From Survey to Ontology, the challenges and opportunities in ontologizing questions and answers

Lauren Chan[1], Jimmy Phuong[2], Anne Thessen[3], Stephanie Hong[4], Shahim Essaid[3], Melissa Haendel[3]

[1]Oregon State University, Corvallis, OR, 97331, USA

[2]University of Washington Medical Center, Seattle, WA, 98195, USA

[3]University of Colorado Anschutz Medical Campus, Aurora, CO, 80054, USA

[4]Johns Hopkins University, School of Medicine, Baltimore, MD, 21218, USA

Abstract

Surveys are an essential tool for gathering data from individuals in a wide variety of fields. Common Data Elements (CDEs) are routinely used in surveys to streamline frequently asked questions and support question reuse. While CDEs are a widespread method for collecting survey data, they face a variety of problems including inconsistent file formats, lack of computational encoding, little to no community development standards, and frequent duplication of CDEs across various registries. Alternative approaches such as ontologies can serve as meaningful translation tools to not only support interoperability between similar/duplicate CDEs, but also to coordinate survey data with other ontology aligned data from heterogeneous methodologies including wet lab, clinical, or field research. To capitalize on the extensive data collected using CDEs, we propose a two-step approach of 1) aligning CDEs with ontology terminology to coordinate similar questions based on primary topic and 2) using standardized enumerations for CDE responses that can support data harmonization for meta analytics. Using this approach, greater data integration can support higher powered analytics and coordination of heterogeneous data types which are often difficult to study in tandem.