Updating the CEMO ontology for future epidemiological challenges

Núria Queralt-Rosinach[1,*], Paul N. Schofield[2], Marco Roos[1] and Robert Hoehndorf[3]

[1]Leiden University Medical Center, Leiden, The Netherlands
[2]University of Cambridge, Cambridge, United Kingdom
[3]King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia

Abstract

The COVID-19 epidemiology and monitoring ontology (CEMO) is an OWL ontology built during the COVID-19 pandemic for better exchange, integration and reuse of epidemiological information. Here, we present an update of the development of the ontology and future directions in order to make it usable under different scenarios and new challenges.

Keywords

Ontologies, Epidemiology, COVID-19, FAIR

1. Motivation

The COVID-19 outbreak seriously challenged worldwide research data infrastructure for patient monitoring and public health surveillance, and sharply exposed our problems on sharing and analyzing health data. Epidemiology is the area of science that uses population statistical analysis to monitor and provide evidence on how disease outbreaks are spread or contained. The COVID-19 epidemiology and monitoring ontology (CEMO) is an OWL ontology built during the COVID-19 pandemic for better exchange, integration and reuse of epidemiological information [1] publicly available on GitHub. Making these data readily available for computational analysis is essential for efficient outbreak surveillance and evidence-based decision making for public health to provide rapid responses. Failure to providing this information unambiguously and in a machine readable way, not only blocks robust national policymaking, but also across borders.

CEMO was developed to fill the gap in the biomedical ontological landscape to represent epidemiological quantitative data. The ontology was developed following knowledge engineering best practices, and importantly the OBO principles in order to maximize its use for efficient biomedical analysis. Furthermore, and as an RDA COVID-19 recommendation on data sharing, we incorporated a patient-population link to enable reasoning and analytics of person-level

SWAT4HCLS 2023: The 14th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences
*Corresponding author.
n.queralt_rosinach@lumc.nl (N. Queralt-Rosinach); pns12@cam.ac.uk (P. N. Schofield); m.roos@lumc.nl (M. Roos); robert.hoehndorf@kaust.edu.sa (R. Hoehndorf)
0000-0003-0169-8159 (N. Queralt-Rosinach); 0000-0002-5111-7263 (P. N. Schofield); 0000-0002-8691-772X (M. Roos); 0000-0001-8149-5890 (R. Hoehndorf)

real world observations over epidemiological surveillance information based on the GA4GH Phenopackets and OHDSI OMOP data model representations. Here, we present an update of the development of the ontology and future directions in order to make it usable under different scenarios and new challenges.

2. Modifications

Firstly, we are improving the logical structure of the ontology. We decided to build an OBO ontology and use the OWL 2, a DL-based formalism and Semantic Web standard for knowledge representation, to enable data sharing and formal reasoning. Interoperability in OBO is fostered by adopting the BFO hierarchy. But, in the first version of the ontology we reused a GFO-based design pattern to represent time courses, which led to logical inconsistency between these two foundational ontologies due to incompatible conceptualizations of time in BFO and GFO. We are also improving the commitment of the ontology to the OBO principles and the epidemiological representation. Secondly, we are extending the ontology to be reused in particular use-cases such as VODAN-Africa and waste-water surveillance. The Virus Outbreak Data Network (VODAN) Africa is a collaboration of researchers and health practitioners across 15 African countries and 83 health facilities to enable access to and analysis of critical data needed to fight the novel COVID-19 in Africa following the data-visiting approach, i.e., data stays under the control of the owner and allows the consumers (e.g. analysts or machine learning algorithms) to come to the data to work with it. Thirdly, to tackle health conditions associated with climate change we are enriching the ontology with extreme climate events epidemiology. Outbreaks of climate-sensitive infectious diseases in the aftermath of extreme climatic events, such as floods and heatwaves, are of high public health concern. We are currently curating climate-related terms and evaluating how to include them in a COVID-19 epidemiology data model. With this new addition we intend to facilitate investigation of the association between extreme climatic events and COVID-19, and potentially any other disease outbreaks. Finally, we are collaborating with the COVID-19 ontology harmonization effort for several OBO ontologies being developed internationally. We expect to use this model in FAIR-based projects such as TWOC.

Acknowledgments

This initiative has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°825575 (the European Joint Programme Rare Diseases), the collaboration project Trusted World of Corona (TWOC) which is co-funded by the PPP Allowance made available by Health Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships, and The Alan Turing Institute.

References

[1] N. Queralt-Rosinach, P. Schofield, R. Hoehndorf, C. Weiland, E. Schultes, C. H. Bernabé, M. Roos, The covid-19 epidemiology and monitoring ontology (2021). doi:https://doi.org/10.5281/zenodo.5752958.