodologia SABiO (Falbo, 2014), a qual propõe processos que devem ser cumpridos para criar uma ontologia de forma sistemática.

A seção 2 apresenta as principais etapas da metodologia SABiO, cuja ontologia de domínio resultante - e principal proposta deste trabalho - é descrita na Seção 3. A Seção 4 apresenta algumas discussões sobre os conceitos explicitados na OntoCOVID, e aponta algumas direções de trabalhos futuros.

2. Metodologia Adotada e Breve Referencial Teórico

Proposta por Falbo (2014), a metodologia SABiO (*Systematic Approach for Building Ontologies* ou Abordagem Sistemática para Construção de Ontologias) foi adotada nesse trabalho. Ainda, como linguagem de representação da ontologia proposta, adotou-se a OntoUML, a qual considera as distinções ontológicas da *Unified Foundational Ontology* (UFO) e as incorpora aos diagramas de classe UML como forma de representação da ontologia como artefato (Guizzardi, 2005).

SABiO propõe cinco atividades no Processo de Desenvolvimento: (1) Levantamento de Requisitos e Identificação de Propósito, (2) Captura e Formalização da Ontologia, (3) Design, (4) Implementação e (5) Teste; além destas, são também 5 atividades no Processo de Suporte ao Desenvolvimento: (1) Aquisição de Conhecimento, (2) Reuso, (3) Documentação, (4) Gerência de Configuração e (5) Avaliação. O presente trabalho se limita às atividades de Levantamento de Requisitos e Identificação de Propósito, e Captura e Formalização da Ontologia, uma vez que o objetivo é desenvolver uma Ontologia a nível conceitual, sem levar em conta questões para sua codificação. Vale destacar a atividade de suporte Aquisição de Conhecimento, pois para a busca dos significados de conceitos na área da Saúde (mais especificamente de doenças infecciosas) o trabalho reusou as definições da Organização Mundial da Saúde (OMS¹) e do Centro Europeu para Prevenção e Controle de Doenças (*European Centre for Disease Prevention*

¹ <https://www.who.int/>

and Control - ECDC²). Há também a atividade "Reuso", uma vez que ontologias podem ser reusadas, totalmente ou parcialmente (Falbo, 2014). Assim, a criação da OntoCOVID observou outras ontologias e trabalhos da área da saúde encontrados na literatura, como a OntoTB (Guarnier et. al, 2020) e a OntoSaúde (Costa et. Al, 2020). A Seção 3 detalha a aplicação das etapas da SABIO para a construção da OntoCOVID.

3. Modelagem Conceitual do Domínio da COVID-19

A primeira etapa da metodologia SABiO determina a identificação de Questões de Competência (QC). Uma QC é representada através de uma pergunta que a ontologia deverá ser capaz de responder. As QC são um meio efetivo de se definir o escopo pretendido para a ontologia. Sendo assim, para a OntoCOVID foi definida a seguinte lista preliminar de questões de competência, endereçadas neste trabalho: (QC1) O que é a COVID-19? (QC2) O que causa a COVID-19? (QC3) O que caracteriza um indivíduo com COVID-19? (QC4) O que é um caso de COVID-19, segundo a OMS e o ECDC?

Na etapa de Captura e Formalização da Ontologia, SABiO determina que se deve identificar e organizar os conceitos relevantes do domínio. Para a OntoCOVID, os conceitos necessários para endereçar as QC foram definidos segundo a OMS ou o ECDC, conforme descrito a seguir e estruturado como ilustrado nas Figuras 1 e 2.

Segundo a OMS (2011), uma **Doença** é caracterizada por uma **Disfunção** (fisiológica ou psicológica), a qual é uma propriedade intrínseca (portanto, um *Mode* da UFO) que caracteriza uma **Pessoa** e que é externamente dependente desta **Doença**. Uma **Doença** é causada por um **Principal Agente Causador** (que pode ser um **Vírus**, um **Protozoário** ou mesmo uma **Predisposição Genética**) e, portanto, caracteriza-se como uma *Category* na UFO). Além disso, um subtipo específico de **Doença** é **Doença Infeciosa**, cujo **Principal Agente Causador** que a causa é especificamente um **Microrganismo Patogênico**, que pode ser uma bactéria, um vírus, um parasita ou um fungo (OMS, 2021). Dentre os vírus existentes e já conhecidos, o **SARS-CoV-2** é um deles. Em específico, a **COVID-19** é um subtipo particular de **Doença Infeciosa**, cujo **Principal Agente Causador** é o vírus **SARS-CoV-2** (OMS, 2020).

Considera-se que uma **Pessoa** está exclusivamente viva (**Pessoa Viva**) ou morta (**Pessoa Morta**), segundo as meta-propriedades do construto *Phase* da UFO. Desta forma, a OntoCOVID assume que uma **Pessoa Morta** é uma **Pessoa**, e o **Óbito** é o evento que dispara a transição de uma **Pessoa Viva** para uma **Pessoa Morta**.

Uma **Pessoa Com COVID-19** é uma **Pessoa Viva** que se encontra infectada pelo **SARS-CoV-2**, ou seja, que é caracterizada por uma **Carga Viral de SARS-CoV-2** quantificada como um valor (*quale*, da UFO) não nulo. Uma **Pessoa Com COVID-19** pode participar em um ou mais eventos que dão origem ao relacionamento material (portanto, um *Relator*) de **Transmissão de SARS-CoV-2**. Neste relacionamento, a **Pessoa Com COVID-19** (*qua Pessoa Transmissora*) é quem transmite parte de sua **Carga Viral de SARS-CoV-2** para outra **Pessoa** (*qua Pessoa Transmitida*). Uma **Transmissão de SARS-CoV-2** (que é intrinsecamente dependente da transmissora e externamente dependente da transmitida) dispara um evento que faz surgir um outro relacionamento material (portanto, outro *Relator*) de **Infecção de SARS-CoV-2**, fazendo com que a **Pessoa Transmitida** assuma também o papel de **Pessoa Com COVID-19**.

² <https://www.ecdc.europa.eu/en>

Vale ressaltar que uma **Infeção de SARS-CoV-2** (tanto nos casos de transmissão direta quanto indireta) é intrinsecamente dependente da **Pessoa Com COVID-19**, e externamente dependente da **Pessoa Transmissora**.

Quanto aos sintomas causados pela **COVID-19**, não há um consenso entre os órgãos, os quais definem diferentes subconjuntos de sintomas típicos como critério clínico para diagnosticar se uma pessoa está infectada. A OMS reconhece que a COVID-19 afeta pessoas de diferentes maneiras, e definiu **Sintomas de Critério Clínico COVID-19 Reconhecidos Pela OMS** incluem, dentre outros: **Febre, Tosse, Fraqueza/Fadiga Geral** (na ontologia proposta, adotamos este conjunto como os **Sintomas COVID-19**). Já segundo o ECDC, os **Sintomas de Critério Clínico COVID-19 Reconhecidos Pelo ECDC** incluem: **Febre, Tosse, Dispneia e Anosmia/Hiposmia/Ageusia Repentina**. Esta diferença no conjunto de sintomas reconhecido por cada órgão impacta a definição do que é um **Caso COVID-19 Segundo ECDC** e **Caso COVID-19 Segundo OMS** (Figura 2).

Ainda, uma **Pessoa Com COVID-19** pode ou não apresentar um ou vários **Sintomas De COVID-19**. Uma **Pessoa Com COVID-19 Sintomática** é uma fase (*Phase*) de **Pessoa Com COVID-19** definida enquanto ela apresenta pelo menos um **Sintoma De COVID-19**; a outra fase (*Phase*) complementar a esta (**Pessoa Com COVID-19 Assintomática**) dura enquanto ela não apresenta nenhum **Sintoma De COVID-19**.

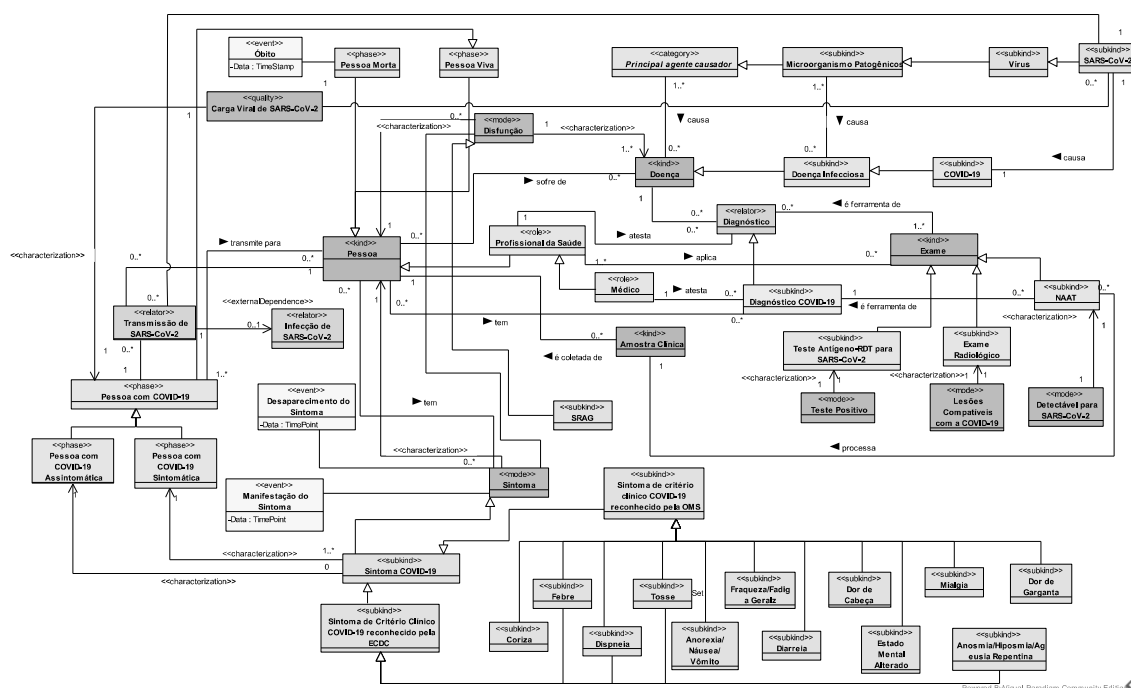


Figura 1. Ontologia OntoCOVID-19 – Doença

De acordo com o ECDC (2020), um **Caso COVID-19 Segundo ECDC** se refere a uma **Pessoa** e pode assumir 3 papéis (*Roles*): **Caso Possível COVID-19 Segundo ECDC**, **Caso Provável COVID-19 Segundo ECDC** e **Caso Confirmado COVID-19 Segundo ECDC**. Para definir precisamente cada um destes subtipos, foram explicitadas as regras (*normative descriptions*) correspondentes ao **Critério Clínico Segundo ECDC**, **Critério Epidemiológico Segundo ECDC**, **Critério de Diagnóstico de Imagem Segundo ECDC** e **Critério Laboratorial Segundo ECDC**.

Uma **Pessoa** que satisfaz ao **Critério Clínico Segundo ECDC** apresenta ao menos dois dentre os **Sintomas de Critério Clínico COVID-19 Reconhecidos Pelo ECDC**. Já a **Pessoa** que satisfaz ao **Critério Epidemiológico Segundo ECDC** é aquela que satisfaz a alguma das seguintes condições: (i) esteve em contato próximo com uma **Pessoa com COVID-19 Confirmado Segundo ECDC** (a qual é uma **Pessoa** a que se refere um **Caso Confirmado COVID-19 Segundo ECDC**) em seus primeiros 14 dias desde a **Manifestação do Sintoma** referente a um **Sintoma COVID-19 Reconhecido Pelo ECDC**; ou (ii) residiu ou trabalhou em uma **Instituição de Tratamento de Pessoas Vulneráveis** localizada em um **Lugar Geográfico** caracterizado por uma **Taxa de Transmissão de COVID-19**.

Uma **Pessoa** satisfaz ao **Critério de Diagnóstico de Imagem Segundo ECDC** se possuir um **Exame Radiológico com Lesões Compatíveis com a COVID-19**. Por último, definimos que a **Pessoa** que satisfaz ao **Critério Laboratorial Segundo ECDC** é aquela que realizou um exame NAAT (*Nucleic Acid Amplifications Tests* ou Teste de Detecção de Ácido Nucleico) cujo resultado foi **Detectável Para SARS-CoV-2** ou realizou um **Teste Antígeno-RDT Para SARS-CoV-2** cujo resultado foi **Teste Positivo**.

Assim, define-se uma **Pessoa Possivelmente com COVID-19 segundo ECDC** como a **Pessoa** que satisfaz ao **Critério Clínico Segundo ECDC**; a **Pessoa Provavelmente com COVID-19 Segundo ECDC** é tanto a **Pessoa** que satisfaz em conjunto ao **Critério Clínico Segundo ECDC** e ao **Critério Epidemiológico Segundo ECDC** quanto à **Pessoa** que satisfaz apenas ao **Critério Laboratorial Segundo ECDC**. Por último, a **Pessoa com COVID-19 Confirmada Segundo ECDC** satisfaz ao **Critério Laboratorial Segundo ECDC**.

De acordo com a OMS (2020), cada subtipo (*Role*) **Caso COVID-19 Segundo OMS** é regido por uma regra diferente. Primeiramente, deve-se definir **Critério Clínico Segundo OMS** e **Critério Epidemiológico Segundo OMS**. Uma **Pessoa** satisfaz ao **Critério Clínico Segundo OMS** se apresenta **Febre e Tosse** ou se apresenta três ou mais **Sintomas de Critério Clínico COVID-19 Reconhecido Pela OMS**. Uma **Pessoa** satisfaz ao **Critério Epidemiológico Segundo OMS** caso satisfaça uma das 3 condições: (i) residiu ou trabalhou em **Estabelecimento** localizado em **Lugar Geográfico** com **Taxa de Transmissão de COVID-19** quantificada como alta; (ii) residiu em **Estabelecimento** localizado em **Lugar Geográfico**, ou viajou para **Lugar Geográfico**, com **Taxa de Transmissão de COVID-19** considerada comunitária; ou (iii) trabalhou em algum **Estabelecimento de Saúde**.

Assim, um **Caso Suspeito de COVID-19 Segundo OMS** refere-se a uma **Pessoa com Suspeita de COVID-19 Segundo OMS** que satisfaz a uma das três condições: (i) satisfaz simultaneamente ao **Critério Clínico Segundo OMS** e ao **Critério Epidemiológico Segundo OMS**; (ii) se encontra acometida de **SRAG** (Síndrome Respiratória Aguda Grave), que é um *subkind* do *mode* **Disfunção**; ou (iii) não se encontra acometida de nenhum **Sintoma de Critério Clínico Segundo OMS** ao mesmo tempo em que realizou um **Teste Antígeno-RDT Para SARS-CoV-2** com resultado **Teste Positivo**.

Um **Caso Provável de COVID-19 Segundo OMS** refere-se a uma **Pessoa com Provável COVID-19 Segundo OMS** que satisfaz a uma das quatro condições: (i) satisfaz ao **Critério Epidemiológico Segundo OMS** e esteve em contato com uma **Pessoa com Provável COVID-19 Segundo OMS** ou uma **Pessoa com COVID-19 Confirmada Segundo OMS**; (ii) realizou **Exame Radiológico com Lesões Compatíveis com a**

COVID-19; (iii) apresentou **Anosmia/Hiposmia/Ageusia Repentina** (*subkind* do *mode Sintoma COVID-19*); (iv) é uma **Pessoa Morta** sem **Motivo** (não explicito no diagrama) explicado para o **Óbito** e, enquanto na sua fase de **Pessoa Viva**, teve contato com **Pessoa com Provável COVID-19 Segundo OMS** ou **Pessoa com COVID-19 Confirmada Segundo OMS** ou esteve presente em um **COVID-19 Cluster**³.

Um **Caso Confirmado de COVID-19 Segundo OMS** refere-se a uma **Pessoa com COVID-19 Confirmada Segundo OMS** que satisfaz a pelo menos uma de três condições: (i) realizou **NAAT** cujo resultado foi **Detectável Para SARS-CoV-2**; (ii) realizou **Teste Antígeno-RDT Para SARS-CoV-2** cujo resultado foi **Teste Positivo** e (é uma **Pessoa com Provável COVID-19 Segundo OMS** ou é uma **Pessoa com COVID-19 Suspeita Segundo OMS** de acordo com a condição (i) ou (ii)); (iii) não apresenta **Sintomas de Critério Clínico COVID-19 Reconhecidos Pela OMS** e teve contato com **Pessoa com Provável COVID-19 Segundo OMS** ou **Pessoa com COVID-19 Confirmada Segundo OMS**.

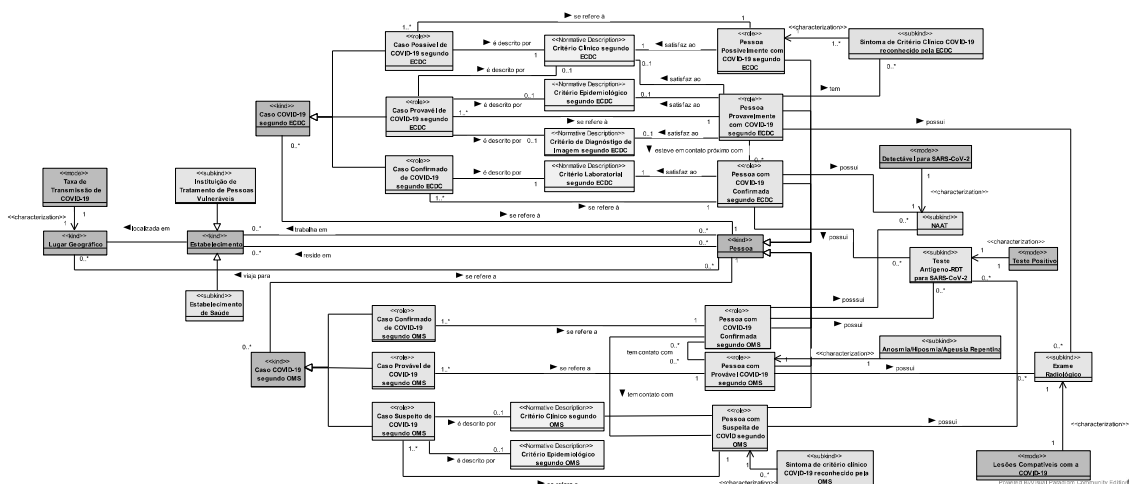


Figura 2. Ontologia OntoCOVID – Caso COVID-19 segundo ECDC e OMS

4. Conclusões

Este trabalho descreveu os primeiros passos da metodologia SABIO no desenvolvimento da OntoCOVID, uma ontologia bem fundamentada do domínio da COVID-19, que contribui para a compreensão semanticamente precisa dos conceitos e para análise adequada dos dados que vêm sendo disponibilizados sobre a pandemia. A OntoCOVID endereça precisamente todas as questões de competência definidas, permitindo distinguir importantes diferenças entre as definições de CASO COVID-19 de duas instituições reguladoras importantes, a OMS e o ECDC. A definição formal dos axiomas da ontologia e a análise do impacto desta distinção ontológica já estão em andamento e serão alvo de trabalhos futuros.

³ Devido a sua complexidade e por limitação de espaço, o conceito COVID-19 Cluster não será tratado.

Referências

- Amaral, G., Baião, F., Guizzardi, G. (2021), “Foundational Ontologies, Ontology-Driven Conceptual Modeling and their Multiple Benefits to Data Mining”, WIREs Data Mining and Knowledge Discovery, Wiley.
- Costa D., Teixeira M., Rissino S., Guarnier T (2020), “O Uso da Abordagem SABiO na Construção do Overview de OntoSaúde”, Engineering Ontologies & Ontologies for Engineering: Celebrating Ricardo Falbo’s Career, ISBN 978-1393963035, p.82-98.
- Guimarães, V. H. A., de Oliveira-Leandro, M., Cassiano, C., Marques, A. L. P., Motta, C., Freitas-Silva, A. L., ... & Oliveira, C. J. F. (2021). Knowledge about COVID-19 in Brazil: cross-sectional web-based study. JMIR Public Health and Surveillance, 7(1), e24756.
- ECDC (2020). “Case definition for coronavirus disease 2019 (COVID-19), as of 3 December 2020”, disponível em <https://www.ecdc.europa.eu/en/covid-19/surveillance/case-definition>, acessado em Ago/2021.
- Falbo, R. (2014). SABiO: Systematic Approach for Building Ontologies. In ONTO.COM/ODISE@ Formal Ontology and Information systems (FOIS).
- Guarino, N., Oberle, D., & Staab, S. (2009). What is an ontology?. In Handbook on ontologies (pp. 1-17). Springer, Berlin, Heidelberg.
- Guarnier T., et al. (2020), “Um Modelo Conceitual Baseado em Ontologia para Doenças Infeciosas com Ênfase em Tuberculose”. Proceedings XIII Seminar on Ontology Research in Brazil (ONTOBRAS 2020), Vitória, Brazil, CEUR-WS 2728.
- Guizzardi G. (2005), “Ontological Foundations for Structural Conceptual Models”, Universal Press.
- Guizzardi G. (2020), “Ontology, ontologies and the ‘I’ of FAIR”, Data Intelligence, MIT Press, Volume 2, Issue 1-2, p.181-191.
- Hartung M., Groß A., Rahm E. (2013), “COnto-Diff: generation of complex evolution mappings for life science ontologies”, Journal of Biomedical Informatics 46 (1), pp. 15-32, ISSN 1532-0464, <https://doi.org/10.1016/j.jbi.2012.04.009>.
- OMS (2011) “Definitions of Key Concepts from the WHO Patient Safety Curriculum Guide”, em: https://www.who.int/patientsafety/education/curriculum/course1a_handout.pdf. Acessado em Ago/2021
- OMS(2020) “WHO COVID-19: Case Definitions”, em: https://www.who.int/publications/i/item/WHO-2019-nCoV-Surveillance_Case_Definition-2020.2. Acessado em Ago/2021
- OMS (2020) “Diagnostic testing for SARS-CoV-2, Interim guidance”
- OMS (2021) “Evaluation of COVID-19 vaccine effectiveness”, em https://www.who.int/publications/i/item/WHO-2019-nCoV-vaccine_effectiveness-measurement-2021.1. Acessado em 10/ago/2021.
- OMS (2021) “WHO Coronavirus (COVID-19) Dashboard”, em: <https://covid19.who.int/>. Acessado em 16/ago/2021.

OMS (2021) “Infectious Diseases”, disponível em: <http://www.emro.who.int/health-topics/infectious-diseases/index.html>. Acessado em 16/Ago/2021.

Tsang T., Wu P., Lin Y., Lau E., Leung G., Cowling B. (2020), “Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study”, *The Lancet Public Health* 5(5), pp. e289-e296, ISSN 2468-2667, [https://doi.org/10.1016/S2468-2667\(20\)30089-X](https://doi.org/10.1016/S2468-2667(20)30089-X).