# Using the Informatics for Integrating Biology and the Bedside Platform to Query OMOP Data in the OHDSI Ecosystem

**Jeffrey G. Klann, PhD[1,2,3], Griffin M. Weber, MD, PhD[1,4], Michele Morris[5], Michael Mendis[2], Diane Keogh[1,6], Shawn N. Murphy, MD, PhD[1,2,3]**

**[1]Harvard Medical School; [2]Mass General Brigham; [3]Massachusetts General Hospital ; [4]Beth Israel Deaconess Medical Center; [6]i2b2-tranSMART Foundation – All in Boston, MA ;**

**[5]University of Pittsburgh – Pittsburgh, PA**

## Background

Clinical data warehouses and research networks each use a specific Common Data Model (CDM) on which software tools and analysis programs are built. To participate in multiple data research initiatives, it is generally necessary to transform data into different models, which creates additional local effort and can lead to information loss due to mapping and transformation issues. For example, OMOP-formatted data is needed for AllOfUs, but i2b2-formatted data is needed for NCATS ENACT (Next-Gen ACT). Here, we demonstrate an approach to tightly integrate these two popular CDMs, potentially eliminating the need for costly data transformation.

The Observational Medical Outcomes Partnership (OMOP) data model and Integrating Biology and the Bedside (i2b2) data model are two popular approaches to clinical data warehousing, and both are ideal for representing standard EHR data for shared analytics.

Integrating Biology and the Bedside (i2b2) is an open-source clinical data warehousing and analytics platform in use at over 200 locations worldwide. Over 15 years, i2b2 has evolved from a single-site cohort identification tool into an enterprise-ready, network-enabled analytics platform with a diverse open-science community.[2] i2b2 emphasizes flexibility, making it ideal for representing new and non-standard healthcare data (e.g., social determinants of health, survey instruments, and patient-reported outcomes). The platform is frequently used as a self-serve clinical research portal for investigators, with graphical preparatory-to-research cohort finding. It can also be used to find cohorts over large data networks using the Shared Health Research Informatics Network (SHRINE) system, such as NCATS' 40+ site ENACT research network, covering 90 million patient lives. [3]

i2b2 consists of three core software components: a user interface that supports graphical cohort design and analysis, an application framework which provides a REST API for queries in a database agnostic way, and a database that houses both the actual patient data and metadata used in building queries. The metadata about query terms is called an i2b2 "ontology," and it organizes patient-related concepts (e.g., ICD-10 and LOINC codes) into a browsable and searchable hierarchy. The ontology provides an "information model" for each i2b2 instance, defining all possible concepts that can be represented.

The user interface and database can be swapped out - in this way, there are several i2b2 user interfaces and several database platforms are supported. The data model in the database can be swapped out as well, as long as it expresses data that support i2b2 queries (e.g., patient- and encounter- centric data), and we have previously demonstrated (in 2018) that i2b2 can query OMOP and PCORnet databases.[4]

## Methods

Our current undertaking is to offer out-of-the-box support for i2b2's tools to run on the OMOP CDM

alongside i2b2 CDM, so that sites already using OMOP can utilize i2b2's self-service cohort finding tools and participate in networks such as ENACT. The 2018 i2b2-on-OMOP proof-of-concept had limitations: queries were slow; it was difficult to configure; and, it could query only in SNOMED, which is not widely used by i2b2 sites or networks.

i2b2 1.8, to be released in early 2024, will include an installation option to use existing OMOP database tables alongside the i2b2 application layer (and, optionally, i2b2 data tables) for seamless integration. This feature is being developed by the i2b2-tranSMART Foundation, which maintains the i2b2 software, in a philanthropic partnership with Dell Technologies, and through the NCATS ENACT (Next-Gen ACT) grant.

The foundation of i2b2-on-OMOP support is the ENACT ontology, a concept hierarchy that is common to all sites in the ENACT network. The ENACT team has created and will maintain an OMOP version of this ontology, which will map all ENACT concepts to OMOP codes and the fields and tables in the OMOP CDM. This is complex because these mappings are one-to-many and non-obvious. OMOP codes and domains are chosen based on simplicity for analytics queries, which means that a single ENACT concept (e.g., "obesity diagnosis") might map to multiple OMOP concepts (e.g., both "diagnosis of obesity" and "observation of obesity") that are recorded in separate tables.

It is further possible to have some ontology elements refer to i2b2's fact table, while other elements refer to OMOP tables. This will allow users to take advantage of OMOP's standardization alongside i2b2's ability to represent non-standard data, all through the i2b2 API and tools.

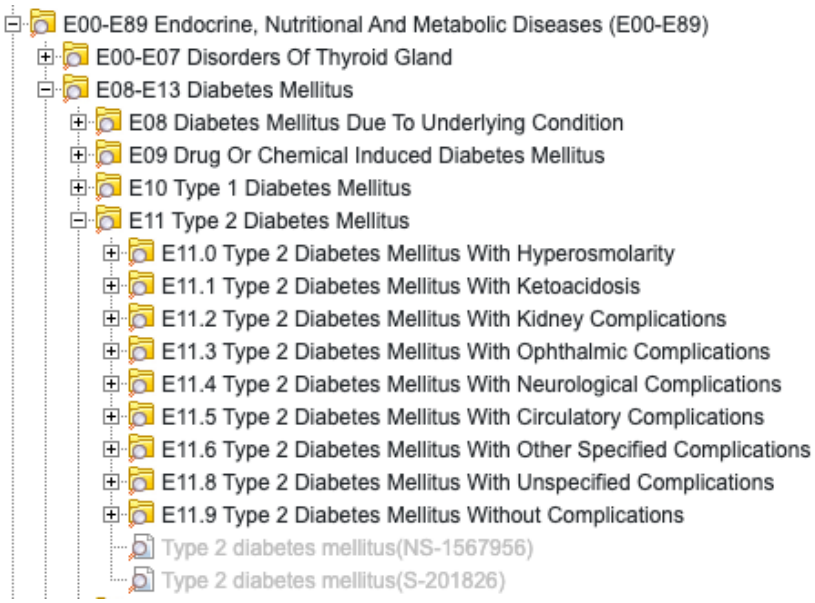


**Figure 1.** A portion of the ACT-OMOP Diagnosis ontology, showing the diabetes hierarchy with OMOP codes hidden within the folders (in light grey)

We developed a 3-million-patient synthetic data source consisting of OHDSI's Medicare SynPUF dataset and an OMOP-formatted Synthea dataset[5], and we loaded these onto a server with the ACT-OMOP ontology, the i2b2 application server, and supporting database views and tables for the i2b2 Query Tool.

**Results**

i2b2's primary presentation layer is a graphical cohort identification and characterization tool ("query

tool") that has evolved over the years to support advanced analytics like breakdowns, temporal queries, and plugins like timeline views and data export. It is driven by a hierarchical "ontology browser" that can be defined to include any type of structured biomedical data, from EHR to social determinants of health. The i2b2 query tool was developed using technology that is now 15 years old, and the user interface was lacking in modern user interface conventions such as drag-and-drop and panel resizing. For 2023, the query tool has been completely re-engineered to provide a new intuitive interface to define queries.

The demo will focus on i2b2's graphical query tool running live queries of the 3-million patient synthetic OMOP database. The demo will include boolean logic queries as well as advanced features like temporal queries, date constraints, and plugins. Additionally, we will provide background on the data models, the tooling design, and the installation process.

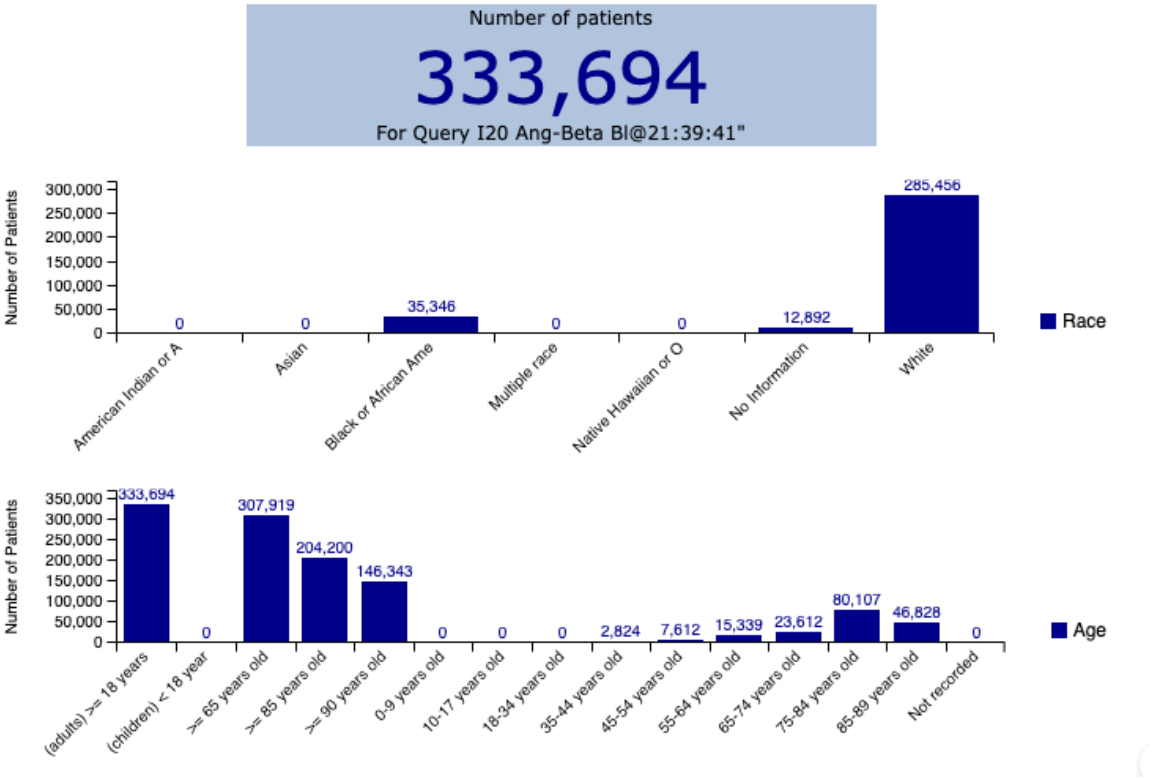


**Figure 2.** Shown here is the count and a bar chart by age and race for all patients who had a diagnosis of angina on or after 1/1/2005 and any medication record of a beta blocker on a synthetic OMOP dataset.

## Conclusion

OHDSI and i2b2 are two of the most popular clinical data warehousing platforms. Here, we take advantage of the i2b2 platform's ability to connect to any data model that is centered around patients and facts. We demonstrate hardened tools to allow the newly-redesigned i2b2 query tool to run on existing OMOP datasets. These will be available in i2b2 1.8, which will be released in early 2024.

## References

1. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI):

Opportunities for Observational Researchers. *Stud Health Technol Inform* 2015; 216: 574–578.

2. Murphy SN, Weber G, Mendis M, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc* 2010; 17: 124–130.

3. Visweswaran S, Becich MJ, D'Itri VS, et al. Accrual to Clinical Trials (ACT): A Clinical and Translational Science Award Consortium Network. *JAMIA Open* 2018; 1: 147–152.

4. Klann JG, Phillips LC, Herrick C, et al. Web services for data warehouses: OMOP and PCORnet on i2b2. *J Am Med Inform Assoc* 2018; 25: 1331–1338.

5. Walonoski J, Kramer M, Nichols J, et al. Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *J Am Med Inform Assoc* 2018; 25: 230–238.