Design COVID-19 Ontology: A Healthcare and Safety Perspective

Hamid Mcheick[1(✉)], Youmna Nasser[2], Farah Al Wardani[2], and Batoul Msheik[2]

[1] Department of Computer Science and Mathematics, University of Québec at Chicoutimi, Chicoutimi, QC G7H 2B1, Canada
Hamid_mcheick@uqac.ca

[2] Department of Applied Mathematics and Ecole Doctorale, Lebanese University, Beirut, Lebanon
{youmna.nasser,f.alwardani}@st.ul.edu.lb,
batoul.msheik@hotmail.com

Abstract. The COVID-19 pandemic has flooded a vast amount of information into the world. To help control this situation, good utilization of the overflow in data is required. However, data come in different forms, posing numerous challenges in subsequent processing. Therefore, a uniform knowledge representation of COVID-19 information is needed, and ontology can play a role. The ontology will model patient healthcare-related data, ranging from symptoms to side effects and medical conditions, and the necessary precautions, especially for healthcare workers, to obtain protection from the COVID-19 virus. We followed Sánchez's methodology to build the vocabularies, which include current ontology concepts, W3C standards RDF, OWL and SWRL. This work shows promising results that can be applied by different organizations.

Keywords: Design ontology · COVID-19 · Healthcare · Knowledge uniform methodology · Knowledge representation collaboration

1 Introduction

The World Health Organization (WHO) [1] defines COVID-19 as an infectious disease that emerged as a new strain of coronavirus. WHO [1] adds that approximately 405 million people were infected with this virus, killing almost six million from the beginning of the pandemic till February 2022. The scientific community is being tasked with controlling the spread of COVID-19, and this paper is a contributor. We aim to provide a knowledge representation ontology for the collection and analysis of data related to COVID-19 to track relevant data and eventually help control this pandemic.

To respond to this health emergency, information-sharing across different platforms and systems is demanded. However, data can be heterogeneous. Systems collect information based on discipline-specific terminologies, which can restrain them from sharing the information between platforms. Hence, ontologies can offer uniform ways of representing knowledge. Data publishers and collectors can use a shared vocabulary to collect

COVID-19 data. An ontology, as presented by Tudorache et al. [2], defines a common jargon for sharing data amongst researchers and incorporates machine-interpretable definitions of essential concepts and relationships in this field. Knowledge represents the foundation of creating an artificial intelligent model. Dou et al. [29] state how research has shown that the frequent integration of semantics presents improved outcomes in data mining and deep learning.

The ontology presented in this paper provides a knowledge representation model of COVID-19 from a healthcare perspective, demonstrating every patient's case and recommending how people, especially healthcare workers, can protect themselves and follow appropriate safety and protection measures to decrease the risk of contracting the virus.

This article is organized as follows. Section 2 presents an overview of the existing COVID-19 ontologies. We classify our ontology in Sect. 3, and then we build a COVID-19 ontology and describe how it differs from other ontologies in Sect. 4. Section 5 specifies a few reasoning rules applied in this ontology. Finally, conclusions and future works are given in Sect. 6.

2   Literature Review

This global pandemic has given rise to several COVID-19–related ontologies in order to cope and take control. Each ontology models this virus from a different perspective. This section discusses some of these ontologies, shedding light on how our ontology differs.

The COVID-19 Surveillance Ontology, presented by de Lusignan et al. [4], is intended to help surveillance in primary care. The fundamental objective of this ontology is to monitor COVID-19 cases and related respiratory conditions using information from various clinical record frameworks. It was built as a taxonomy with classes including exposure to COVID-19, knowledge related to COVID-19, definite and indefinite contraction of COVID-19. However, this ontology does not comprise any property, hence lessening its semantic expressivity.

Another ontology is CODO, presented by Dutta and DeBellis [5], which is the most similar to our work. This ontology was designed for cases and patient information representation to help in publishing COVID-19 data using Findability, Accessibility, Interoperability, and Reuse (FAIR) standards. It was built to facilitate the organization and illustration of daily-produced COVID-19 data, the relationships between the datasets, and the surrounding factors, for the further analysis of data. Our work differs in that it also covers data related to the safety and protection measures that should be applied.

Many ontologies cover the medical perspective of COVID-19. He et al. [3] presented the well-known CIDO, which is a disease ontology that presents the etiology, transmission, pathogenesis, diagnosis, prevention and treatment of this virus. The information includes the nature of the virus, means of transmission, common symptoms, and medical treatments. In contrast, our approach looks to this disease from a healthcare perspective.

3   Classification of Ontologies

Ontologies can be classified based on content, goals, application technique and timing, domain representation, reusability, and field of application. Many ontology classification methods have been defined over the decades that vary in their perspectives. Ajami and Mcheick [7] expanded on OntoCube and added more performance standards, including machine-readability, i.e., whether the ontology could be easily understood and processed by the computer; reusability of concepts and classes to accomplish an objective; and complexity, i.e., measures of time and resources needed to achieve a certain task. Applying their classification technique, our ontology is considered formal as it employs the Web Ontology Language (OWL). Also, it is domain-specific since it describes terminology in healthcare, particularly COVID-19, allowing it to be semi-reusable. Finally, ours is a heavyweight ontology containing classes and relations, in addition to axioms and rules.

4   Designing a COVID-19 Ontology

Studer et al. [6] define an ontology as a "formal, explicit specification of a shared conceptualization" that is a clearly defined, simple, machine-readable interpretation of real-world concepts and their interrelationships, providing shared knowledge for the target community. In this paper, we adopt this definition of ontology.

This section provides a description of the design and development methodology of COVID-19 ontology. Well-known methods of designing an ontology include METHON-TOLOGY [8], TOVE [9], Cyc 101 [10] and YAMO [11]. We follow Sánchez's methodology [25] for building a medical ontology, also used by Ajami and Mcheick [7], since our ontology focuses on the medical field of the COVID-19 disease. This methodology combines both METHONTOLOGY [8] and Cyc 101 [10] and consists of five main steps that we follow to build our ontology, that are as follows: determine the domain scope, reuse the ontology, develop the conceptual model, implement the ontology, and evaluate it.

4.1   Determine the Domain and Scope of the Ontology

The domain and scope refer to the main field this ontology covers, putting boundaries around the conceptualization, and the ontology's purpose. According to Ajami and Mcheick [7], researchers pose a set of questions to determine the domain of the ontology, which we answer:

- *What is the domain that the ontology will cover?*
    The COVID-19 is the main domain this ontology covers.
- *What is the purpose of this ontology?*
    The purpose of this ontology is to facilitate the gathering and publication of COVID-19–related data as semantic services. Our ontology tracks COVID-19 patient's medical status and predicts their severity level for a better understanding of the nature of the virus and how patients could be treated. It also tries to reduce the risk of contracting COVID-19 by tracking the essential safety and control measures taken by people.

- *Who will use the ontology?*

     This ontology can be used by organizations willing to collect COVID-19–related data to help control this pandemic and know more about this disease, such as hospitals, government agencies, health organizations and researchers.
- *What types of questions should the information in the ontology answer?*

     1) What are the most common symptoms of patients with COVID-19?
     2) How severe is a patient's case?
     3) Are the safety measures effective in protecting against contracting COVID-19?
     4) Is there a relationship between COVID-19 and a certain disease?

4.2   Reuse the Ontology

To the best of our knowledge, the ontological representation of COVID-19 patient clinical healthcare and safety measures is poor since many current COVID-19–related ontologies tackle the disease from a medical standpoint. However, we do integrate some medical terms that are used to represent health data. In our ontology, we integrate concepts from: Schema.org, Friend of a Friend (FOAF), and SNOMED CT.

4.3   Develop a Conceptual Model

The following is a set of guidelines for the development of a conceptual model, as suggested by [7].

- *Enumerate key terms in the ontology.*

     The crucial terms that describe a context need to be defined. These terms include nouns that represent a specific concept (e.g., a patient is described by the noun "Patient"), attributes that describe the type and value of what is being modelled (e.g., the value of temperature is a float), verbs that describe the relationships between nouns or between nouns and attributes (e.g., a patient "is a" Person). Since standard terminology shall be used t model medical terms, we used SNOMED CT ontology in building our ontology to model the concepts of drugs and symptoms.
- *Define classes and class hierarchy.*

     This step starts by defining the classes used in the ontology, then defining the taxonomy of these classes by matching subclasses to classes.
- *Define class properties.*

     The two main class properties are object and datatype properties. These are used to model the relationships among different elements of the ontology, as classes alone do not provide enough information to represent the context behind this ontology. Object properties build relationships between classes by specifying the class domain of the relationship and its class range. The datatype property models the value and type of the concept, such as string, integer, and boolean.
- *Define the facet of slots.*

     According to [7], a slot shall be assigned different kinds of facets that frame its value type, allowed values and cardinalities, to be added as required. They are mostly represented as string, integer, and float in our ontology.

- *Create instances.*

    Individuals of a certain class are created by choosing a class, then filling the value slots. An example is creating Steve as an instance of Patient.
- *Develop our ontology domain.*

    We model our COVID-19 medical ontology with respect to the information and guidelines provided by WHO. It contains information related to a patient with COVID-19, including symptoms and treatment, the patient's medical history and the safety measures to control this virus by healthy people, patients, or healthcare workers. Our ontology consists of four main sub-ontologies, as presented in Fig. 1.

In developing our ontology, we used OWL-DL, a descriptive logic ontology language. The Protégé ontology editor developed at Stanford University with a Pellet reasoner plugin was employed.

We divided our ontology into four sub-ontologies based on the recursive algorithm by Le Pham et al. [15], using Even's algorithm from Amir and McIlraith [16].
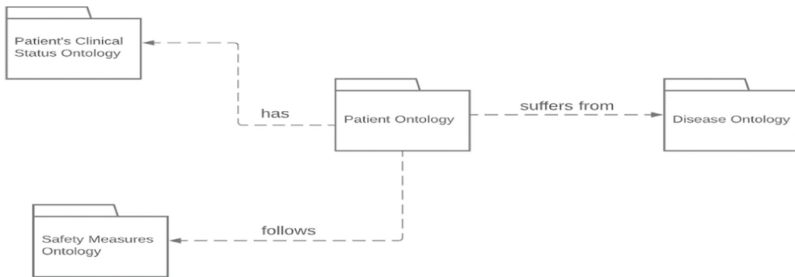

Fig. 1. COVID-19 sub-ontologies

In our ontology, "Patient" is the minimum vertex separator; hence, our ontology is divided based on this. The first subgraph includes patient personal information and physical factors, in addition to the precautions to follow, while the other subgraph accounts for the patient's clinical status in addition to the COVID-19 disease ontology that include the common symptoms and treatments of COVID-19 patients. Each sub-ontology can be further divided into two sub-ontologies following the same reverse algorithm suggested by Le Pham et al. [15]. The first subgraph can be divided through the "Person", which is the minimum vertex separator, thus separating our ontology into two, one dealing with the precautions a person should follow and the second one deals with the physical state of a person. The second ontology can also be divided into two sub-ontologies following the same methodology, with the "Patient" as a minimum vertex separator; one concerns the patient's clinical status and the other with the COVID-19 disease. Descriptions of these ontologies follow.
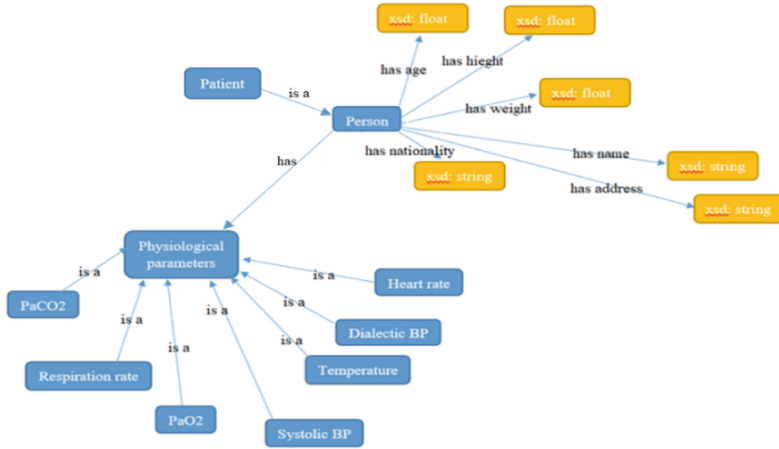
Fig. 2. Part of the patient ontology

Patient Ontology. As shown in Fig. 2, it includes a person's personal information and the relevant physical vital signs that are alerted when a person contracts COVID-19, as suggested by [1, 12, 13]. We use SNOMED CT to represent classes such as the Patient and some physical attributes.
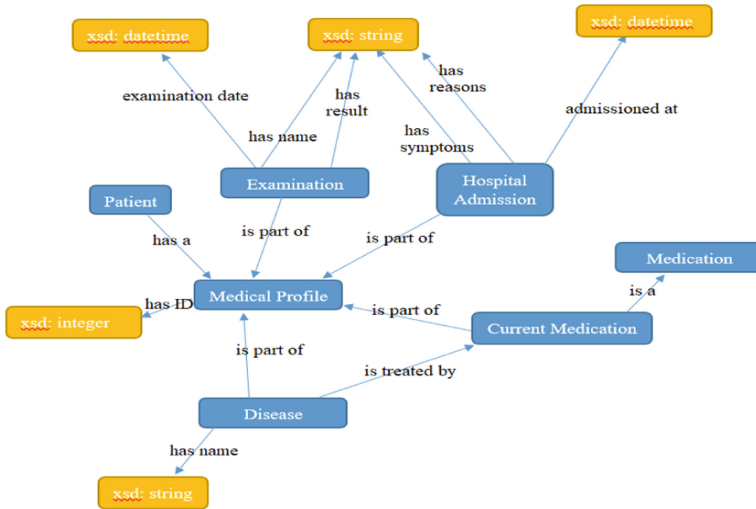

Fig. 3. Part of the patient's clinical status ontology

Patient's Clinical Status Ontology. The clinical status ontology demonstrates a patient's medical history, past diseases, current diseases, hospitalizations, examinations, and current medications, as illustrated in Fig. 3. According to [14], people with some medical conditions may have worse cases of COVID-19 and require more care and attention.

Disease Ontology. The disease ontology mainly represents COVID-19, as shown in Fig. 4, modelling its variant type, symptoms, current stage, severity, potential risk factors, the location where the patient is monitored, and the patient's medication for COVID-19 treatment. SNOMED CT is used to model the symptoms and treatments of a patient with COVID-19.
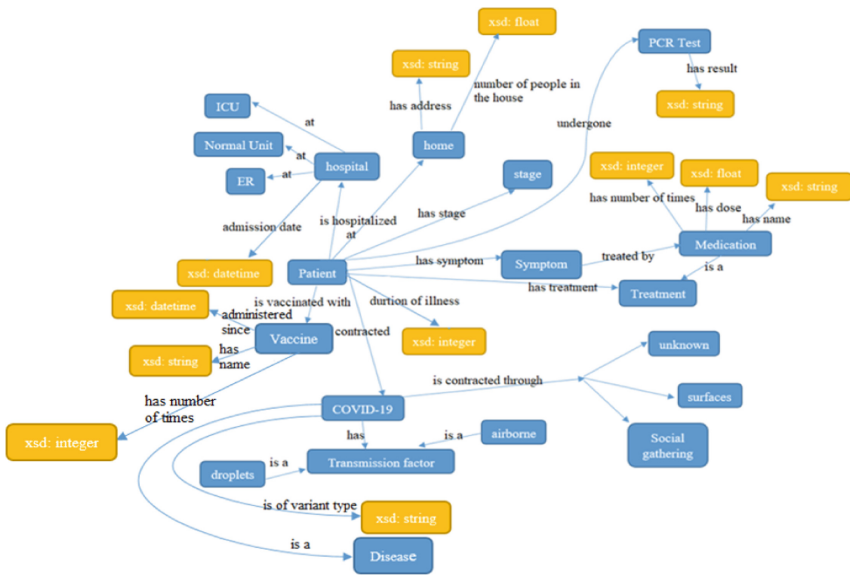

Fig. 4. Part of the disease ontology

Safety Measures Ontology. The precautions ontology, shown in Fig. 5, models the protection mechanisms needed to follow to protect oneself against contracting COVID-19. These precautions are for healthy people, patients and healthcare workers, especially those working on the frontline fighting this pandemic. All these mechanisms are modelled in accordance with WHO's instructions [17, 18].
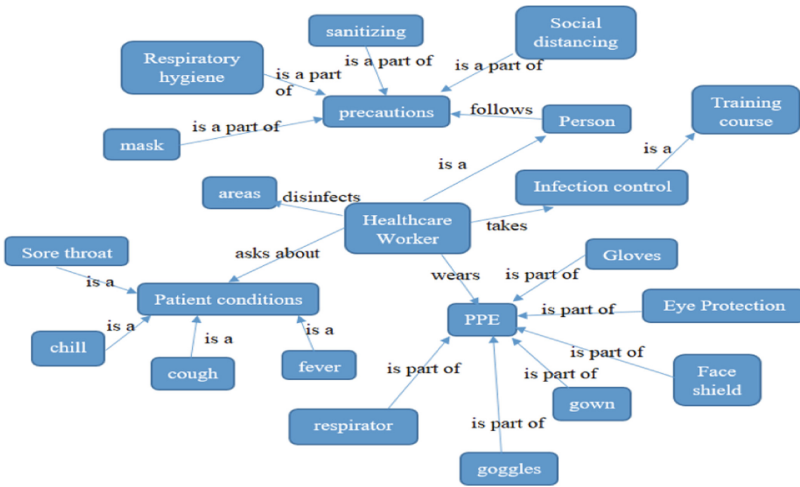
Fig. 5. Part of the safety measures ontology

4.4 Implement the Ontology

As mentioned before, we used the Protégé tool to build our ontology. We formalized our ontology in OWL-DL so that it can be highly expressive, and hence, it enables us to apply appropriate reasoning techniques. We chose the Pellet reasoner for our ontology as it ensures that an ontology does not contain any contradictory facts, checks if any instances of a class are possible, computes the subclass relations between every named class to create the complete class hierarchy and computes the direct types for each of the individuals, as stated by Sirin et al. [24].

4.5 Evaluate the Ontology

Several ontology evaluation measures have been proposed. Evaluation measures the quality of an ontology and if the constraints and requirements have been met. This section covers some of the proposed evaluation metrics and assesses our COVID-19 ontology accordingly.

Yu [19] suggested the following ontology evaluation criteria:

- *Consistency* is achieved when the ontology's set of definitions and axioms have no contradictions between them, according to [27]. Running Pellet reasoner shows that our ontology is consistent and coherent with no conflicting knowledge.
- *Completeness* occurs when the represented knowledge by the ontology covers the domain it represents sufficiently [27]. Our ontology is complete in terms of its purposes and constraints. However, COVID-19 is still a new disease that is continuously studied, and not enough confirmed information exists. Hence, from this aspect, our ontology can be considered incomplete.
- *Conciseness* ensures that the ontology has no redundancy, as stated by [7]. We tried to minimize the number of definitions of our ontology to eliminate redundant ideas while representing the idea fairly. Hence, our ontology is concise.

- *Expandability* measures whether the ontology can be expanded to describe further knowledge without affecting the current, built ontology. Since much knowledge can still be discovered in COVID-19 disease, we built our ontology such that the core concepts are not altered if new knowledge is added, much as how we added the vaccination part to our initial ontology after vaccination data surfaced. Hence, our ontology is expandable.
- *Sensitiveness* indicates if any changes could affect the core of the ontology. As mentioned previously, any alterations or addition of new concepts will not affect our representation as in our classes and axioms; hence, our ontology is considered non-sensitive.

Ontology-Level Evaluation. Srinivasulu et al. [20] set four ontology-level metrics to measure the complication of an ontology's purpose. We compare our ontology to the metrics against the gold metric specified in Ajami and Mcheick [7].

- The *size of vocabulary* (SOV) represents the overall number of classes, properties and individuals in our ontology. In our case, SOV is approximately 300, which is low, thus indicating that our ontology is not significantly large or complex.
- The *edge–node ratio* (ENR) represents the ratio of the number of edges to the number of nodes. The ENR of our ontology is one, hence indicating that the ontology is simple and straight-forward.
- *Tree impurity* (TIP) measures the divergence of the ontology inheritance hierarchy. The TIP of our ontology is approximately 0.5, which means that the inheritance hierarchy of our ontology has not deviated significantly from the rooted tree, implying that our ontology is not complex.
- *The entropy of ontology graph* (EOG) measures the number of structural models. A low EOG denotes more than one structural model, thus a less difficult ontology. Calculated using the formula mentioned in [7], our EOG is almost one, which means that the class structure is fine.

Class-Level Evaluation. Brewster et al. [21] suggested class-level metrics to evaluate the complexity of the ontology, and we use four for our ontology:

- The *number of classes* (NOC) Our NOC is 76, which is relatively good.
- The *number of properties* (NOP) Our NOP is roughly 130, which indicates that the ontology has strong reasoning.
- The *number of root classes* (NORC) Our NORC is 14, indicating that our ontology is diverse.
- *Relationship richness* (RR) measures the overall number of relationships divided by the overall sum of numbers of subclasses and relationships. Our RR is approximately 0.5, indicating the richness of our ontology with COVID-19–related content.

In addition, a dataset from Carbon Health data lab [28] that included cases from people who contracted COVID-19 was used to validate our ontology. We were able to partially test our reasoning rules and have gotten promising results in terms of consistency

and accuracy. For future work, we're working on fully validating our ontology with a complete dataset and with different scenarios.

Overall, the evaluations show that our ontology is adequate and can be reused or expanded.

5   COVID-19 Reasoning Rules

We use the Semantic Web Rule Language (SWRL) to formulate those rules, combined with an OWL knowledge base, which, according to Horrocks et al. [22], extends the OWL abstract syntax to a high level.

5.1   Reasoning in Our Ontology

First Case. Used to determine the severity level of a patient's condition and the place where they shall be monitored in. The National Institutes of Health (NIH) [23] specifies the severity level of a patient diagnosed with COVID-19 based on symptoms, physiological signs, and hospitalization requirements. [23] classifies the severity level of a patient into these four categories: asymptotic, mild, moderate and severe. An asymptotic person doesn't exhibit any symptoms though having a positive polymerase chain reaction (PCR) test. A patient with mild severity level is one that exhibits symptoms but doesn't have shortness of breath, dyspnea, or abnormal chest imaging. A patient with moderate severity level has an oxygen saturation (SpO_2) percentage above or equal to 94%, while a severe case is when the patient's respiratory rate is above 30 breaths/min. Nonetheless, patients with asymptotic or mild severity level shall be monitored at home, while those with moderate and severe cases shall be hospitalized.

Second Case. The second case determines whether the safety measures imposed by the European Centre for Disease Control [26] are being applied, such as wearing a mask, sanitizing, and social distancing. These rules also aid in showing the effect of these measures on the health and lives of people not diagnosed with COVID-19 in terms of contracting the virus, especially healthcare workers who need to be more cautious when working with COVID-19 patients. Our ontology can be used in different frameworks, such as an Internet-of-Things (IoT) system that detects if people are actually following the COVID-19 safety precautions, and evaluating the effectiveness of such precautions.

5.2   Defined Classes

We built axioms in OWL that define the necessary information to declare an individual as a member of a class. Reasoners use these axioms and build the class hierarchy in accordance while adding further reasoning to these individuals. Most of our axioms were built inside the EquivalentTo field in Protégé. Hence, whenever an individual fulfills these conditions, the person is considered an instance of the equivalent class. An example in our case is a patient – being a patient is equivalent to being a person who has undergone a PCR test and received a positive result.

6 Conclusion

All the data generated from the pandemic can be rendered useless if not organized and translated into meaningful information. An ontology is a key element to extract the concepts from emerging COVID-19 data to help control the pandemic. We designed this ontology to model concepts related to patient healthcare data and help with the collection and analysis of the symptoms and effects on patients with COVID-19, in addition to reducing the risk of contracting the virus by setting well-defined precautionary measures. The built ontology was compared against evaluation metrics and shown to be of good quality and ready to be used or expanded. With the onset of vaccinations and new vaccination data emerging daily, future work can include extending our ontology with a detailed vaccine sub-ontology to cover the assessment and side effects of each vaccine. Nonetheless, we are in the works of validating our ontology with a complete dataset that tests all the reasoning rules in to fully prove the ontology's effectiveness in the healthcare and safety domain.

References

1. Coronavirus disease (COVID-19): https://www.who.int/health-topics/coronavirus/corona virus
2. Tudorache, T., Noy, N.F., Tu, S., Musen, M.A.: Supporting Collaborative Ontology Development in Protégé. In: Sheth, A., et al. (eds.) ISWC 2008. LNCS, vol. 5318, pp. 17–32. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88564-1_2
3. He, Y., et al.: CIDO, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis. Sci. Data **7**(1), 1–5 (2020)
4. de Lusignan, S., et al.: COVID-19 surveillance in a primary care sentinel network: in-pandemic development of an application ontology. JMIR Public Health Surveill. **6**(4), e21434 (2020)
5. Dutta, B., DeBellis, M.: CODO: an ontology for collection and analysis of COVID-19 data. In: 12th International Conference on Knowledge Engineering and Ontology Development (KEOD) (2020)
6. Studer, R., Benjamins, V.R., Fensel, D.: Knowledge engineering: principles and methods. Data Knowl. Eng. **25**(1–2), 161–197 (1998)
7. Ajami, H., Mcheick, H.: Ontology-based model to support ubiquitous healthcare systems for COPD patients. Electronics **7**, 371 (2018)
8. Fernández-López, M., Gómez-Pérez, A., Juristo, N.: METHONTOLOGY: from ontological art towards ontological engineering. In: Spring Symposium on Ontological Engineering of AAAI, pp. 33–40. Stanford University, California (1997)
9. Grüninger, M., Fox, M.: Methodology for the design and evaluation of ontologies. In: Proceedings of the IJCAI Workshop on Basic Ontological Issues in Knowledge Sharing, Menlo Park, CA, USA. AAAI Press, Menlo Park (1995)
10. Lenat, D., Guha, R.: Building Large Knowledge-based Systems: Representation and Inference in the cyc Project. Addison-Wesley, Boston MA USA (1999)
11. Dutta, B., Chatterjee, U., Madalli, D.P.: YAMO: yet another methodology for large-scale faceted ontology construction. J. Knowl. Manag. **19**(1), 6–24 (2015)
12. Ran, J., et al.: Blood pressure control and adverse outcomes of COVID-19 infection in patients with concomitant hypertension in Wuhan, China. Hypertension Res. **43**(11), 1267–1276 (2020)

13. Lee Lewis, D.: COVID-19 and the heart: what have we learned? (2021). https://www.health.harvard.edu/blog/covid-19-and-the-heart-what-have-we-learned-2021010621603

14. People with Certain Medical Conditions, Centers for Disease Control and Prevention CDC, Saving Lives, Protecting People (2022). https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html [Last updated 29 April 2022]

15. Le Pham, T.A., Le-Thanh, N., Sander, P.: Some approaches of ontology decomposition in description logics. In: Loureiro G., Curran, R. (eds.) Complex Systems Concurrent Engineering. Springer, London (2007)

16. Amir, E., McIlraith, S.A.: Partition-based logical reasoning for first-order and propositional theories. Artif. Intell. **162**(1–2), 49–88 (2005)

17. World Health Organization: Prevention, identification and management of health worker infection in the context of COVID-19 (2020)

18. World Health Organization: Infection prevention and control and preparedness for COVID-19 in healthcare settings (2020)

19. Yu, J.: Requirements-oriented methodology for evaluating ontologies. Ph.D. Thesis. RMIT University, Melbourne, Victoria, Australia (2008)

20. Srinivasulu, S., Sakthivel, P., Balamurugan, E.: Measuring the ontology level and class level complexity metrics in the semantic web. Int. J. Adv. Comput. Eng. Netw. **2**, 68–74 (2014)

21. Brewster, C., Alani, H., Dasmahapatra, S., Wilks, Y.: Data-driven ontology evaluation. In: Proceedings of LREC (2004)

22. Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B., Dean, M.: SWRL: a semantic web rule language combining OWL and RuleML. W3C Member Submission (2004). http://www.w3.org/Submission/SWRL/

23. Clinical Spectrum of SARS-CoV-2 Infection: https://www.covid19treatmentguidelines.nih.gov/overview/clinical-spectrum/

24. Sirin, E., Parsia, B., Cuenca Grau, B., Kalyanpur, A., Katz, Y.: Pellet: a practical OWL-DL reasoner. J. Web Semant. **5**(2), 51–53 (2007)

25. Sánchez, D., Batet, M.: Semantic similarity estimation in the biomedical domain: an ontology-based information-theoretic perspective. J. Biomed. Inform. **44**, 749–759 (2011)

26. European Centre for Disease Prevention and Control: Infection prevention and control and preparedness for COVID-19 in healthcare settings. Technical report, Fifth update (2020)

27. Raad, J., Cruz, C.: A survey on ontology evaluation methods. In: Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, pp. 179–186 (2015)

28. Carbon Health and Braid Health: Coronavirus Disease 2019 (COVID-19) Clinical Data Repository (2020). https://covidclinicaldata.org/

29. Dou, D., Wang, H., Liu, H.: Semantic data mining: a survey of ontology-based approaches. In: IEEE ICSC, pp. 244–251. IEEE (2015)