ON TO A Better Path to Choose Your Best Ontologies - Abstract

Asiyah Yu Lin[1, *], John Graybeal[2, *], Anna Maria Masci[3], Juliane Schneider[4], Ruth Duerr[5], Eric G. Stephan[4], Hande Kũçük McGinty[6, *]

[1] *Axle Research and Technology, Rockville, Maryland, U.S.*

[2] *Graybeal.SKI Consulting, Stanford, California, U.S.*

[3] *University of Texas MD Anderson Cancer Center , Huston, Texas, U.S.*

[4] *Pacific Northwest National Laboratory,Richland, Washington, U.S.*

[5] *Ronin Institute for Independent Scholarship, Montclair, New Jersey, U.S.*

[6] *Kansas State University, Manhattan, Kansas, U.S.*

Abstract

The development of biomedical ontologies has been rising exponentially in the last 10 years. In 2010, 181 ontologies were listed in BioPortal. As of June 2024, this number has increased to 1124, and many other open ontology registration systems exist. When an ontology engineer starts developing an ontology related to clinical studies, or anyone who wants to reuse similar ontologies, he/she may need to navigate multiple similar ontologies, such as Clinical Study Ontology, Clinical Trial Ontology, Clinical Research Ontology, SNOMED CT or NCIT, and more. A search of the term "clinical trial" shows that 35 ontologies have the term labeled "clinical trial".

This presents a challenge for data engineers and ontologists: how does one select the best fit-for-purpose ontologies that work well for specific use cases? Often the selection process is unstructured and highly biased, influenced by the domain experts involved, the background of the development team, and the team's theoretical approach to ontology adoption. While pursuing FAIR Principles for ontology development, the reuse and interoperability of ontologies have been a serious concern among ontologies, decision makers, and funders.

Beginning with work in the Research Data Alliance (RDA) Vocabulary and Semantic Services Interest Group (VSSIG) over several years, our group has been gathering published and lived evaluation techniques for choosing terms and ontologies. Here we share our observations of structured guidance and specific criteria to choose the right ontologies that fit various semantic purposes. We distinguish between criteria that can be evaluated with existing tools and services, criteria that can be evaluated manually or subjectively, and criteria that could be evaluated with as-yet-unimplemented techniques.

Finally, we invite collaboration in ongoing development of the evolving materials through the RDA VSSIG group, its slack channel, and its collaborative development in our Google Drive folder.

Keywords
Ontology, Vocabulary, Terms, Best Practices, Semantics, FAIR, Reusability