2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-1179-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/EMBC46164.2021.9629555

CIDO-COVID-19: An Ontology for COVID-19 Based on CIDO

Yu Xiao†, Xiangwen Zheng†, Wei Song, Fan Tong, Yiqing Mao, Sheng Liu, Dongsheng Zhao

***Abstract*— To realize integration, organization and reusability of knowledge related to COVID-19, an ontology for COVID-19 (CIDO-COVID-19) was constructed which extended the Coronavirus Infectious Disease Ontology (CIDO) by adding terms of COVID-19 related to symptoms, prevention, drugs and clinical domains. First, terms from the existing ontologies, literature, clinical guidelines and other resources about COVID-19 were merged. Then, the Stanford seven-step approach was used to define and organize the acquired terms. Finally, the CIDO-COVID-19 was built on basis of the terms mentioned above using Protégé. The CIDO-COVID-19 is a more comprehensive ontology for COVID-19, covering multiple areas in the domain of COVID-19, including disease, diagnosis, etiology, virus, transmission, symptom, treatment, drug and prevention.**

***Keywords*— COVID-19; Ontology; CIDO; OBO Foundry**

***Clinical Relevance*— The CIDO-COVID-19 covers multiple areas related to COVID-19, including diseases, diagnosis, etiology, virus, transmission, symptoms, treatment, drugs, prevention. Compared with the CIDO, it is expanded to cover drugs, prevention, and clinical domain. The definition of terms in CIDO-COVID-19 refers to biomedical ontologies, Clinical glossaries and clinical guidelines for COVID-19, which can provide clinicians with standard terminology in the clinical domain.**

I. Introduction

The pandemic of COVID-19 has taken a heavy toll on mankind. There has been increasing research on COVID-19, so that data and concepts related to COVID-19 have surged, and a concept contains multiple terms. Facing the wide range of concepts in the field of COVID-19 and the relationships between concepts, it is urgent to organize concepts and relationships in the field of COVID-19 in an orderly manner and ensure the interoperability of terms. An ontology is a tool that can handle related concepts and relationships in a field [1]. In information science, ontology is a clear specification of the conceptual model [2]. Applying an ontology to COVID-19 can organize its domain knowledge in an orderly manner, form a domain knowledge system, and provide knowledge support for health care providers and researchers. Ontologies have a wide range of application scenarios. Literature [3] verifies the feasibility of application of ontologies in the field of clinical decision support, and literature [4] demonstrates the feasibility of SNOMED-CT, NCIT and other ontologies in natural language processing and text mining.

Several ontologies related to COVID-19 have been constructed. The Coronavirus Infectious Disease Ontology (CIDO) [5] is an open-source biomedical ontology relating to coronavirus infectious diseases. It is intended to provide standardized annotations and representations for various coronavirus infectious diseases. CIDO mainly focuses on common terms of the coronavirus category, and it can be applied to discovery of coronavirus pathogenic factors and development of therapeutic drugs. COVID-19 Ontology [6] is a domain ontology for COVID-19, which mainly describes the role of molecules and cells in the virus-host interaction and virus life cycle. It aims to provide support for drug development and repurposing of COVID-19. COVID-19 Infections Disease Ontology (IDO-COVID-19) [7] is an extension of infectious disease ontology (IDO) and virus infectious disease ontology (VIDO), focusing on epidemiology, classification and pathogenesis of diseases. IDO-COVID-19 plays an important role in the research on COVID-19 at the disease level. Although the above ontologies have organized COVID-19 terms respectively from their own perspective, there is still a lack of comprehensive organization of COVID-19 terms. Besides, there is still room for improvement in the field of symptoms, drugs and preventions for COVID-19, so further supplementation is still needed.

We built CIDO-COVID-19 based on existing ontologies and other resources, focusing on the expansion of the terms of prevention, symptoms, drugs, and clinical domains compared with CIDO. The rest of this paper was organized as follows: Chapter 2 introduces the materials and methods of ontology construction; Chapter 3 introduces the terms, organization, and relationships in CIDO-COVID-19; Chapter 4 summarizes the work in this paper.

II. Material and Methods

A. Methods

The methods used are based on the existing mature ontology construction specifications, taking the OBO Foundry guidelines as the development principle, and the Basic Formal Ontology (BFO) as the top-level ontology, to construct an open, well-expressed, and verifiable ontology for COVID-19 to integrate and reuse COVID-19 terms. OBO Foundry is an open-use biomedical controlled vocabulary and ontology-based collaboration network supported by the National Institutes of Health [8]. Most of ontologies included in OBO follow the common ontology development principle, using Basic Formal Ontology (BFO) as the top-level ontology [9]. By providing a common top-level architecture, BFO regulates the interoperability among domain ontologies on the top-level structure.

† These authors contributed equally to this work.

Yu Xiao (xiaoyuchn@foxmail.com), Xiangwen Zheng (zhengxw@bmi.ac.cn), Fan Tong (tongf@bmi.ac.cn) and Dongsheng Zhao (dszhao@bmi.ac.cn) are with the Information Center, Academy of Military Medical Sciences, Beijing, China, and Dongsheng Zhao is the corresponding author.

Wei Song (songwei@medpeer.cn), Yiqing Mao (polo_simon@163.com) and Sheng Liu (liusheng@medpeer.cn) are with the Beijing MedPeer Information Technology Co., Ltd., Beijing, China.

We used the seven-step approach [10] to construct CIDO-COVID-19. Firstly, we determined the concept coverage of CIDO-COVID-19, and then retrieved, filtered, and reused terms from existing ontologies, including classes, relations, and instances. Next, the logical relationships between these concepts were further defined. Based on literature investigation, the scope of the core concepts for COVID-19 was determined, including disease, diagnosis, virus, etiology, transmission, symptoms, treatment, drugs, and prevention. Aiming at the above scope, CIDO-COVID-19 gave priority to the reuse of terms in existing ontologies. We reused the terms which were of high quality and widely used under the circumstance of a term appearing in different ontologies. Correspondence between reused terms and ontologies is shown in TABLE I. For terms not covered in existing ontologies, CIDO-COVID-19 took advantage of the Aristotelian form [9] to define these terms and added them to our ontology. For example, the term of COVID-19 preventive intervention was defined as: *COVID-19 preventive intervention=def. a preventive intervention that can be used to prevent COVID-19.*

As for relationships between concepts, the strategy is to prioritize the reuse of defined relationships in reused ontologies and Relation Ontology (RO) [11]. For relationships that did not exist in resources above, we used Protégé to define them.

For instances of a concept, CIDO-COVID-19 referred to the current authoritative resources in the clinical domain of COVID-19, such as Diagnosis and treatment of novel coronavirus pneumonia (trial version 8) of China [12], BMJ [13] and DrugBank [14], from which we obtained specific COVID-19 treatments and drugs as instances.

B. Tools

We chose Protégé [15] as the tool to build CIDO-COVID-19. Protégé is currently a relatively mature and widely-used tool in the field of building biomedical ontologies. It can support the reuse of ontologies, and provides functions such as reasoning and visual interface.

We used Ontofox for terms acquisition. Ontofox [16] is a web-based tool to obtain terms and axioms in ontologies, which supports the reuse of ontology. We reused the classes, attributes, and annotations in ontologies using Ontofox.

The reasoner can assess the consistency of the ontology. Commonly used reasoners are Pellet (*http//clarkparsia.com/pellet/*) and Hermit Reasoner (*http://hermit-reasoner.com/*) [9], which can be used in Protégé. The reasoner we chose is Pellet.

III. Results

Currently, CIDO-COVID-19 covers the terms in the field of disease, diagnosis, etiology, virus, transmission, symptom, treatment, drug, and prevention related to COVID-19, contains more than 8000 classes, 356 relationship types and 448 instances, and reuses more than 15 ontologies. Reasoner verification results showed that CIDO-COVID-19 had good consistency. Fig. 1 shows the class structure diagram of the first five levels of the class hierarchy in CIDO-COVID-19. The top-level terms in the figure come from BFO such as entity, continuant, occurrent, etc. The bottom-level terms are also further divided into categories in CIDO-COVID-19. Fig. 2 shows the hierarchical structure of diagnosis.

TABLE I. THE SCOPE OF CORE CONCEPTS AND THE DOMAIN ONTOLOGY OF REUSE

Concepts	Core Concepts	Reused Ontologies
Disease	disease	OGMS [17]
Diagnosis	diagnosis	OGMS
	diagnostic process	OGMS
Virus	viruses	NCBI Taxonomy [18]
Pathogen	pathogen	IDO [19]
Transmission	transmission process	TRANS
Symptom	symptom	SYMP
Treatment	treatment	OGMS
Drug	drug product	DRON [20]
	drug substance	CIDO
	pharmaceutical preparations	NDF-RT [21]
	drug role	CHEBI [22]
	pharmacology	NCIT [23]
	drug product therapeutic function	DRON
	adverse drug effect	OAE [24]
	drug pathway	PW [25]
Prevention	preventive intervention	NCIT
	vaccine	VO [26]
Others	gene	SO [27]
	protein	PR [28]
	host	IDO

TABLE II. THE NUMBER OF NEW CLASSES AND INSTANCES

Core Concepts	Number of Increased Class	Number of Increased Instance
Disease	4	0
Pathogen	0	1
Symptom	847	0
Treatment	0	16
Drug Product	3	11
Drug Substance	0	23
Pharmacology	39	0
Adverse Drug Effect	1	0
Drug Pathway	834	0
Preventive Intervention	1	14
Host	0	5

Compared with CIDO, it includes more than 2,000 new terms in prevention, symptoms, drugs, clinical domains. The numbers of added classes and instances are shown in TABLE II.

For symptoms, we organize terms according to the human system, which covers all possible symptoms of COVID-19, including head and neck symptoms, blood and immune system symptoms, respiratory and chest symptoms. For diagnosis, four clinical types of COVID-19 have been added. According to symptoms of the patient diagnosed with COVID-19, a patient can be classified into 4 categories: mild, moderate, severe, and critical. For drugs and treatments, CIDO-COVID-19 includes 34 new instances of drug substance and drug product, including COVID-19 human immunoglobulin, Glucocorticoid, Tocilizumab. Sixteen treatment methods have been added, including general treatments such as Oxygen nasal cannula, and treatments for severe and critically patients such as Extracorporeal Membrane Oxygenation (ECMO). Meanwhile, CIDO-COVID-19 has also expanded the terms of pathways, pharmacology, adverse effect, and preventive intervention for COVID-19.

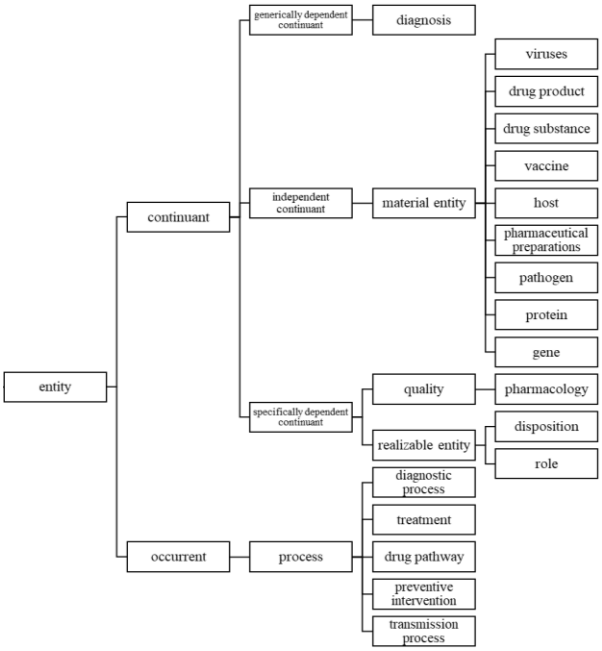

Figure 1. The first five levels of CIDO-COVID-19 hierarchy

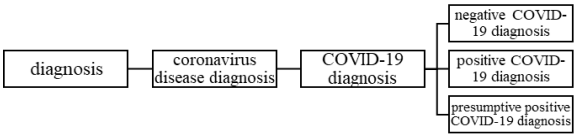

Figure 2. Diagnosis hierarchy

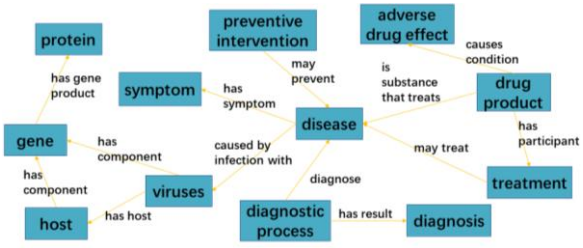

Figure 3. Schematic diagram of the relationships between concepts

Thirteen new relationships have been added in CIDO-COVID-19 compared with CIDO, including two newly defined relationships, and eleven reused relationships. CIDO-COVID-19 has associated core concepts with newly added relations. The relationships between concepts are shown in Fig. 3. There are many types of relationships between concepts, the above content ensures semantic consistency of relationships between concepts and facilitates definitions of logical axioms and reasoning.

CIDO-COVID-19 not only supports the expression of hierarchical structure between concepts but also demonstrates logical relationships between concepts. For example, a logical axiom has been defined by the relationship *has symptom*, which connects COVID-19 with its clinical categories:

critical COVID-19 infection：

COVID-19

and ('has symptom' some shock)

and ('has symptom' some respiratory failure)

and ('has symptom' some multiple organ failure)

According to the definition of critical COVID-19 infection, when a patient diagnosed with COVID-19 has symptoms of shock, respiratory failure, and multiple organ failure, his/her clinical category is critical. The definitions of above axioms will allow computers to infer the clinical category of patients with COVID-19. Such axioms expand the reasoning ability of CIDO-COVID-19.

The ontology is a kind of prior knowledge, and can guide construction of knowledge graph from top to bottom. We defined the schema of a COVID-19 knowledge graph based on CIDO-COVID-19, which stipulated types of entities and relationships in the knowledge graph with biomedical significances, and further built a knowledge graph for COVID-19 *(http://covid19.medpeer.cn/home/)*.

IV. Discussion and Conclusion

Based on OBO Foundry guidelines and the guidance of seven-step approach [10], targeting at the field of COVID-19, especially clinical domain, we refined and expanded CIDO to build an open-source and well-expressed ontology for COVID-19, which covers terms of diseases, diagnosis, etiology, virus, transmission, symptoms, treatment, drugs, prevention in COVID-19. CIDO-COVID-19 uses BFO as the top-level ontology, reuses over 15 biomedical ontologies widely used in OBO Foundry, and formally expresses related concepts into a standard ontology presentation using Protégé software. CIDO-COVID-19 is characterized by wide concept coverage and data interoperability.

CIDO-COVID-19 is optimized and supplemented in the following aspects:

- In terms of building specifications, we have conducted in-depth research on the applicable rules of BFO, and structurally optimized the irregularities of CIDO.
- In terms of coverage of concepts, current research on COVID-19 is more targeted at diagnosis and treatment. However, CIDO has some deficiencies in terms of

diagnosis, treatment, and drugs. Thus, CIDO-COVID 19 mainly expands terms of diseases, symptoms, diagnosis, treatment, and drugs based on CIDO. CIDO-COVID-19 also adds relationships and instances.

- In terms of clinical applicability, when constructing CIDO-COVID-19, we referred to resources such as Diagnosis and treatment of novel coronavirus pneumonia (trial version 8) [12] and SNOMED-CT [29], etc., from which terms of diagnosis, treatment and drugs were obtained to make CIDO-COVID-19 clinically applicable.

CIDO-COVID-19 has a variety of application scenarios. In addition to the guidance to build knowledge graphs, CIDO-COVID-19 can also be used to support reasoning process, the COVID-19 literature mining and clinical decision.

At present, CIDO-COVID-19 still has the following limitations: Firstly, some concepts can be further expanded, such as vaccines, traditional Chinese medicine treatment, and gene-protein interactions. Secondly, CIDO-COVID-19 is an ontology for a subdivided domain, the number of instances directly related to COVID-19 is not enough. We will follow up the researches and enrich relevant instances in time.

CIDO-COVID-19 is open-source, and can be accessed at https://github.com/xiaoyuchn/CIDO-COVID-19. Your valuable comments are welcomed for us to improve it.

ACKNOWLEDGMENT

Our gratitude goes to the developers of CIDO, COVID-19 Ontology, IDO-COVID-19, and all the ontologies mentioned in this article. Their excellent work and the public resources enable us to engage in this research.

Data involved in this paper was open-source, and there was no involvement of any procedure on human subjects or animals.

REFERENCES

[1] L. M. Serra, W. D. Duncan, and A. D. J. B. b. Diehl, "An ontology for representing hematologic malignancies: the cancer cell ontology," vol. 20, no. 5, pp. 231-236, 2019.

[2] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition,* vol. 5, no. 2, pp. 199-220, 1993.

[3] Y. Shen, J. Colloc, A. Jacquet-Andrieu, and K. Lei, "Emerging medical informatics with case-based reasoning for aiding clinical decision in multi-agent system," *Journal of biomedical informatics,* vol. 56, pp. 307-317, 2015.

[4] I. Spasić, J. Livsey, J. A. Keane, and G. Nenadić, "Text mining of cancer-related information: review of current status and future directions," *International journal of medical informatics,* vol. 83, no. 9, pp. 605-623, 2014.

[5] Y. He, H. Yu, E. Ong, Y. Wang, Y. Liu, A. Huffman, H.-h. Huang, J. Beverley, J. Hur, and X. Yang, "CIDO, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis," *Scientific Data,* vol. 7, no. 1, pp. 1-5, 2020.

[6] A. Sargsyan, A. T. Kodamullil, S. Baksi, J. Darms, S. Madan, S. Gebel, O. Keminer, G. M. Jose, H. Balabin, L. N. DeLong, M. Kohler, M. Jacobs, and M. Hofmann-Apitius, "The COVID-19 Ontology," *Bioinformatics,* vol. 36, no. 24, pp. 5703–5705, Dec 15, 2020.

[7] S. Babcock, L. G. Cowell, J. Beverley, and B. Smith, "The Infectious Disease Ontology in the Age of COVID-19," *OSF Preprints*: Center for Open Science, 2020.

[8] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, and C. J. Mungall, "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration," *Nature biotechnology,* vol. 25, no. 11, pp. 1251-1255, 2007.

[9] R. Arp, B. Smith, and A. D. Spear, *Building ontologies with basic formal ontology*: Mit Press, 2015.

[10] N. F. Noy, and D. L. McGuinness, "Ontology development 101: A guide to creating your first ontology," Stanford knowledge systems laboratory technical report KSL-01-05 and …, 2001.

[11] B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A. L. Rector, and C. Rosse, "Relations in biomedical ontologies," *Genome biology,* vol. 6, no. 5, pp. 1-15, 2005.

[12] O. o. N. A. o. T. C. M. General Office of National Health Commission of the People's Republic of China, "Diagnosis and treatment of novel coronavirus pneumonia (trial version 8）," 2020.

[13] N. Beeching, and R. Fowler, "Coronavirus disease 2019 (COVID-19). BMJ Best Practice 2020," 2020.

[14] D. S. Wishart, C. Knox, A. C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam, and M. Hassanali, "DrugBank: a knowledgebase for drugs, drug actions and drug targets," *Nucleic acids research,* vol. 36, no. suppl_1, pp. D901-D906, 2008.

[15] J. H. Gennari, M. A. Musen, R. W. Fergerson, W. E. Grosso, M. Crubézy, H. Eriksson, N. F. Noy, and S. W. J. I. J. o. H.-c. s. Tu, "The evolution of Protégé: an environment for knowledge-based systems development," vol. 58, no. 1, pp. 89-123, 2003.

[16] Z. Xiang, M. Courtot, R. R. Brinkman, A. Ruttenberg, and Y. He, "OntoFox: web-based support for ontology reuse," *BMC research notes,* vol. 3, no. 1, pp. 1-12, 2010.

[17] W. Ceusters, and B. Smith, "Biomarkers in the ontology for general medical science." pp. 155-159.

[18] S. J. N. a. r. Federhen, "The NCBI taxonomy database," *Nucleic Acids Research,* vol. 40, no. D1, pp. D136-D143, 2012.

[19] L. M. Schriml, C. Arze, S. Nadendla, Y.-W. W. Chang, M. Mazaitis, V. Felix, G. Feng, and W. A. J. N. a. r. Kibbe, "Disease Ontology: a backbone for disease semantic integration," vol. 40, no. D1, pp. D940-D946, 2012.

[20] J. Hanna, E. Joseph, M. Brochhausen, and W. R. J. J. o. b. s. Hogan, "Building a drug ontology based on RxNorm and other sources," vol. 4, no. 1, pp. 1-9, 2013.

[21] J. Pathak, and C. G. J. J. o. t. A. M. I. A. Chute, "Analyzing categorical information in two publicly available drug terminologies: RxNorm and NDF-RT," vol. 17, no. 4, pp. 432-439, 2010.

[22] K. Degtyarenko, P. De Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcántara, M. Darsow, M. Guedj, and M. J. N. a. r. Ashburner, "ChEBI: a database and ontology for chemical entities of biological interest," vol. 36, no. suppl_1, pp. D344-D350, 2007.

[23] S. de Coronado, L. W. Wright, G. Fragoso, M. W. Haber, E. A. Hahn-Dantona, F. W. Hartel, S. L. Quan, T. Safran, N. Thomas, and L. J. J. o. b. i. Whiteman, "The NCI Thesaurus quality assurance life cycle," vol. 42, no. 3, pp. 530-539, 2009.

[24] Y. He, S. Sarntivijai, Y. Lin, Z. Xiang, A. Guo, S. Zhang, D. Jagannathan, L. Toldo, C. Tao, and B. J. J. o. b. s. Smith, "OAE: the ontology of adverse events," vol. 5, no. 1, pp. 1-13, 2014.

[25] V. Petri, P. Jayaraman, M. Tutaj, G. T. Hayman, J. R. Smith, J. De Pons, S. J. Laulederkind, T. F. Lowry, R. Nigam, and S.-J. J. J. o. b. s. Wang, "The pathway ontology–updates and applications," vol. 5, no. 1, pp. 1-12, 2014.

[26] Y. He, L. Cowell, A. Diehl, H. Mobley, B. Peters, A. Ruttenberg, R. Scheuermann, R. Brinkman, M. Courtot, C. Mungall, Z. Xiang, F. Chen, T. Todd, L. Colby, H. Rush, T. Whetzel, M. Musen, B. Athey, G. Omenn, and B. Smith, "VO: Vaccine Ontology," *Nature Precedings*, Aug 5, 2009.

[27] K. Eilbeck, S. E. Lewis, C. J. Mungall, M. Yandell, L. Stein, R. Durbin, and M. J. G. b. Ashburner, "The Sequence Ontology: a tool for the unification of genome annotations," vol. 6, no. 5, pp. 1-12, 2005.

[28] D. A. Natale, C. N. Arighi, W. C. Barker, J. A. Blake, C. J. Bult, M. Caudy, H. J. Drabkin, P. D'Eustachio, A. V. Evsikov, and H. J. N. a. r. Huang, "The Protein Ontology: a structured representation of protein forms and complexes," vol. 39, no. suppl_1, pp. D539-D545, 2010.

[29] J. Millar, "The Need for a Global Language - SNOMED CT Introduction," *Studies in Health Technology & Informatics,* vol. 225, pp. 683, 2016.