CIDO: The Community-based Coronavirus Infectious Disease Ontology

Yongqun "Oliver" He

University of Michigan Medical School

Ann Arbor, MI 48109 USA

Acknowledgements

Collaborators & co-authors:

University of Michigan:

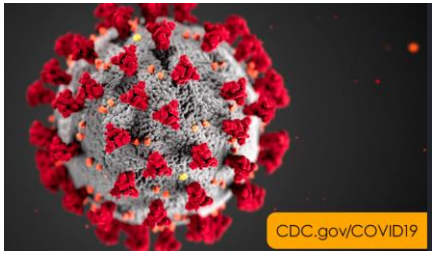
- Yongqun “Oliver” He
- Edison Ong
- Yang Wang (also Guizhou Medical School)
- Yingtong Liu
- Anthony Huffman
- Hsin-Hui Huang (also National Yang-Ming University, Taiwan)
- Gil Omenn
- Brian Athey

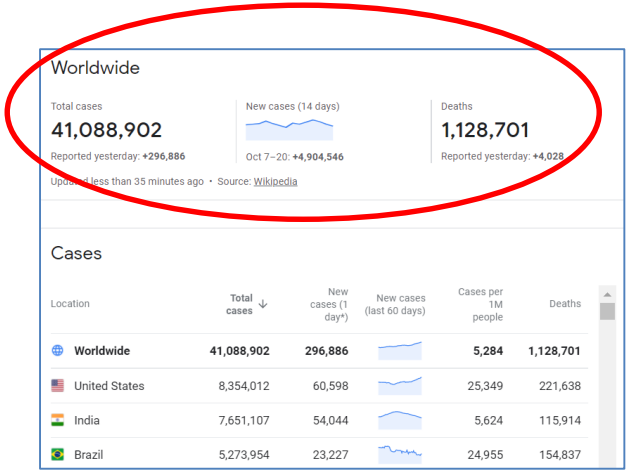
- Hong Yu (Guizhou Medical School)
- John Beverley (Northwestern University)
- Asiyah Yu Lin (National Center for Ontological Research)
- William D. Duncan (Lawrence Berkeley National Lab)
- Sivaram Arabandi (OntoPro LLC)
- Liwei Wang, Hongfang Liu (Mayo Clinic)
- Yehoshua Perl（New Jersey Institute of Technology）
- Fatima Z. Smaili, Robert Hoehndorf (King Abdullah Univ.)
- Jiangan Xie (Chongqing Univ. of Posts and Telecommun.)
- Junguk Hur (Univ. of North Dakota)
- Xianglin Yang (Chinese Academy of Medical Sciences)
- Luonan Chen (Shanghai Institutes for Biological Sciences)
- Barry Smith (Univ. at Buffalo)

Funding: U24CA210967 and P30ES017885 (to GSO); R01GM080646, 1UL1TR001412, 1U24CA199374, and 1T15LM012495 (to BS); the National Natural Science Foundation of China 61801067 (to JX); the Natural Science Foundation of Chongqing CSTC2018JCYJAX0243 (to JX); the non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences 2019PT320003 (to HY); and University of Michigan Medical School Global Reach award (to YH).

COVID-19: Coronavirus Disease 2019

- Caused by SARS-CoV-2 coronavirus
- First found in Wuhan, China, in December 2019
- COVID-19 Pandemic
  - Declared by WHO on March 11
  - At the time: 118,326 confirmed cases and 4,292 deaths globally
  - As of 10/21/2020: >41 mill. cases, >1.1 mill deaths





As of 10/21/2020

Other Coronavirus Diseases

- **SARS**: Severe Acute Respiratory Syndrome
  - Emerged in China in Nov. 2002, lasted 8 months
  - 8,098 confirmed cases in 29 countries, 774 deaths
  - Case-fatality rate: 9.6%

- **MERS:** Middle East Respiratory Syndrome
  - Emerged Saudi Arabia in June 2012
  - 2,260 confirmed cases in 27 countries, 803 deaths
  - Case-fatality rate: 35.5%

- Many other coronaviruses cause mild disease in humans, similar to the common cold

Investigating Coronavirus Diseases

- Coronavirus diseases share common features, but differ in many significant ways
- Researchers have made impressive progresses investigating coronavirus structures, pathogenesis to associated diseases, and transmission
- Well-designed ontology representation is a crucial ingredient in computer-aided coronavirus investigations

CIDO: Coronavirus Infectious Disease Ontology

- Community Effort:
  - https://github.com/CIDO-ontology/cido
  - Accepted into the Open Biological and Biomedical Ontology (OBO) Foundry library

- Integrates coronavirus data concerning:
  - Coronaviruses (etiology); Hosts (phenotypes); Reservoirs (Transmission); Host-Coronavirus Interactions;
  - Diagnosis; Drugs; Vaccines;
  - Metadata; Data Standardization

Ref: He Y, Yu H, Ong E, Wang Y, Liu Y, Huffman A, Huang H, Beverley J, Hur J, Yang X, Chen L, Omenn GS, Athey B, Smith B. **CIDO**, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis. ***Scientific Data***. (2020) 7:181.

CIDO Statistics

Version: 1.0.159:

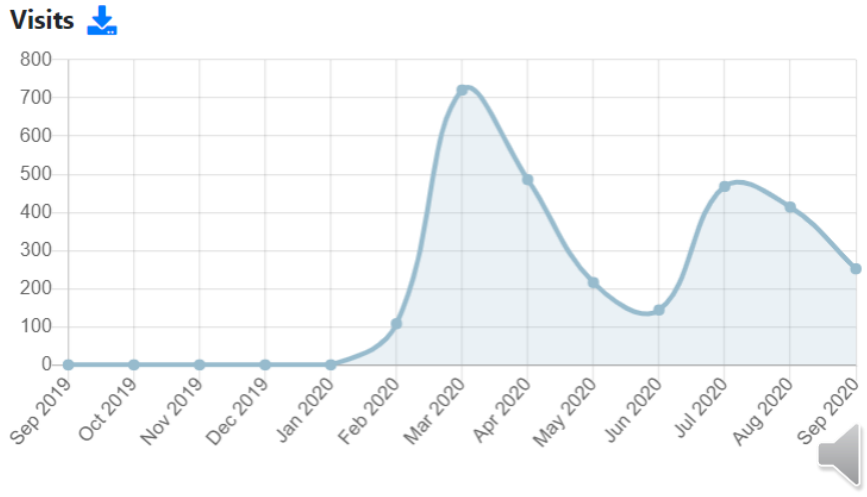
- Class (6556)
- ObjectProperty (345)
- DatatypeProperty (18)
- AnnotationProperty (136)
- Instance (446)

- Importing from 48 ontologies
- 244 CIDO-specific terms

http://www.ontobee.org/ontology/CIDO

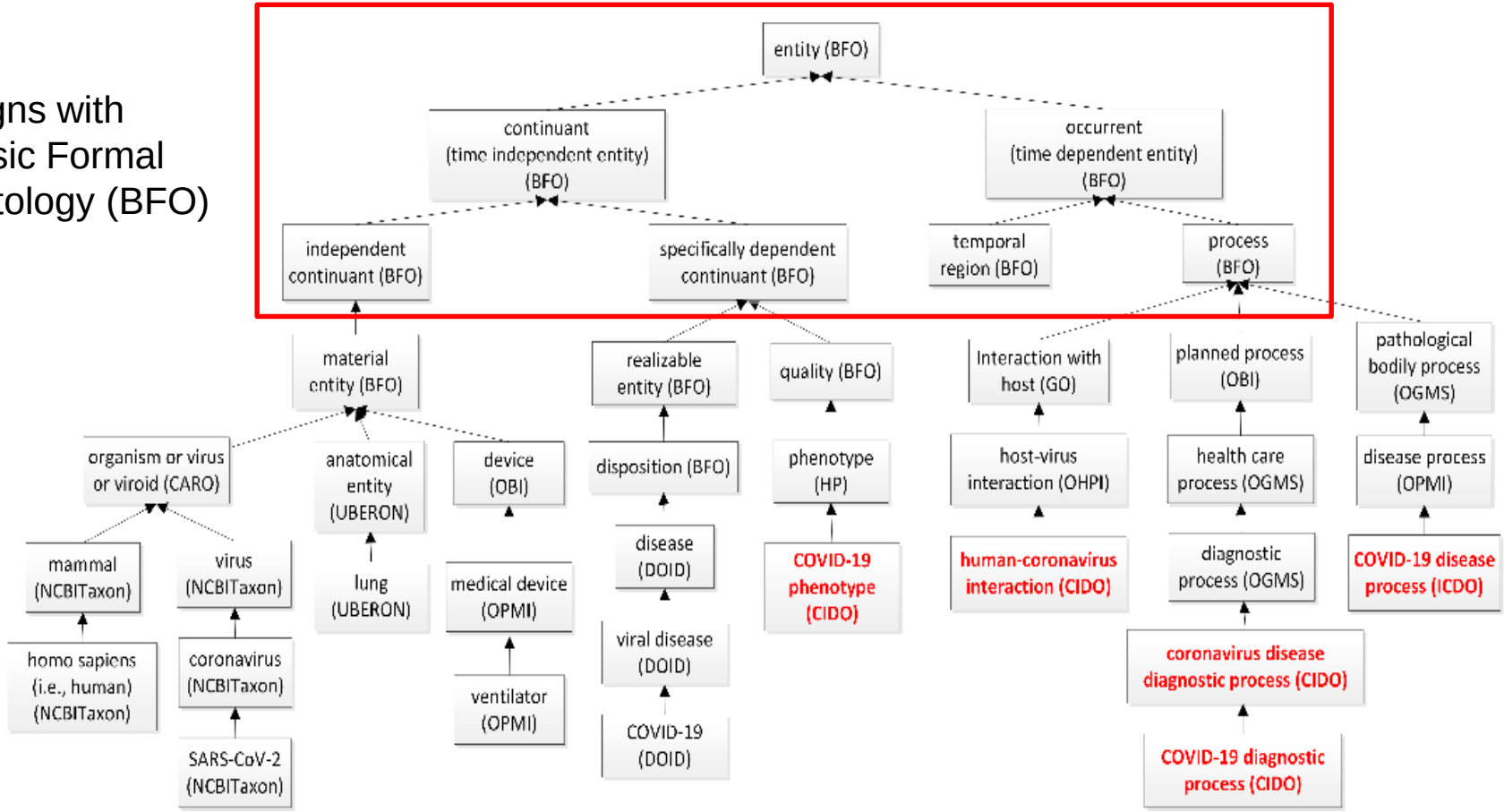
http://bioportal.bioontology.org/ontologies/CIDO

- Used in NLP and drug repurposing, among others



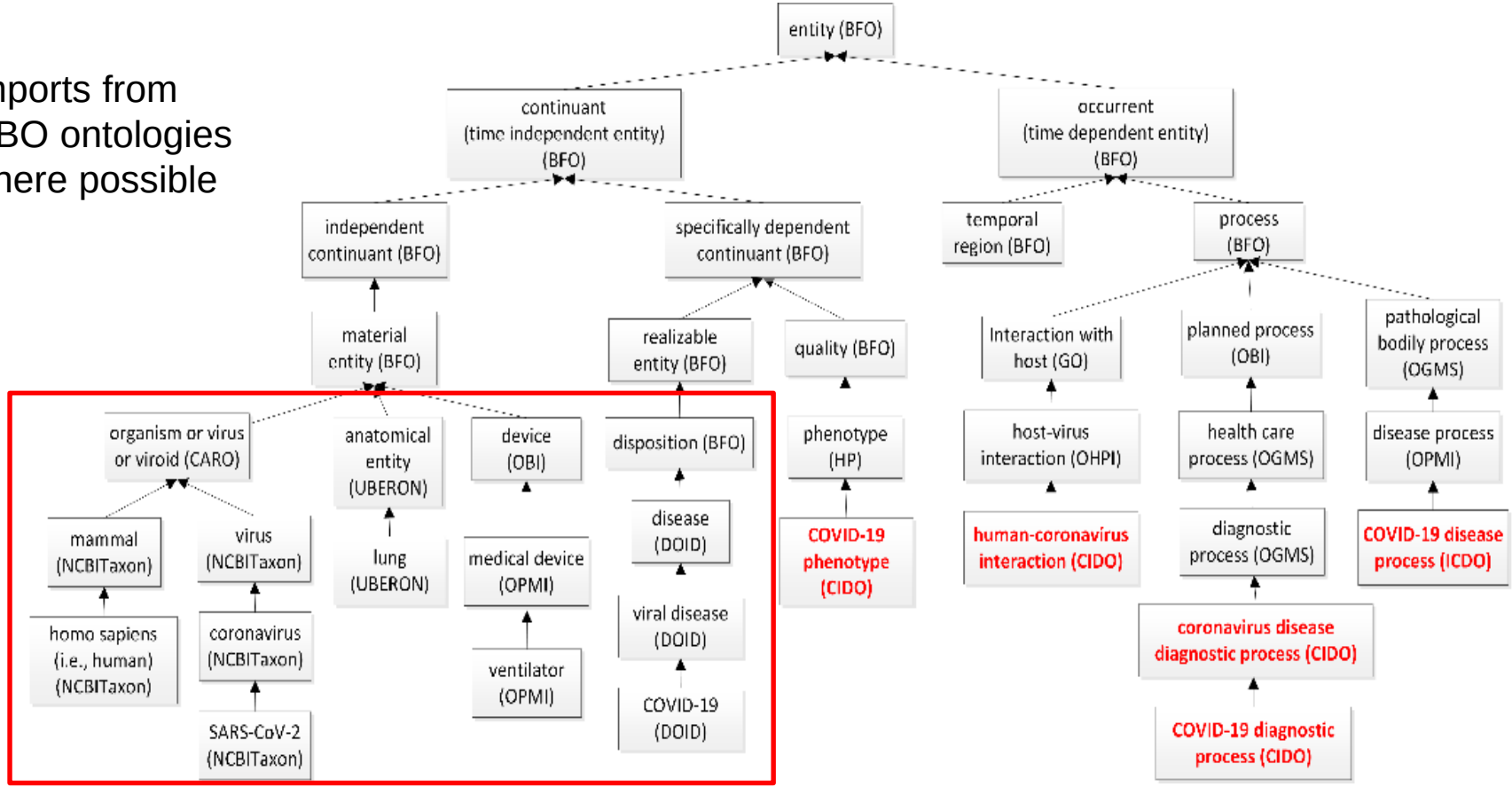
Architecture of CIDO

Aligns with Basic Formal Ontology (BFO)



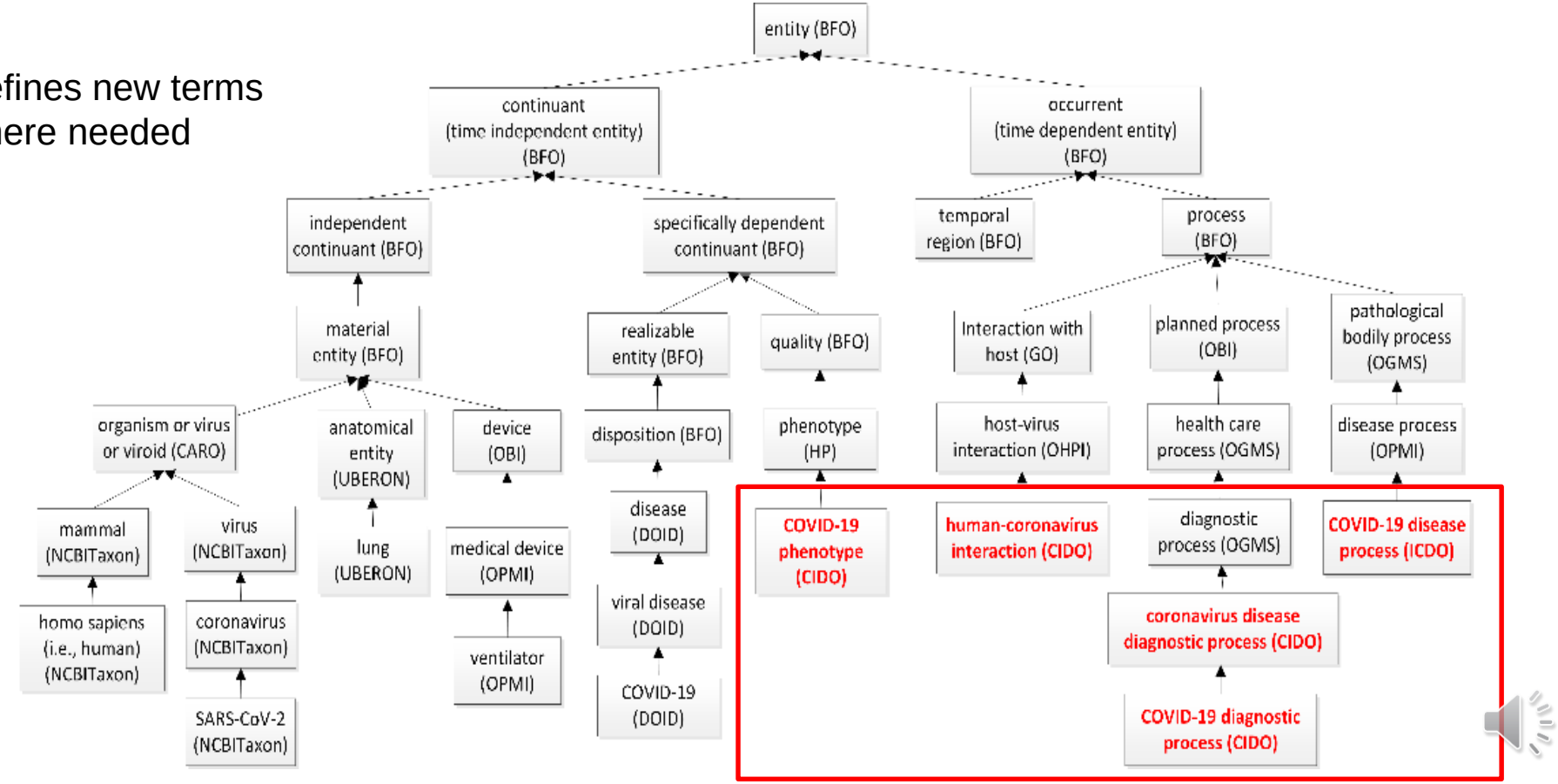
Architecture of CIDO

Imports from OBO ontologies where possible

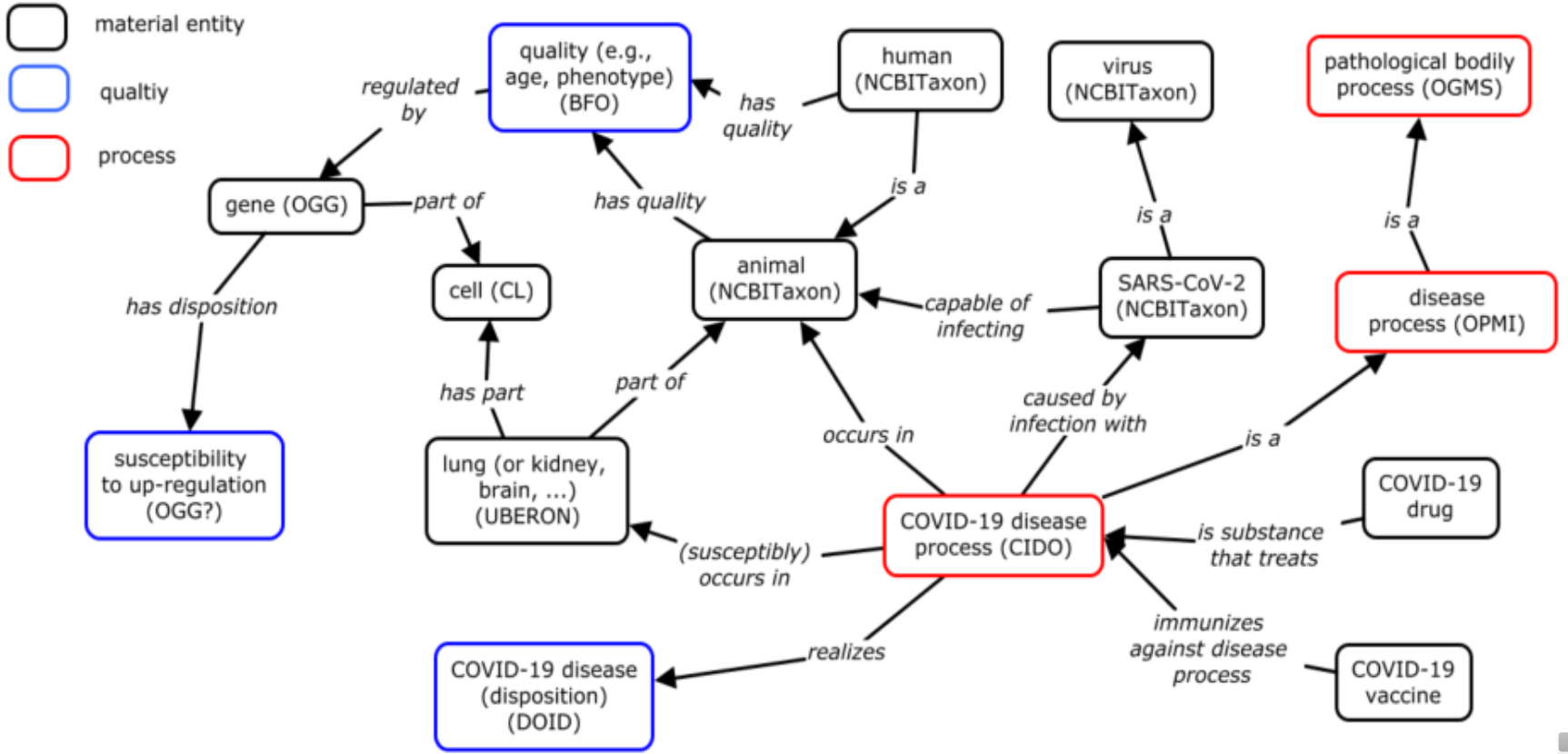


Architecture of CIDO

Defines new terms where needed

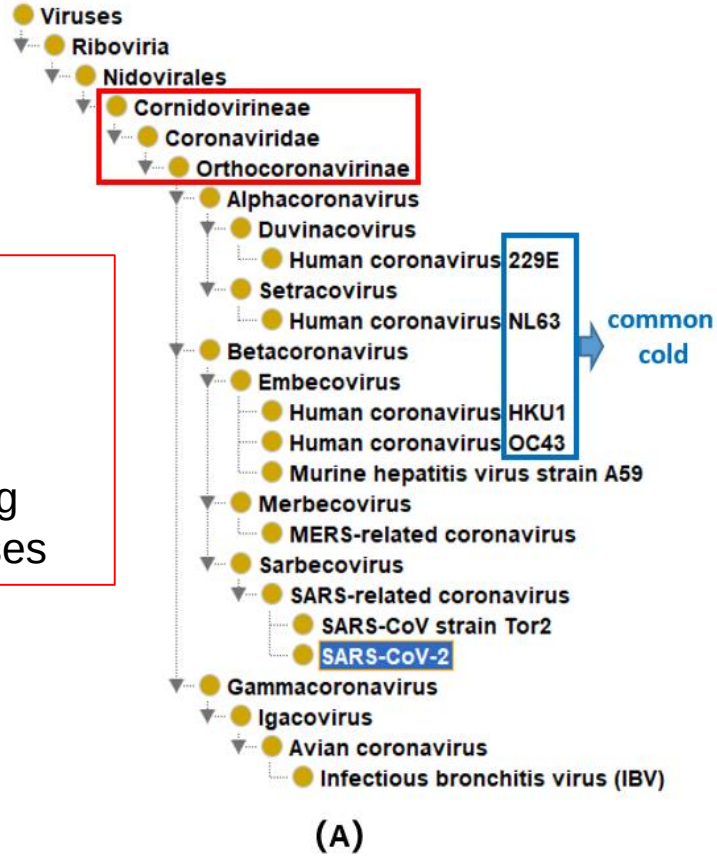


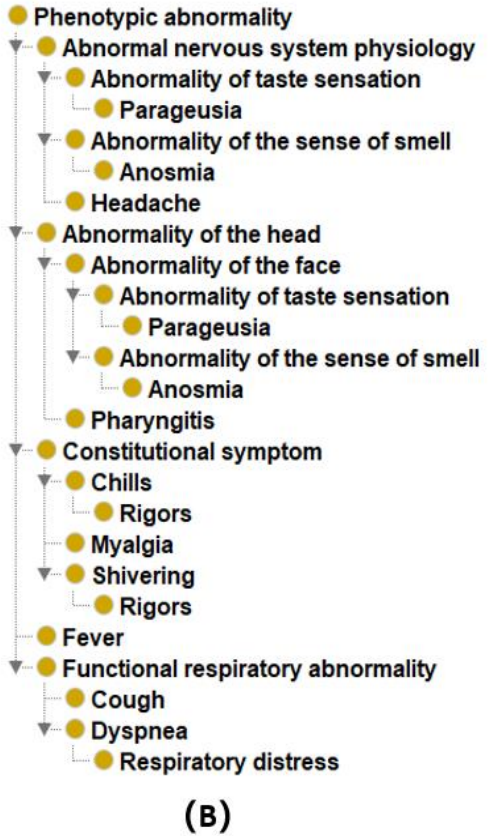
Formal Connections among Coronavirus Disease Terms



Representing Coronaviruses and COVID-19 Phenotypes

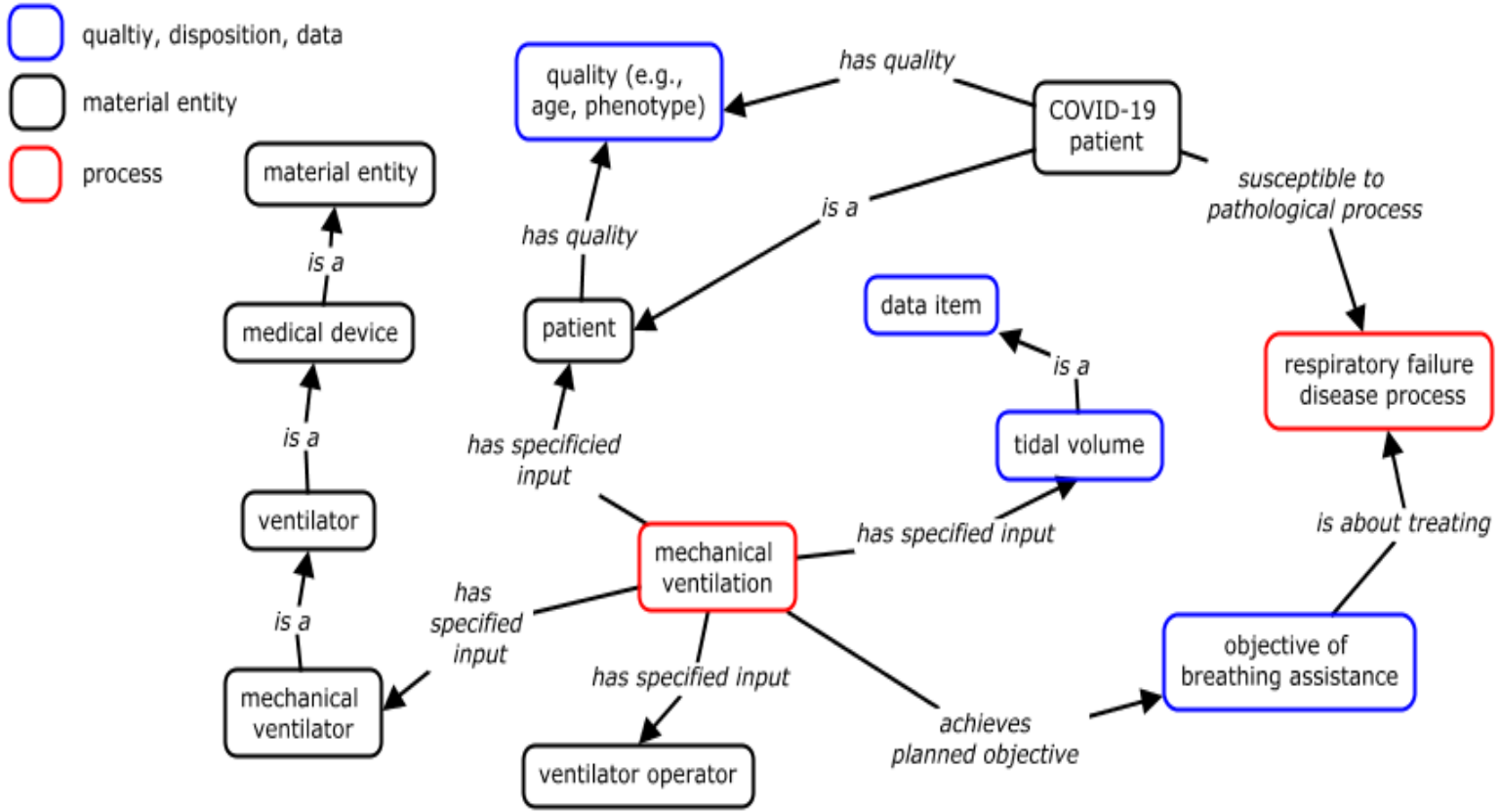
NCBITaxon hierarchy imported to CIDO, representing coronaviruses



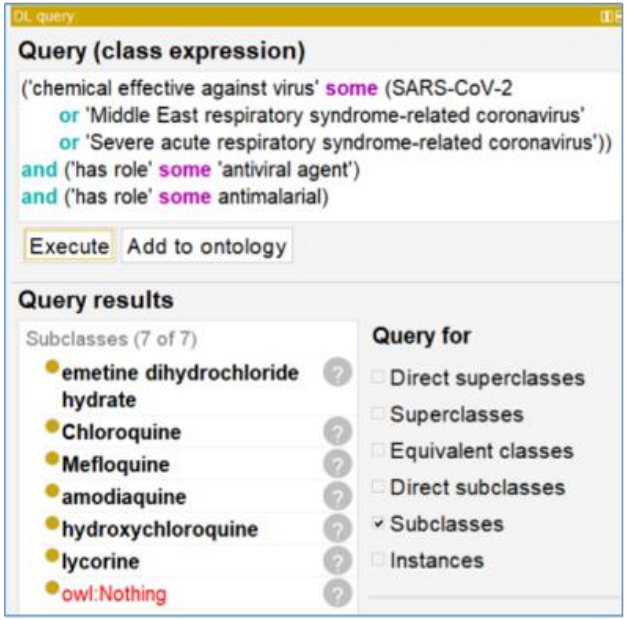


HP imports represent phenotypes shown in COVID-19

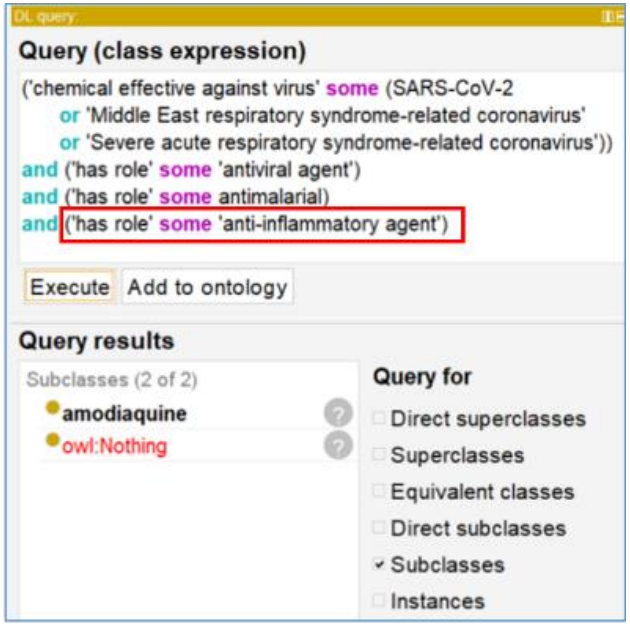
Representing Medical Devices in COVID-19 Treatment



DL Query of Potential COVID-19 Drug Treatment Options



(A)



(B)

Query of chemicals with both antiviral and antimalarial roles

Query with antiviral, antimalarial, and anti-inflammatory roles

CIDO Visualization & CIDO-based NLP and ML

- Yehoshua Perl, et al: CIDO visualization and comprehension using summarization network
- Liwei Wang, Hongfang Liu: CIDO development and usage for COVID-19 clinical data NLP
- Fatima Z. Smaili, Robert Hoehndorf: CIDO-based machine learning (ML) for drug repurposing analysis

Conclusion & Discussion

- CIDO integrates terms for coronaviruses, associated diseases, phenotypes, medical devices, treatments, vaccines, etc.
- CIDO supports computational analysis of coronavirus and disease data, evidenced by drug repurposing applications
- Representations of important coronavirus mechanisms are crucial for our progress
- **Collaborations welcome!**