

OMOP2OBO: Semantic Integration of Standardized Clinical Terminologies to Power Translational Digital Medicine Across Health Systems

PRESENTER: Tiffany J. Callahan

Background

Despite significant progress in biobanking, translational use of electronic health records (EHRs) remains largely aspirational due to its disconnectedness from biomedical knowledge. Open Biomedical Ontologies (OBOs) provide detailed representations of biological domains, are logically verifiable using description logics, and can be easily integrated with basic science data and clinical research (Figure 1).

MAPPING CHALLENGES

- Limitations of existing work in this domain:
 - Focused on specific diseases and biological domains
 - Largely limited to one-to-one mappings
 - Rarely include external validation
- Existing algorithms cannot automatically capture complex biological semantics underlying clinical concepts

GOAL: Develop **OMOP2OBO**, the first health system-wide integration and alignment between OMOP standardized clinical terminologies and OBO ontologies.

Methods

- OMOP-normalized Children's Hospital Colorado EHR data.
- OBOs were selected by domain experts and included diseases, phenotypes, anatomical entities, cell types, organisms, small molecules, vaccines, and proteins.
- Mappings were performed using the pipeline in Figure 2.
- 20% of the most challenging mappings were verified by a panel of clinical and molecular domain experts.
- Mapping generalizability was assessed by comparing the coverage of mapped concepts to 2 independent EHRs.

Results

- 20,850 condition concepts were mapped to 4,661 phenotypes and 3,614 diseases (Figure 3).
- 1,574 of drug ingredient concepts were mapped to 1,422 chemicals, 91 proteins, 39 organisms, and 54 vaccines.
- 11,072 measurement results mapped to over 920 phenotypes, 25 anatomical entities, 27 cell types, 338 chemicals, 194 organisms, and 113 proteins.

VALIDATION

- Domain expert agreement was found for 91.6% of measurements, 75.8% of ingredients, and 73.8% of conditions.
- 80-92% for conditions, 91-96% for ingredients, and 50-55% for measurement concepts on EHR from two independent health systems revealed.

Discussion

OMOP2OBO is the first health system-wide integration of OMOP clinical terminology concepts and OBO biomedical ontologies.

FUTURE WORK

We are currently working on expanding the mapping provenance to include mechanisms of actions and conducting an expanded coverage study, using data from the OHDSI Concept Prevalence Study.

Aligning molecular data to OMOP standardized terminologies will support biologically meaningful analysis of EHR data, which can be achieved by integrating knowledge from biomedical ontologies.

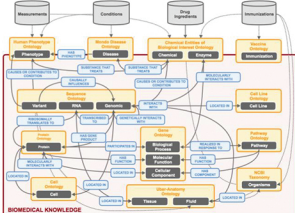


Figure 1. A knowledge representation demonstrating how different OMOP clinical domains (i.e. conditions, drug ingredients, measurements, and immunizations) can be linked with biological mechanisms of human disease using biomedical ontologies.

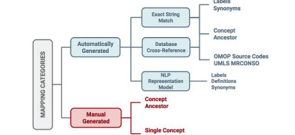


Figure 2. An overview of the OMOP2OBO mapping algorithm. There are two primary mapping strategies: Automatic and manual. The automatic approach uses all OMOP standard concepts, ancestors, labels, and synonyms and all ontology labels, synonyms, definitions, and database cross-references.

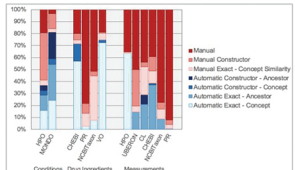


Figure 3. Mapped concepts for each ontology by clinical domain (i.e. conditions, drug ingredients, and measurements) and mapping category. HPO (Human Phenotype Ontology), MONDO (Mondo disease Ontology), ChEBI (Chemical Entities of Biological Interest), PR (Protein Ontology), NCITax (NCBI Organism Taxonomy), VO (Vaccine Ontology), UBERON (Uber-Anatomy Ontology), CL (Cell Ontology).

AUTHORS:

Tiffany J. Callahan, MPH¹, Jordan M. Wyrwa, DO¹, Nicole A. Vasilievsky, PhD², Peter N. Robinson, MD, PhD³, Melissa A. Haendel, PhD⁴, Lawrence E. Hunter, PhD¹, Michael G. Kahn, MD, PhD¹
¹University of Colorado Anschutz Medical Campus, Aurora, CO, USA; ²Oregon Health Sciences University, Portland, OR, USA; ³The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA; ⁴Oregon State University, Corvallis, OR, USA



Take a picture to download the full paper

tiffany.callahan@cuanschutz.edu

<https://github.com/callahantiff/OMOP2OBO>

ACKNOWLEDGMENTS
This work would not have been possible without support from the Colorado Health Data Warehouse and Children's Hospital Colorado Research Informatics, led Sara Costelloe Davies MPH. We also acknowledge gratitude to our domain experts for their help in validating the mappings: Teller D. Barrett MD, James A. Fennell MD, Barry Martin MD, Jessica Snider PharmD, Kate Straley PharmD, William A. Baumgartner Jr PhD. We also thank the Open Biomedical Ontologies community for their continued dedication in supporting the ontologies leveraged in this work. This work was funded by a Training Grant from the NIH NLM (TJ15LM000651) (TJC) and Google Cloud Research Credits (TJC).