Ontologizing Health Systems Data At Scale: Making Translational Discovery A Reality

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 Observational Medical Outcomes Partnership (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 24 independent EHRs.

Results

- 92367 conditions were mapped to 5661 phenotypes and 9643 diseases (Figure 3).
- 49294 drug ingredients were mapped to 4074 chemicals, 145 proteins, 2,739 organisms, and 134 vaccines.
- 11,072 measurement results mapped to 1118 phenotypes, 48 anatomical entities, 41 cell types, 446 chemicals, 428 organisms, and 176 proteins.

VALIDATION

- Domain expert agreement was found for 82.5% of conditions, 75% of ingredients, and 90.9% of measurements.
- 92.9% for conditions, 96% for ingredients, and 70% for measurement concepts on EHR from 24 independent health systems revealed.

Discussion

OMOP2OBO is the first health system-wide resource to provision interoperability between 105020 OMOP clinical concepts and 142249 concepts in eight OBO 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 standardized clinical terminologies will support **biologically meaningful analysis** of medical record data, which can be achieved by **integrating** external sources of **biomedical knowledge.**

Take a picture to **download** the **full paper**

tiffany.callahan@cuanschutz.edu
https://github.com/callahantiff/OMOP2OBO

ACKNOWLEDGEMENTS
This work would not have been possible without support from the Compass Health Data Warehouse and Children's Hospital Colorado Research Informatics, led Sara Deakyne-Davies MPH. We are also incredibly grateful to our domain experts for their help in validating the mappings: Tellen D. Bennett MD, James A. Feinstein MD, Blaine Martin MD, Jessica Sinclair PharmD, Katy Trinkley PharmD, William A. Baumgartner Jr. PhD. We also thank the Open Biomedical Ontologies Foundry for their continued dedication in supporting the ontologies leveraged in this work. This work was funded by a Training Grant from the NIH NLM [No. T15LM009451 (LEH)] and Google Cloud Research Credits (TJC).

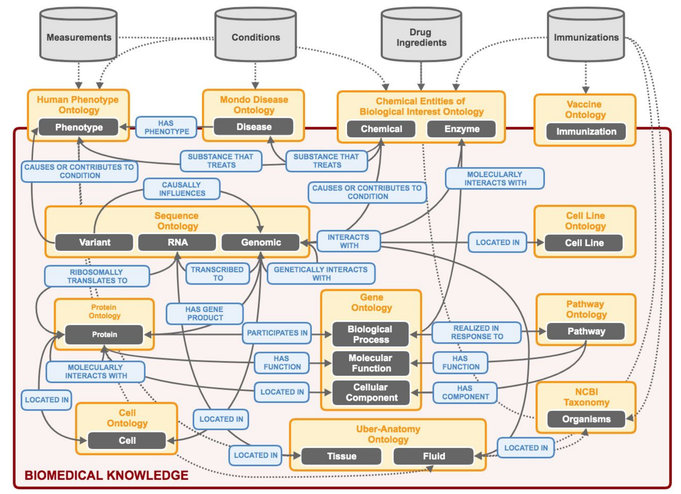

Figure 1. A Knowledge representation demonstrating how different OMOP clinical domains (i.e. conditions, drug ingredients, measurements, and immunizations) can be linked with biological mechanisms of human disease using biomedical ontologies.

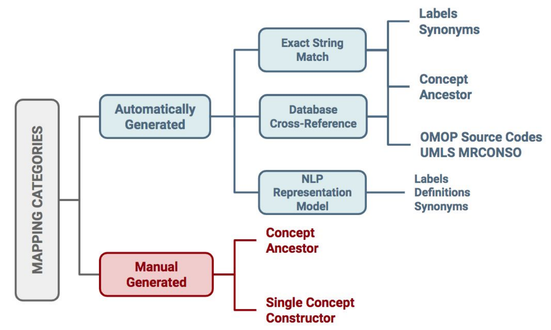

Figure 2. An overview of the OMOP2OBO mapping algorithm. There are two primary mapping strategies: Automatic and manual. The automatic approach uses all OMOP standard concepts, ancestors, labels, and synonyms and all ontology labels, synonyms, definitions, and database cross-references.

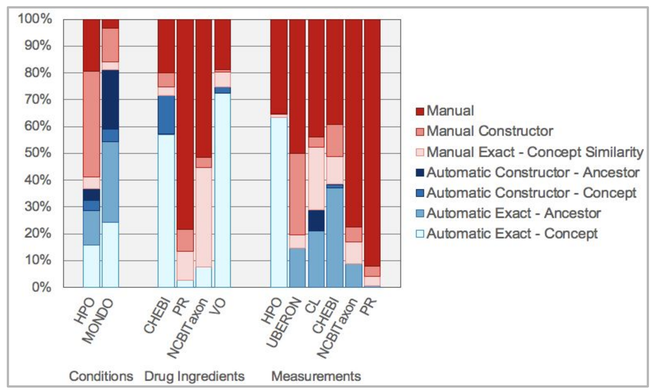

Figure 3. Mapped concepts for each ontology by clinical domain (i.e. conditions, drug ingredients, and measurements) and mapping category. HPO (Human Phenotype Ontology), MONDO (Mondo disease Ontology), CHEBI (Chemical Entities of Biological Interest), PR (Protein Ontology), NCBITaxon (NCBI Organism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

AUTHORS:

Tiffany J. Callahan, MPH[1], Jordan M. Wyrwa, DO[1], Nicole A Vasilevsky, PhD[2], Peter N. Robinson, MD, PhD[3], Melissa A Haendel, PhD[4], Lawrence E. Hunter, PhD[1], Michael G. Kahn, MD, PhD[1]

[1]University of Colorado Anschutz Medical Campus, Aurora, CO, USA; [2]Oregon Health Sciences University, Portland, OR, USA; [3]The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA; [4]Oregon State University, Corvallis, OR, USA