The Cardiovascular Disease Ontology (CVDO)

Mercedes Arguello Casteleiro[1], Julie Klein[2], and Robert Stevens[1]

[1]School of Computer Science, University of Manchester, Oxford Road,Manchester, UK M13 9PL.
[2]Institut National de la Sante et de la Recherche Medicale (INSERM), U1048, Toulouse, France 24105.

ABSTRACT

According to the World Health Organisation, four out of five cardiovascular diseases (CVDs) deaths are due to heart attacks and strokes. Asymptomatic vascular damage accumulates for years before patients are identified and subjected to therapeutic measures. The SysVASC project seeks to provide a comprehensive systems medicine approach to elucidate pathological mechanisms for CVD, which will yield molecular targets for therapeutic intervention. To achieve this aim it is necessary to gather and integrate a large amount of data from omics experiments in CVD biology. This data is currently underexploited and scattered over many publications or held in supplemental data.

The CVD ontology (CVDO) is developed as part of the SysVASC project to provide the infrastructure to integrate omics data collected from the literature. The omics experiments contain data from multiple biological levels, such as genes, mRNA, proteins, miRNA, and metabolites. The CVDO is a rich application ontology supporting complex queries that reuse biological resources, such as the UniProt Knowledgebase, the miRBase from EMBL-EBI, and the Human Metabolome Database (HMDB). At the core of CVDO is the Ontology for Biomedical Investigations (OBI) along with other reference ontologies produced by the OBO Consortium, such as the Protein Ontology (PRO), the Sequence Ontology (SO), the three Gene Ontology (GO) sub-ontologies, the Chemical Entities of Biological Interest Ontology (ChEBI), the NCBI Taxonomy Ontology, the Cell Ontology (CL), the Uber Anatomy Ontology (UBERON), the Phenotypic Quality Ontology (PATO), and the Relationship Ontology (RO).

The CVDO's OWL is converted into JSON-LD and exploited for an inverted flat index that enables the autocompletion of queries. It is essential to provide a simple tool that enables biologists and clinicians to quickly survey a wide range of omics datasets, where the connection between miRNA and genes, or metabolites and proteins is embedded into a formal representation of knowledge in OWL-DL. The CVDO is another example of re-using a group of application ontologies, with minimal additions, to create a knowledge base for a field of interest.