

# An ontology-driven web platform for publishing data specifications to support data sharing

Damion M. Dooley<sup>a</sup>, Ivan S. Gill<sup>b</sup>, Emma J. Griffiths<sup>a</sup>, William L. Hsiao<sup>abc</sup>

<sup>a</sup>*Department of Pathology, The University of British Columbia, Vancouver, BC, Canada*

<sup>b</sup>*Department of Molecular Biology and Biochemistry, SFU, Burnaby, Canada*

<sup>c</sup>*BCCDC Public Health Laboratory, Vancouver, Canada*

## Abstract

Ontologies are seen as a possible lingua-franca for enabling data sharing across databases from different agencies where database content is actually about the same subject material, as for example arises in public health epidemiological food-borne disease outbreak investigations that cut across international borders and different levels of government. However, coming to consensus about how to encode database content according to OWL-driven vocabulary is not finished business. To explore a possible solution, the Hsiao Public Health Bioinformatics Lab has created a system to enable the crafting of ontology-driven data specifications in alignment with the *Ontology for Biomedical Investigations* data item and value specification framework. An initial vision was presented in an ICBO 2016 poster which led to funding for the effort, followed by a mid-project presentation at JOWO 2017, and now our working first release, called the Genomic Epidemiology Entity Mart (GEEM) portal, named after the bulk of the subject area that we have focused on so far, and which is represented in our Genomic Epidemiology Ontology (GenEpiO). The GEEM project update will elaborate on the finalized specification structure that we propose, including details on recent OBI data representation work.

GEEM specifications detail what information content entities are about and how they are measured. An ontology curator works with data architects of existing or new databases to determine what content the specifications must cover to enable data sharing. The curator then sets up each specification component as an OBI value specification axiom, detailing what it is about (e.g. a quality of an entity or entity part, like mass or color, or a calculation like BMI or age) and how it is measured (as a categorical, ordinal, numeric, or string value). The GEEM

software converts much of this content into a JSON representation which software implementers can then use (and customize and refresh periodically) to craft data conversion scripts, and even data management forms. In this way the GEEM specifications act as a hub that “spoke” software applications consume and/or generate compatible data with. Data architects and software implementers can use GEEM specifications without needing ontology curation or owl logic experience. The GEEM portal provides an OAuth2 login for users so that they can set up, customize and download their own “packages” of specifications over time.

## Keywords:

ontology, database, web portal, value specification, data sharing

## Reference

- Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, Clancy K, Courtot M, Derom D, Dumontier M, Fan L. The ontology for biomedical investigations. 2016. <https://doi.org/10.1371/journal.pone.0154556>
- Dooley DM, Griffiths EJ, Brinkman FS, Hsiao WW. An OBI Ontology Datum Proof Sheet. 2016. [http://ceur-ws.org/Vol-1747/IP24\\_ICBO2016.pdf](http://ceur-ws.org/Vol-1747/IP24_ICBO2016.pdf)
- Dooley DM, Griffiths EJ, Gosal G, Brinkman FS, Hsiao WW. The Genomic Epidemiology Ontology and GEEM Ontology Reusability Platform. 2017. [http://ceur-ws.org/Vol-2050/ODLS\\_paper\\_12.pdf](http://ceur-ws.org/Vol-2050/ODLS_paper_12.pdf)
- Griffiths E, Dooley D, Graham M, Van Domselaar G, Brinkman FS, Hsiao WW. Context is everything: harmonization of critical food microbiology descriptors and metadata for improved food safety and surveillance. *Frontiers in microbiology*. 2017. <https://doi.org/10.3389/fmicb.2017.01068>