

Argos: Simple Cohort Builder and Analysis Tool for Federated Networks Using OMOP

Griffin M Weber, MD, PhD^{1,2}, Jeffrey Klann, PhD³, Shawn N. Murphy, MD, PhD^{3,4},
Aaron Abend, MBA⁵, Diane Keogh^{1,6}

¹Harvard Medical School, Boston, MA; ²Beth Israel Deaconess Medical Center, Boston, MA; ³Mass General Brigham, Boston, MA; ⁴University of Washington, Seattle, WA; ⁵Prognosis Data; ⁶i2b2 tranSMART Foundation

Background

The Massachusetts Consortium on Pathogen Readiness (MassCPR) was formed in 2020 as a collaboration between Massachusetts hospitals, academic institutions, and public health agencies to address COVID-19, but has now expanded to other pathogens with pandemic potential. The COVID-19 pandemic demonstrated the importance of up-to-date high-quality data during a public health crisis. However, obtaining these data is often challenging. Although, collectively, MassCPR institutions locally have clinical data on millions of patients, these data are stored in siloed repositories. Various technical obstacles, privacy and regulatory concerns, and governance roadblocks have limited the full potential of the data. We therefore launched an effort to construct a MassCPR federated data network, which now connects electronic health records (EHRs) and public health databases across MassCPR institutions.

Methods

The MassCPR network is originally based on the i2b2¹ and SHRINE² platforms. Though we extended these tools to OMOP, creating **Argos**, to support MassCPR sites using OMOP and to make the platform available to the broader OHDSI Community. The Argos software combines i2b2's easy-to-use cohort builder and query tool with SHRINE's ability to perform real-time queries across a federated network of institutions using OMOP. We added features to import public health datasets, such as wastewater pathogen levels, and created new analysis modules for spatiotemporal data visualizations and reports. **Figure 1** illustrates the network structure, with patient-level data remaining locally within each site. Federated queries are distributed to sites and run automatically on local databases, returning only obfuscated aggregate counts to users for research and publication health projects related to infectious disease.

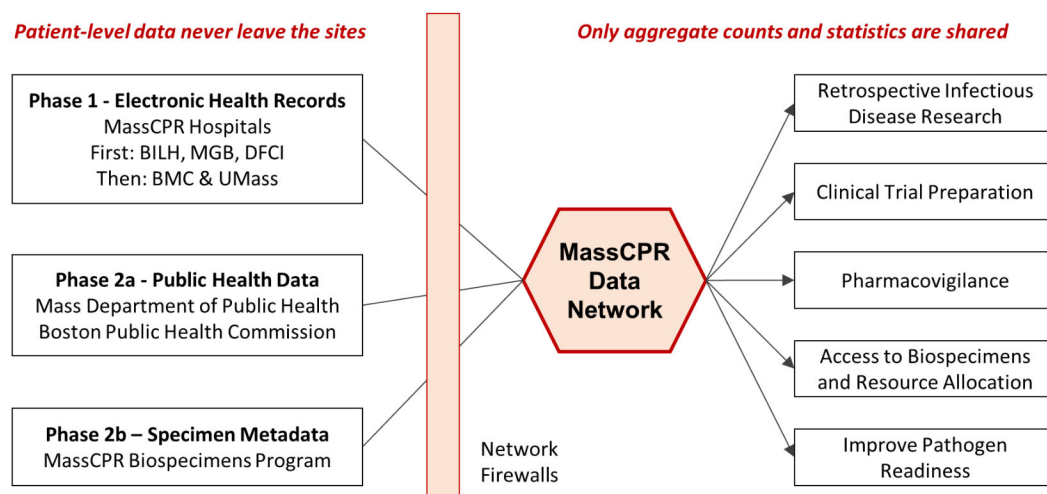


Figure 1. MassCPR Federated Data Network.

Results

The MassCPR Federated Data Network went live in February, 2026, with data on more than 7 million patients from two dozen hospitals. Users login to a central website to define cohorts and analyses. These are then distributed and run automatically at each site, returning aggregate results within seconds to minutes.

Figure 2 illustrates the Argos query builder. Users browse or search medical concepts using structured terminologies (left) and select cohort inclusion and exclusion criteria on the right. Cohorts might define a specific group of patients or all patients in the network. Various analysis tools and interactive spatial and temporal visualizations help users explore the cohort data across the network or drill down into an individual site (**Figure 3**). These include the top diagnosis categories in a cohort, patient demographics displayed as an NIH-style Enrollment Table, maps showing where patients are located, and monthly counts of the number of patients with different infectious diseases. The “pathogen timeline” module is an example where we can overlay counts from hospital EHRs with public health data.

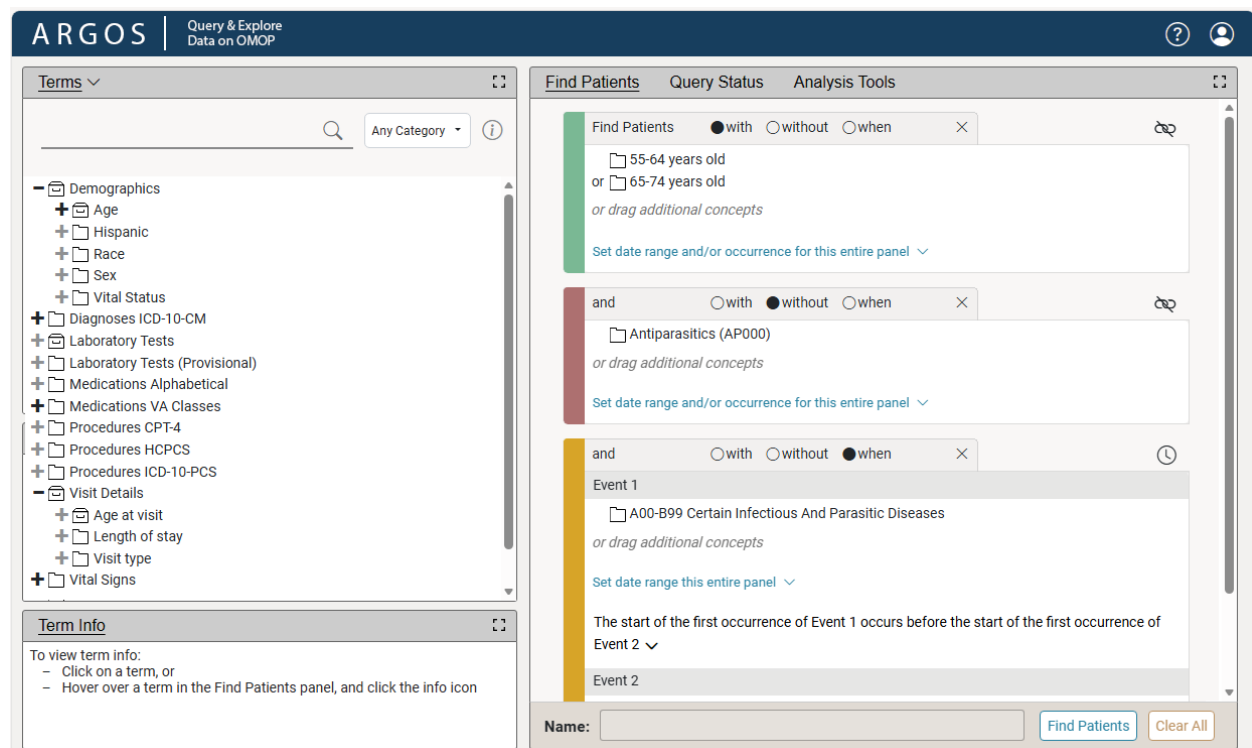


Figure 2. Argos user interface for defining a patient cohort. Medical concepts (left) can be dragged-and-dropped into the cohort builder (right). Queries can be run either locally within an institution on its own OMOP database or distributed to other sites across a federated network.

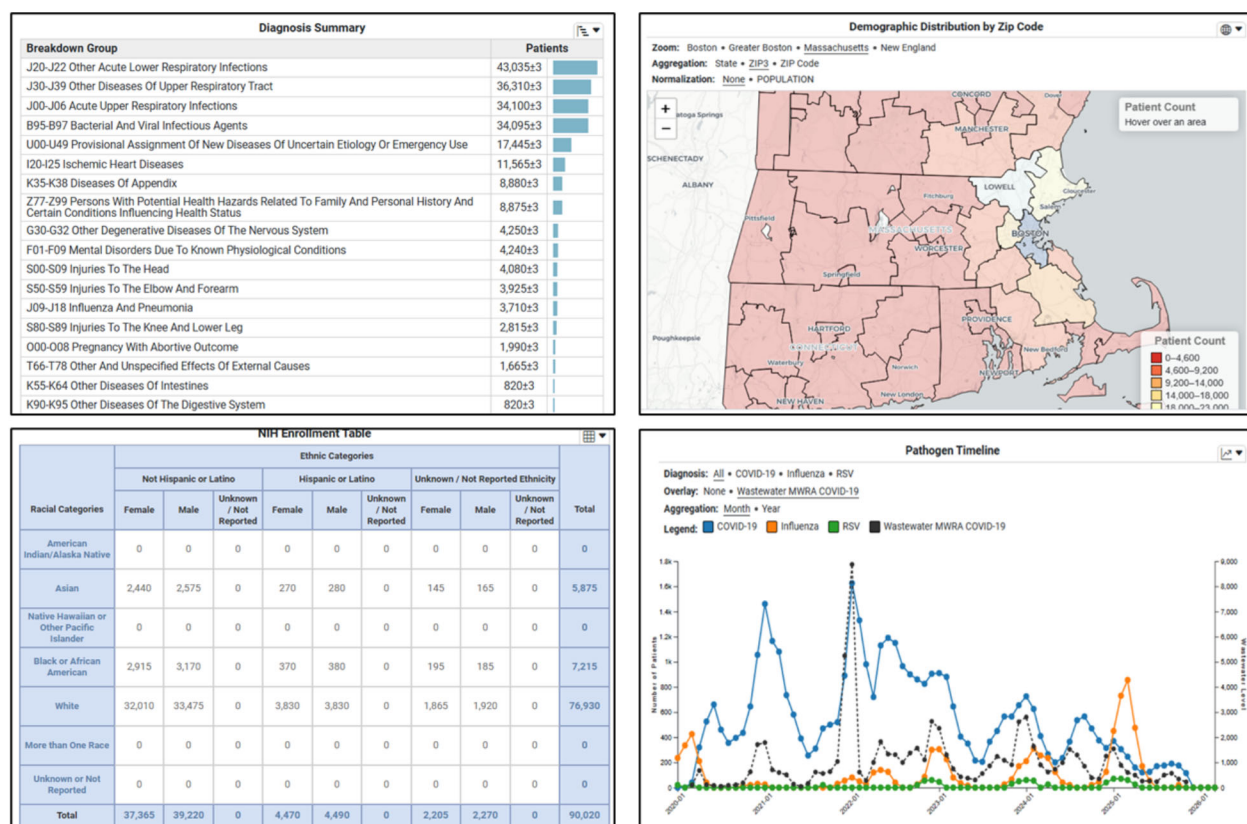


Figure 3. Example interactive data visualizations in Argos. Users can explore patient comorbidities, demographics, geographic location, and pathogen timelines.

Conclusions

The MassCPR Federated Data Network is a unique partnership between academic medical centers and public health agencies in Massachusetts, providing the infrastructure to support both research and public health use cases. The Argos software is freely available as an open-source tool for other OMOP sites to use either internally within their organization or across federated networks.

References

1. Murphy SN, et al. "Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2)." JAMIA. 2010 Mar 1;17(2):124-30.
2. Weber GM, et al. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. J Am Med Inform Assoc. 2009 Sep-Oct;16(5):624-30. doi: 10.1197/jamia.M3191. Epub 2009 Jun 30. PMID: 19567788; PMCID: PMC2744712.