The COVID-19 epidemiology and monitoring ontology

Núria Queralt Rosinach <n.queralt_rosinach@lumc.nl>
June 21, 2021

Abstract

The COVID-19 Epidemiology and Monitoring Ontology (CEMO) provides a common ontological model to make epidemiological quantitative data for monitoring the COVID-19 outbreak machine readable, and to facilitate its exchange, integration and analysis across research centres and across life sciences domains.

---

1 Introduction

We are developing CEMO with the aim to provide a tool that facilitates the work of epidemiologists and aids to perform their research more efficiently. Especifically, make epidemic parameters' datasets structured and computationally ready to process and share. This is the first release of the ontology, which we expect to be refined iteratively with domain experts and users. In the next sections we describe the organization of knowledge in the ontology.

2 Epidemiological parameters in COVID-19 outbreak

The first question we studied was to define the basic quantitative parameters used by epidemiologists for surveillance and how they are represented in the existing biomedical ontologies. We found less than 20 terms spread in different ontologies such as incidence, incidence proportion, incidence rate, prevalence, mortality rate, odds, relative risk, and risk difference.

2.1 Curated terms

To make rapid and informed outbreak interventions, we then looked for quantitative terms used in official public health disease control websites such as the WHO and the most current epidemiological scientific studies on COVID-19 in the literature using repositories such as medRxiv.

We curated around 140 terms which is a significant bigger number from what we found in biomedical ontologies. The curated list of terms needs to be reviewed by epidemiologists both to determine if important terms for outbreak surveillance have to be added and to check and improve terms descriptions. The curated terms can be accessed following the link to the quantitative indicators sheet.

3 CEMO ontology

The curation task was not meant to achieve a comprehensive list of epidemic parameters. However, it was enough to realize that there was a gap of epidemiological concepts in the biomedical ontology landscape. To fill this gap we created a new ontology, CEMO, to organize epidemic parameters for surveillance in outbreaks using the COVID-19 use case.

3.1 Structure

CEMO was designed to structure epidemiological data and to integrate it with the current biomedical knowledge for exchange and analysis. For that reason we structure it following the Open Biological and Biomedical Ontology (OBO) Foundry principles for ontology development[1] (The OBO Foundry is a community that aims to develop interoperable and scientifically accurate ontologies for the biological sciences).

In Figure 1 we show the basic structure of the ontology, which is composed of four main top levels: *continuant*, *occurrent*, *epidemiological core concepts* and *other indicators*. OBO requires 'continuant' and 'occurrent' hierarchies, while 'epidemiological core concepts' is a new level to organize core domain concepts for epidemiology surveillance, and 'other indicators' is a new level to organize other indicators related to epidemiology surveillance. CEMO model is structured in 'continuant', 'occurrent' and 'epidemiological core concepts', which we describe in the next section.

[1] http://www.obofoundry.org/principles/fp-000-summary.html

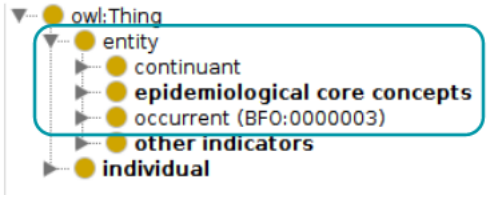

Figure 1: Ontology structure

3.2 Epidemiological classes

We organize two main epidemiological classes: domain basic classes and quantitative parameters classes.

3.2.1 Domain basic classes

Domain basic classes are organized under the 'epidemiological core concepts' parent class, see Figure 2. These classes define the scope of the ontology and are needed to define and interpret the quantitative parameters classes used in epidemiology surveillance. Both are related by axiom patterns. More information is available on GitHub.

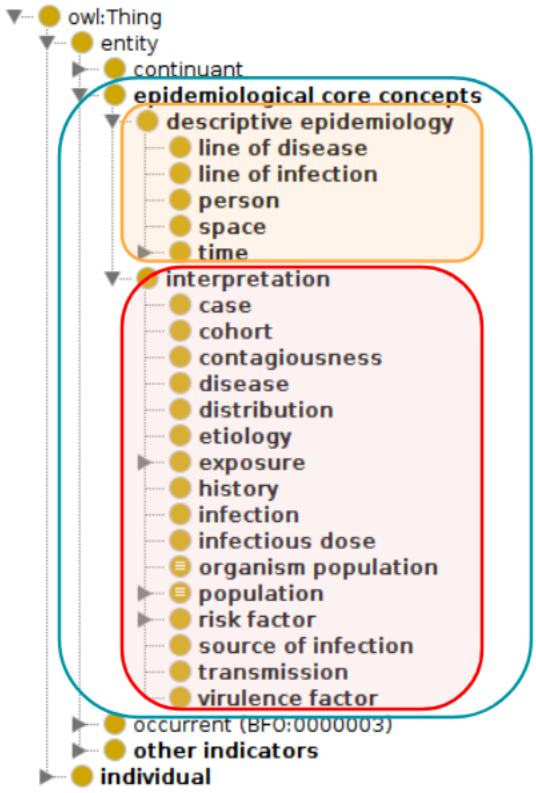

Figure 2: Epidemiology surveillance basic classes

Descriptive epidemiology Some quantitative parameters in epidemiology surveillance are used to describe the outbreak status. These parameters are typically attributes of person, space and time. Other concepts used in this sense are the line of disease and the line of infection.

Interpretation Core concepts in epidemiology surveillance, which are organized under the parent class 'interpretation', and are used to relate every epidemiological quantitative parameter with its interpretation within the epidemiology surveillance scope.

3.2.2 Quantitative parameters classes

Here we focus on epidemiological quantitative parameters or epidemic parameters. As we can see in Figure 3, these parameters are defined as classes in the ontology and organized under the parent class 'quality', like in the OBO ontology Infectious Disease Ontology (IDO)[2]. These classes are the 140 manually curated terms extracted from surveillance websites and epidemiological studies in the literature.

Figure 3: Epidemiological quantitative parameters classes

3.3 Quantitative model

These quantitative parameters are modelled following the SIO design pattern for measurements. This is a pattern that describes every measurement as the output of a process that can be performed by persons, roles or machines. The ontological model of classes and relations for epidemiological measures or quantitative parameters can be seen at the top of Figure 4, and the same model within the structure of the ontology at the bottom.

3.4 Patient-population link

An important aspect of epidemiology surveillance is to link population statistics with patient clinical information. This is an RDA recommendation[3] and we modelled this in our ontology inspired by health data standards such as OHDSI OMOP common data model and GA4GH Phenopackets, see Figure 5.

3.5 Other indicators

We also curated epidemic indicators used in COVID-19, but not necessarily of quantitative type. This list is still under development in the other indicators sheet.

[2] https://www.ebi.ac.uk/ols/ontologies/ido
[3] https://doi.org/10.15497/rda00052

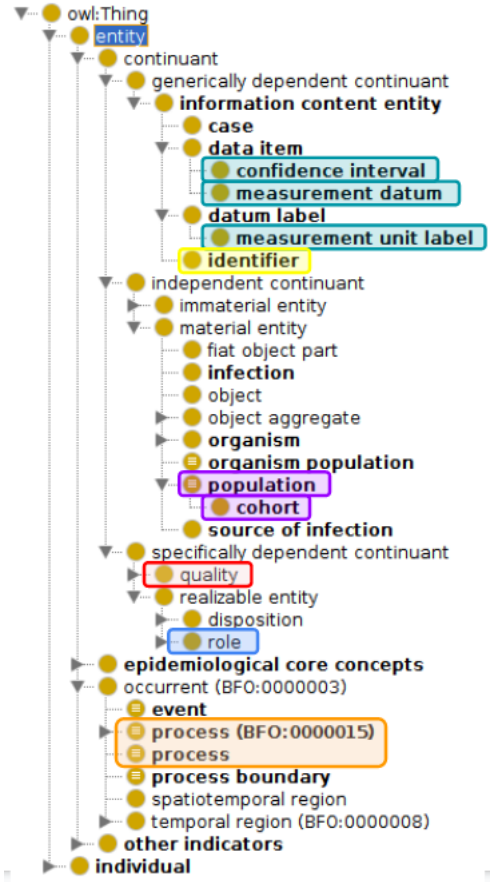

Figure 4: Ontological model

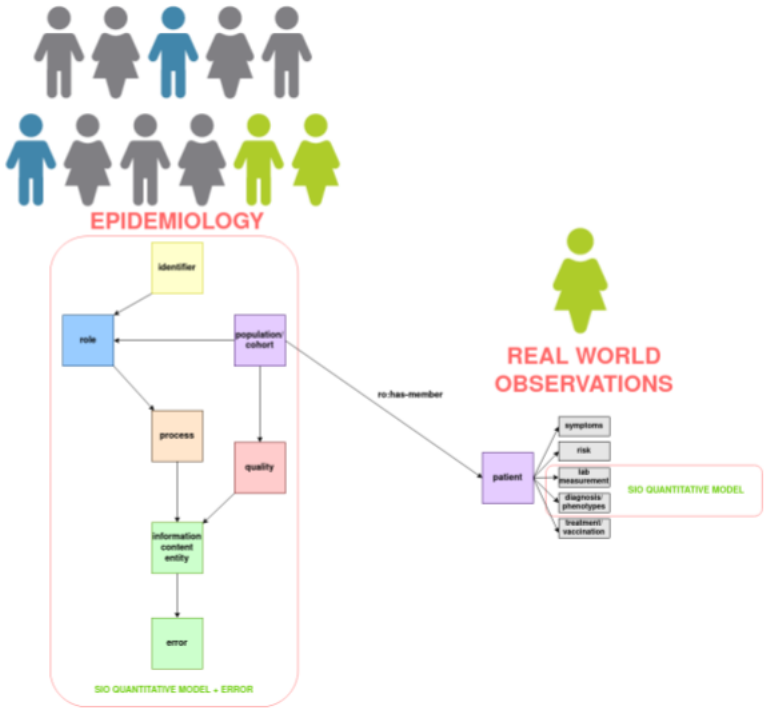

Figure 5: Patient-population link for precision medicine and precision epidemiology