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

Griffin M Weber, MD, PhD

weber@hms.harvard.edu

https://i2b2.org/OMOP/

Griffin M Weber, MD, PhD[1,2], Jeffrey Klann, PhD[3], Shawn N. Murphy, MD, PhD[3,4], Aaron Abend, MBA[5], Diane Keogh[1,6]

[1]Harvard Medical School; [2]Beth Israel Deaconess Medical Center; [3]Mass General Brigham;
[4]University of Washington; [5]Prognosis Data; [6]i2b2 tranSMART Foundation

Acknowledgements & Disclosures

- Harvard Catalyst/CTSA (UL1TR002541 & UM1TR004408)
- Massachusetts Consortium on Pathogen Readiness (MassCPR)
- NIH EnACT (U24TR004111)
- NIH RECOVER (OT2HL161841)
- NIH AIM-AHEAD (OT2OD032581)
- NIH HuBMAP (OT2OD026671)
- 4CE Consortium (3U01HG008685-05S2)
- NIH R01 awards R01GM137410 and R01LM013345
- i2b2-tranSMART Foundation (Dell)
- Tianxi Cai, Xin Xiong, SuChun Cheng, Jeff Klann, Shawn Murphy, Parse Health, Prognosis Data, other i2b2 collaborators

Massachusetts Consortium on Pathogen Readiness (MassCPR)

MASS CPR | Mass Consortium on Pathogen Readiness

https://masscpr.hms.harvard.edu

History: Formed in 2020 to address SARS-CoV-2 and COVID-19, now expanding to other pathogens with pandemic potential

Major Partners: HMS, BIDMC, MGB, BMC, UMass Med School, Tufts, Broad, Ragon, BU/NEIDL, Mass DPH, BPHC, CDC

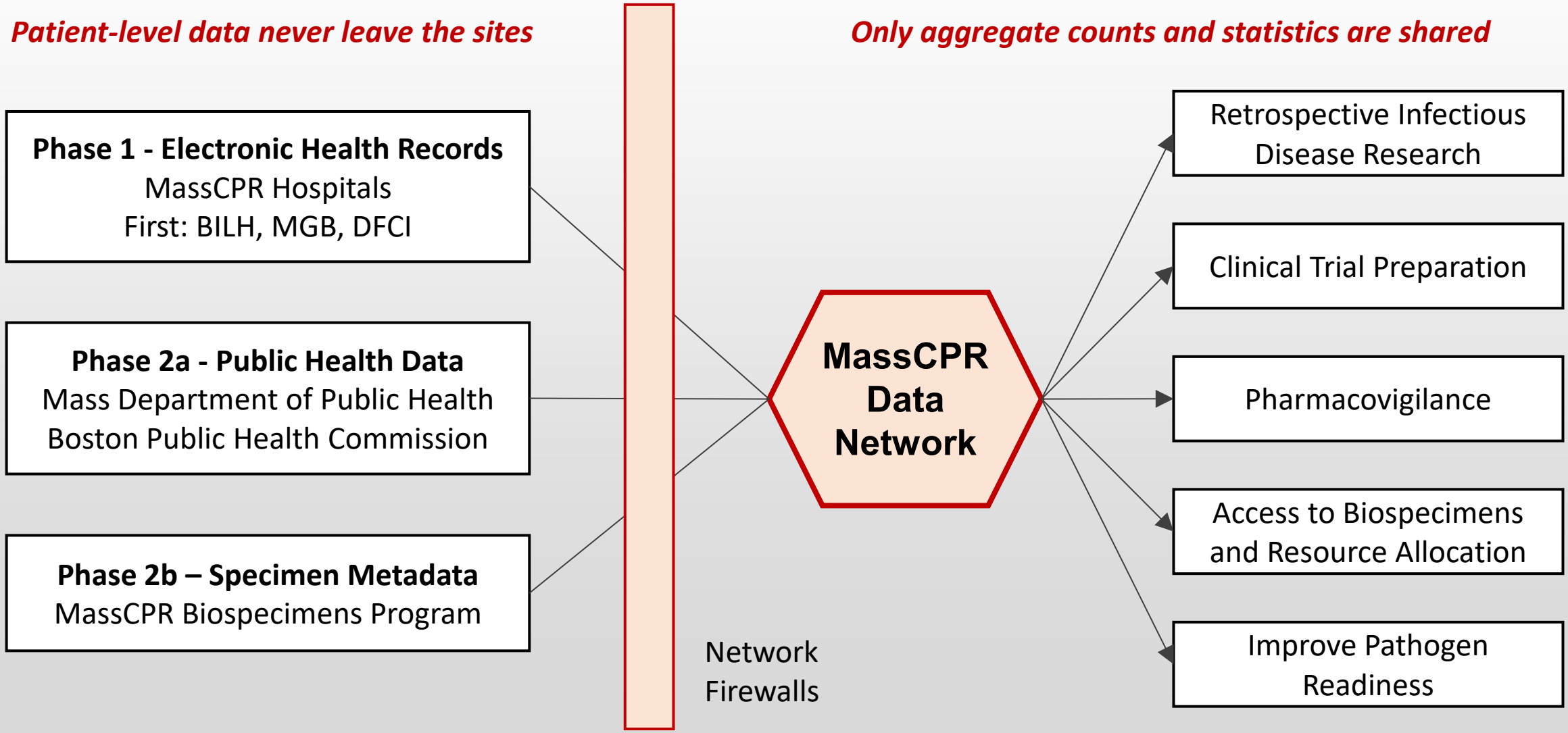
Funding: Awarded funding to 125 research projects since 2020, with 1000+ publications

Programs: Biospecimens and data network

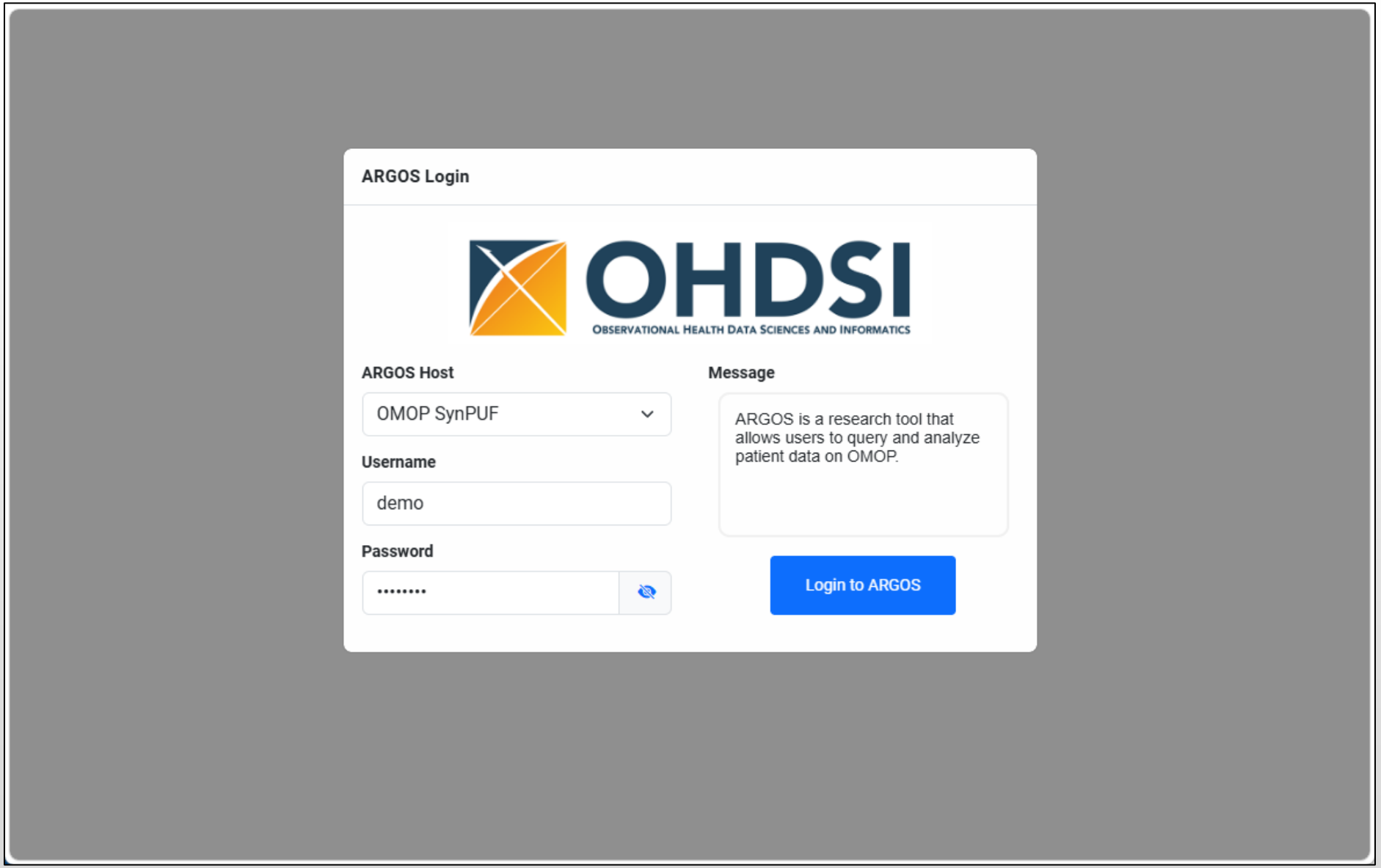


MassCPR Federated Data Network

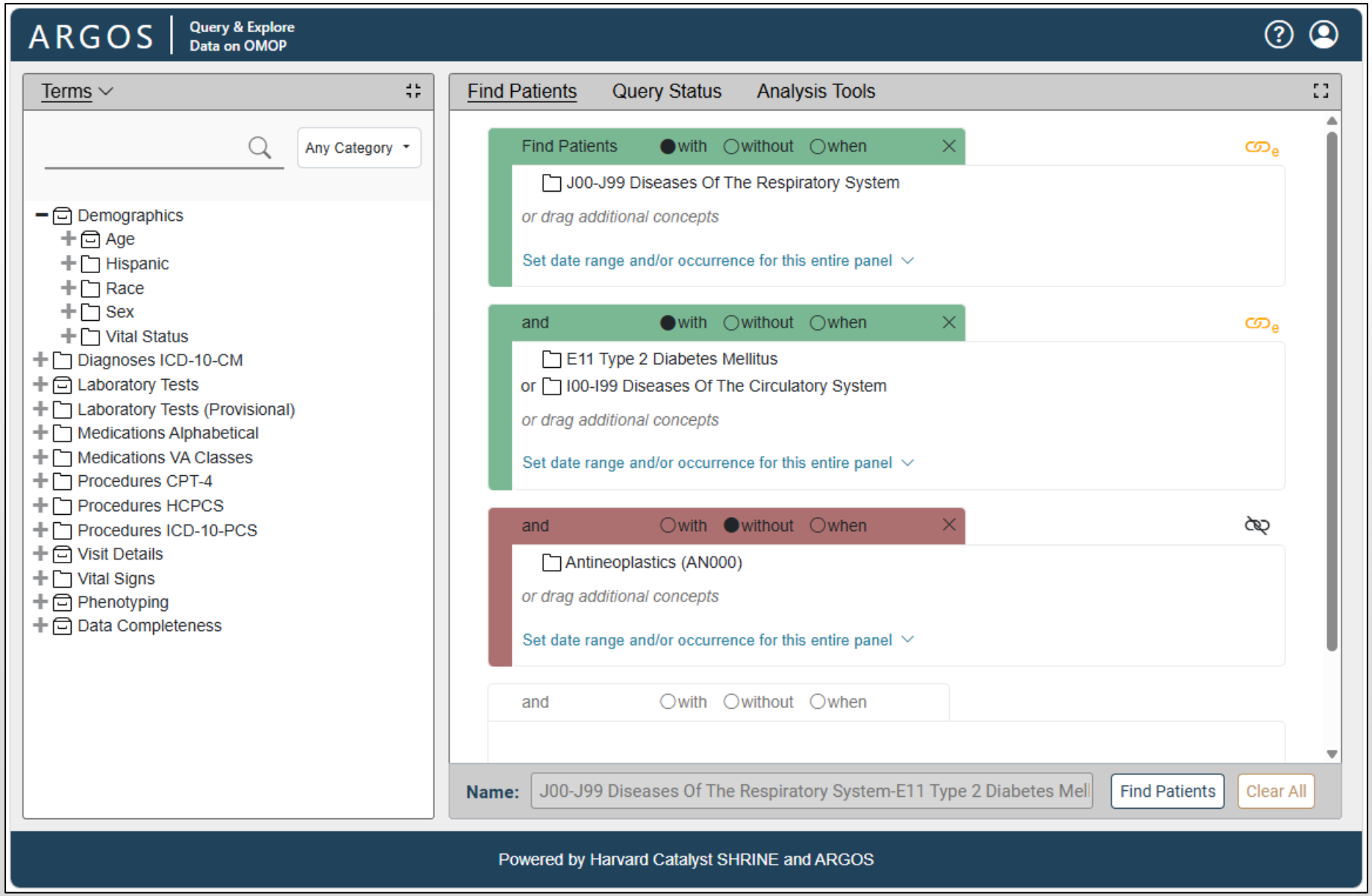
For research and public health use cases related to infectious disease



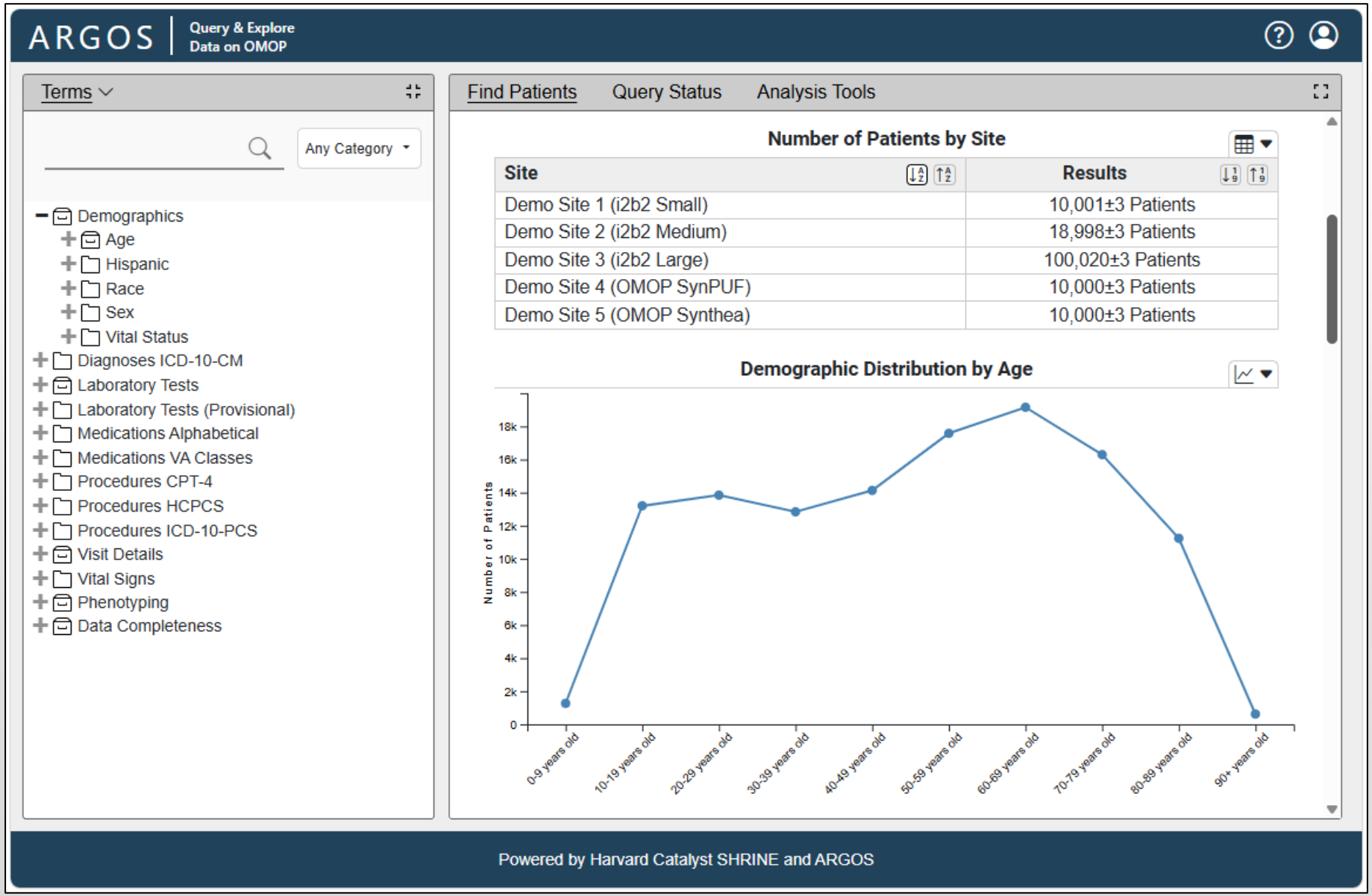
Argos: Login


ARGOS Login
OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS
ARGOS Host
OMOP SynPUF
Username
demo
Password
Message
ARGOS is a research tool that allows users to query and analyze patient data on OMOP.
Login to ARGOS

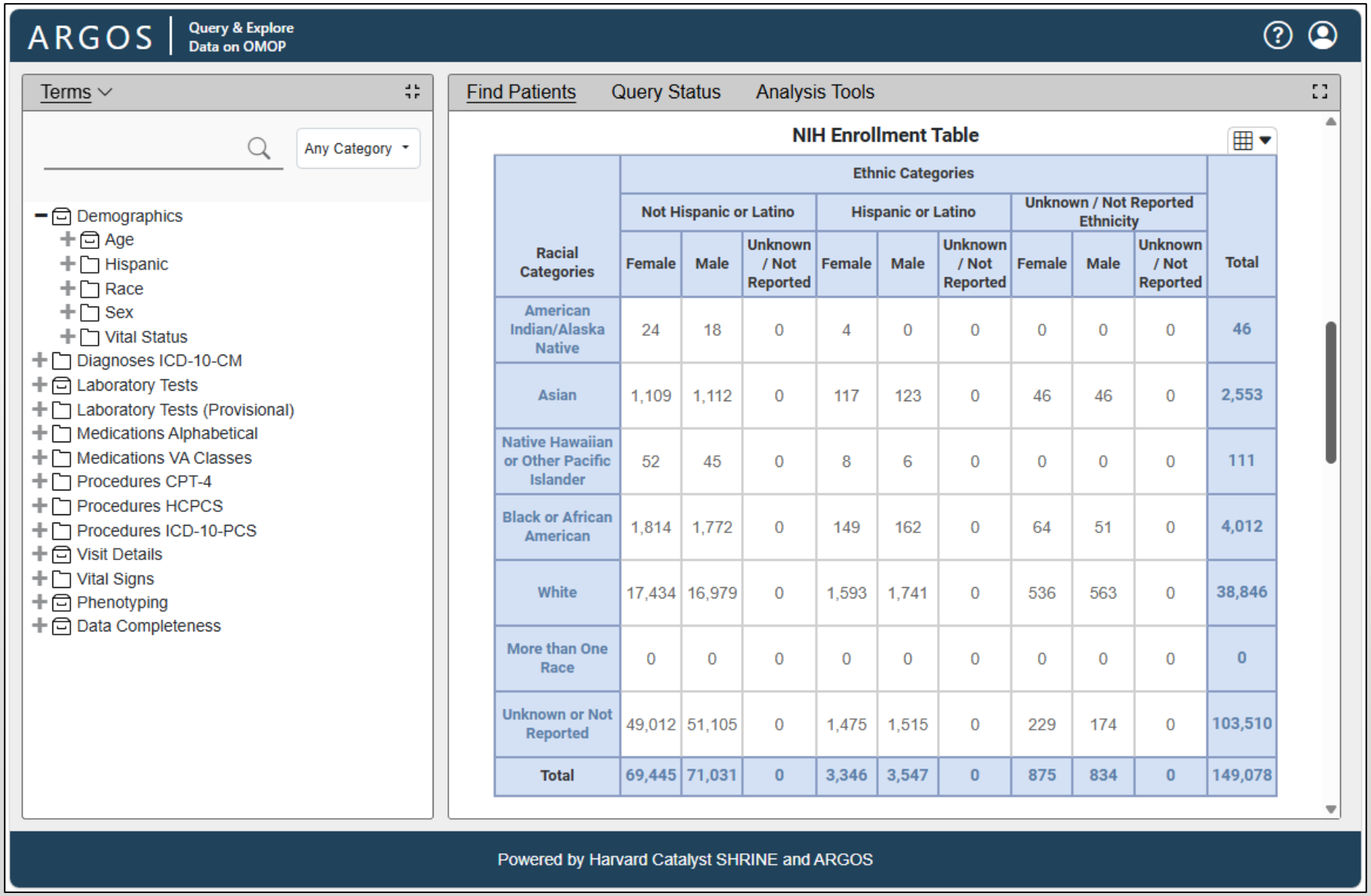
Argos: Specify Search Criteria



Argos: Site Counts and Demographic Distributions



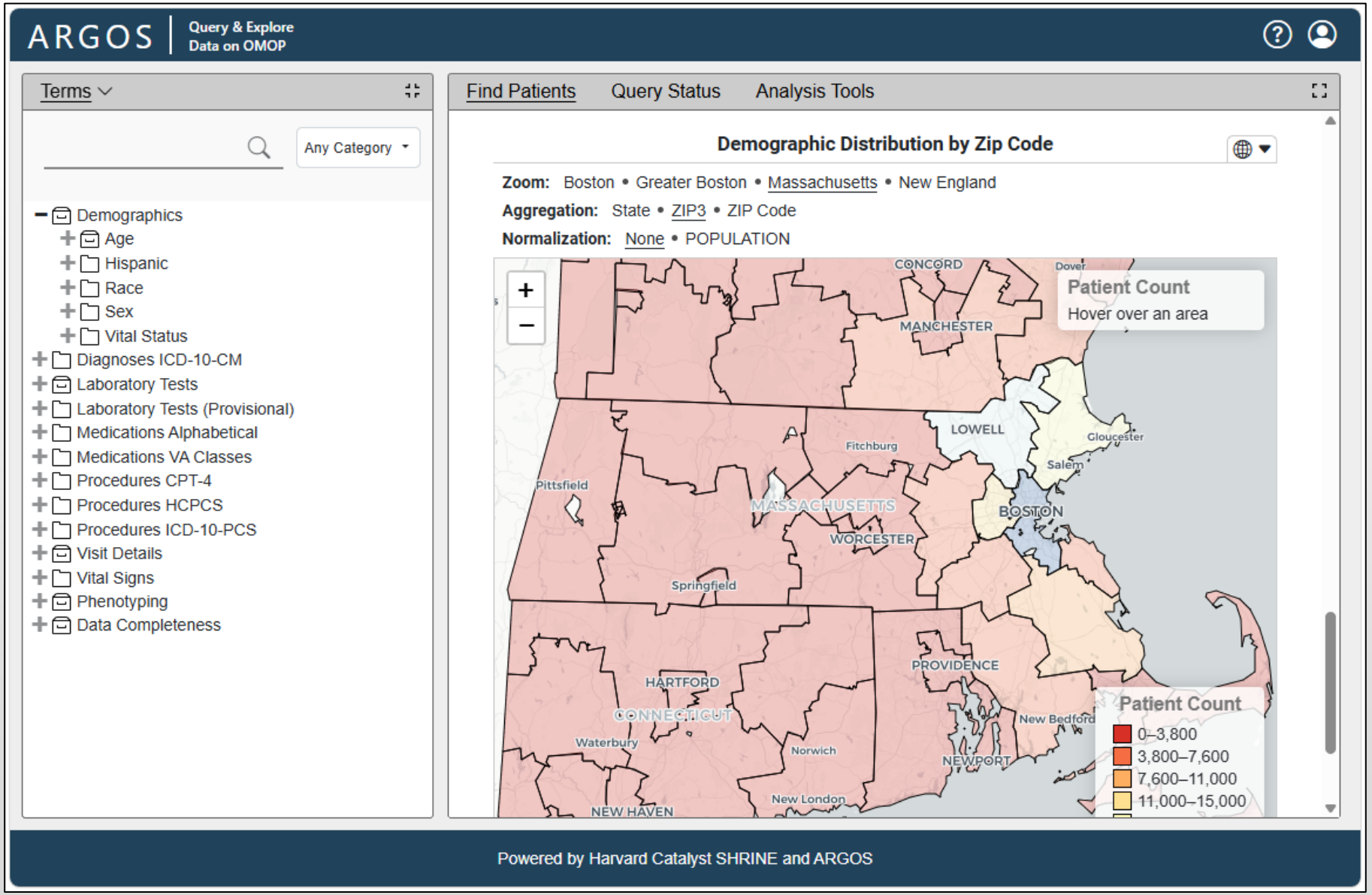
Argos: NIH Enrollment Table



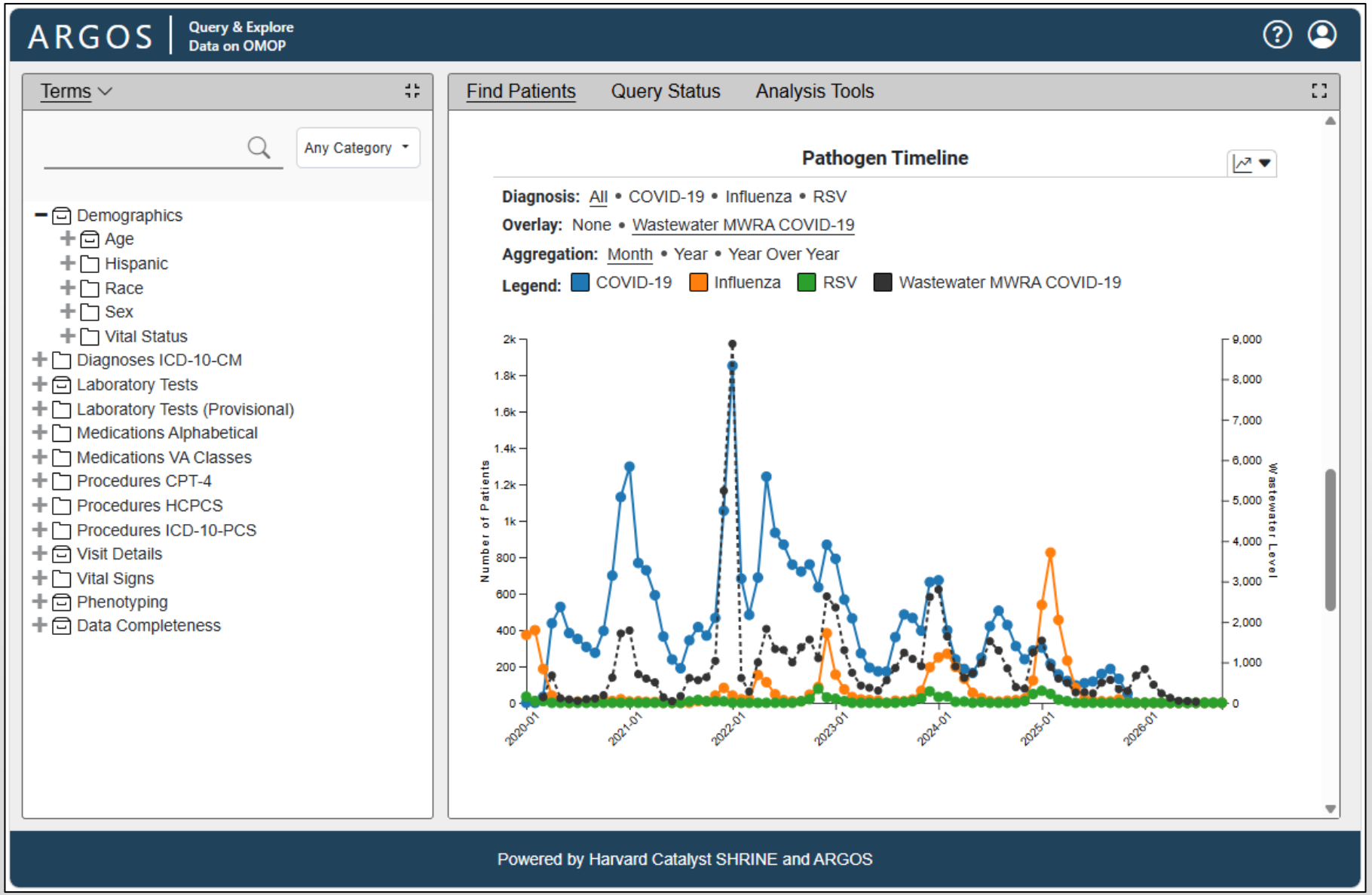
NIH Enrollment Table

Racial Categories	Ethnic Categories									Total
	Not Hispanic or Latino			Hispanic or Latino			Unknown / Not Reported Ethnicity			
	Female	Male	Unknown / Not Reported	Female	Male	Unknown / Not Reported	Female	Male	Unknown / Not Reported	
American Indian/Alaska Native	24	18	0	4	0	0	0	0	0	46
Asian	1,109	1,112	0	117	123	0	46	46	0	2,553
Native Hawaiian or Other Pacific Islander	52	45	0	8	6	0	0	0	0	111
Black or African American	1,814	1,772	0	149	162	0	64	51	0	4,012
White	17,434	16,979	0	1,593	1,741	0	536	563	0	38,846
More than One Race	0	0	0	0	0	0	0	0	0	0
Unknown or Not Reported	49,012	51,105	0	1,475	1,515	0	229	174	0	103,510
Total	**69,445**	**71,031**	**0**	**3,346**	**3,547**	**0**	**875**	**834**	**0**	**149,078**

Argos: Spatial Distributions



Argos: Pathogen Timelines (EHR + Wastewater Data)

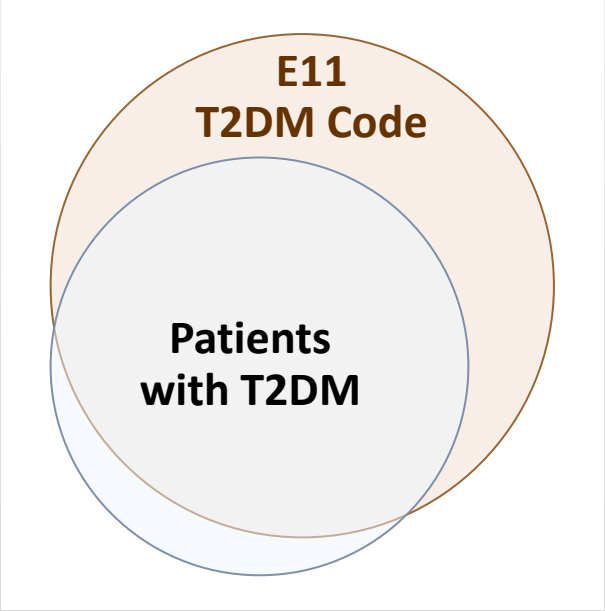


Automatic “Silver-Standard” Phenotype Generator in Argos

Separate AI/ML phenotype models are created for 1000+ different medical conditions

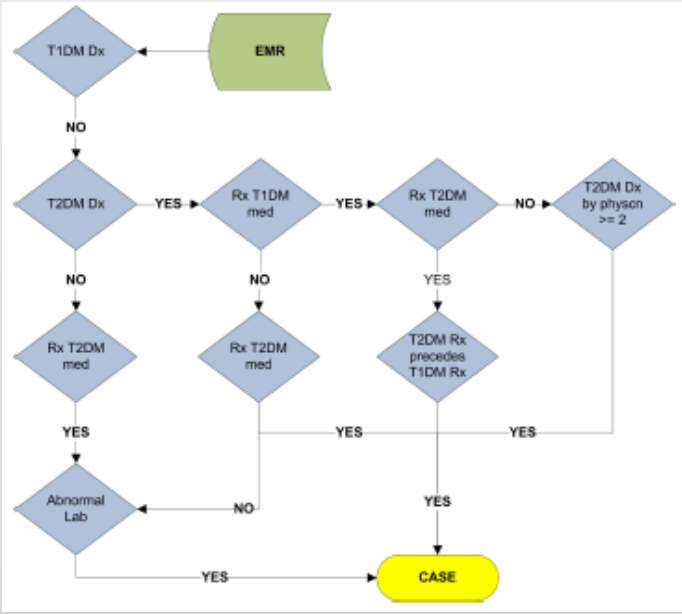
Raw Diagnosis Codes

Low precision for predicting the patient’s true medical condition or “phenotype”



Expert-Based Phenotypes

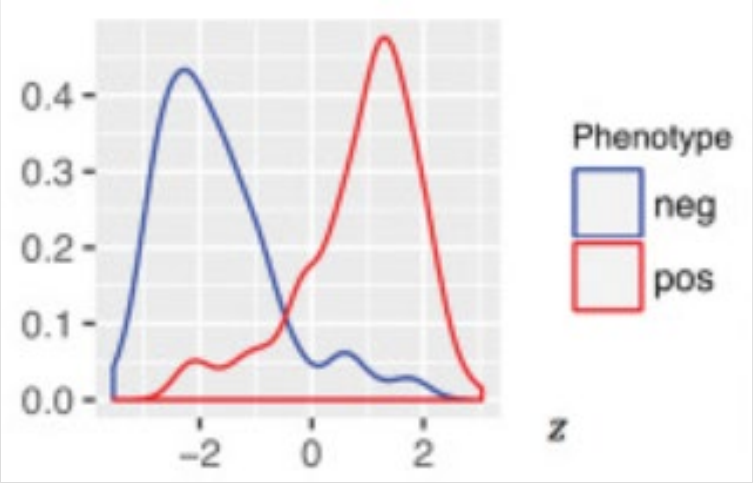
“Gold-Standard” rule-based models; slow and expensive to implement and maintain



From eMERGE

Probabilistic Phenotypes

AI/ML algorithms automatically estimate the likelihood that a patient has the medical condition



Model Score

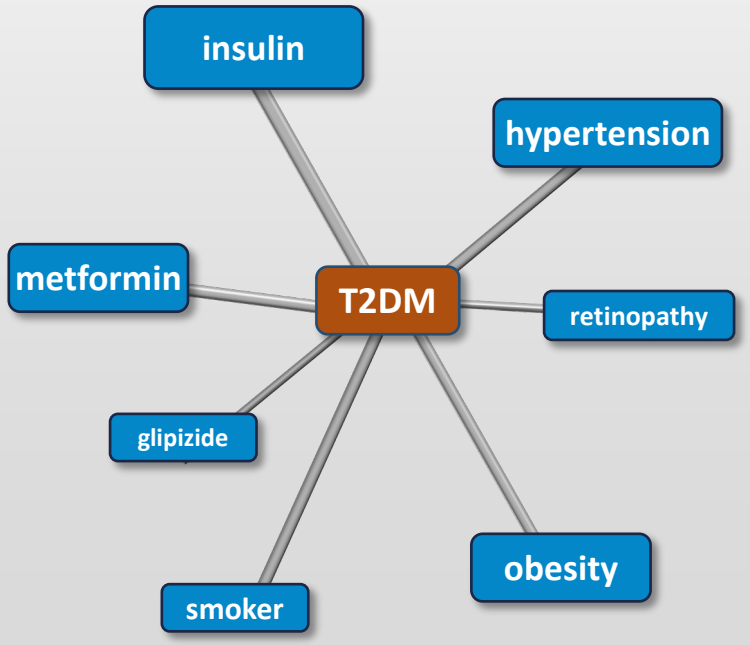
Yu S, et al. JAMIA. 2018 Jan 1;25(1):54-60

Automatic “Silver-Standard” Phenotype Generator in Argos

Separate AI/ML phenotype models are created for 1000+ different medical conditions

Knowledge Graph Derived from EHR

Embedding regression identifies ~200 features that co-occur often with the condition in the EHR



Hong C, et al. NPJ Digit Med. 2021 Oct 27;4(1):151

AI/ML Generated Probabilistic Models

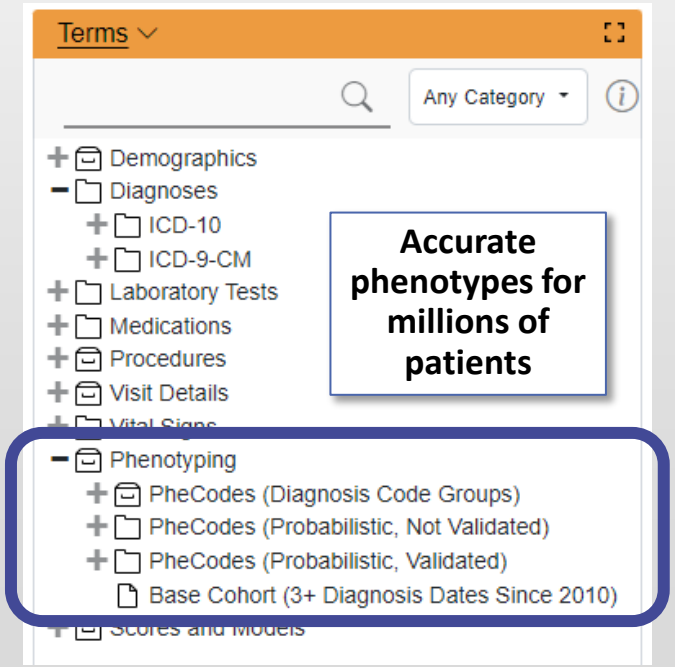
Weakly-supervised machine learning assigns weights to those features to build a model

Feature	Weight
Primary features	
Type 2 diabetes	0.5804
Utilization ICD Dates	-0.3662
Additional top features	
metformin	0.0846
Hypertension	0.0615
Insulin pump user	0.0409
Diabetic retinopathy	0.0403
Smoker	-0.0228
BMI 30 plus	-0.0255
Age 65 plus	-0.0366
Female	-0.0613

Liao KP, et al. JAMIA. 2019 Nov 1;26(11):1255-1262

Computed Medical Conditions in Argos

Patients whose model score is above a threshold is listed in Argos as having that phenotype



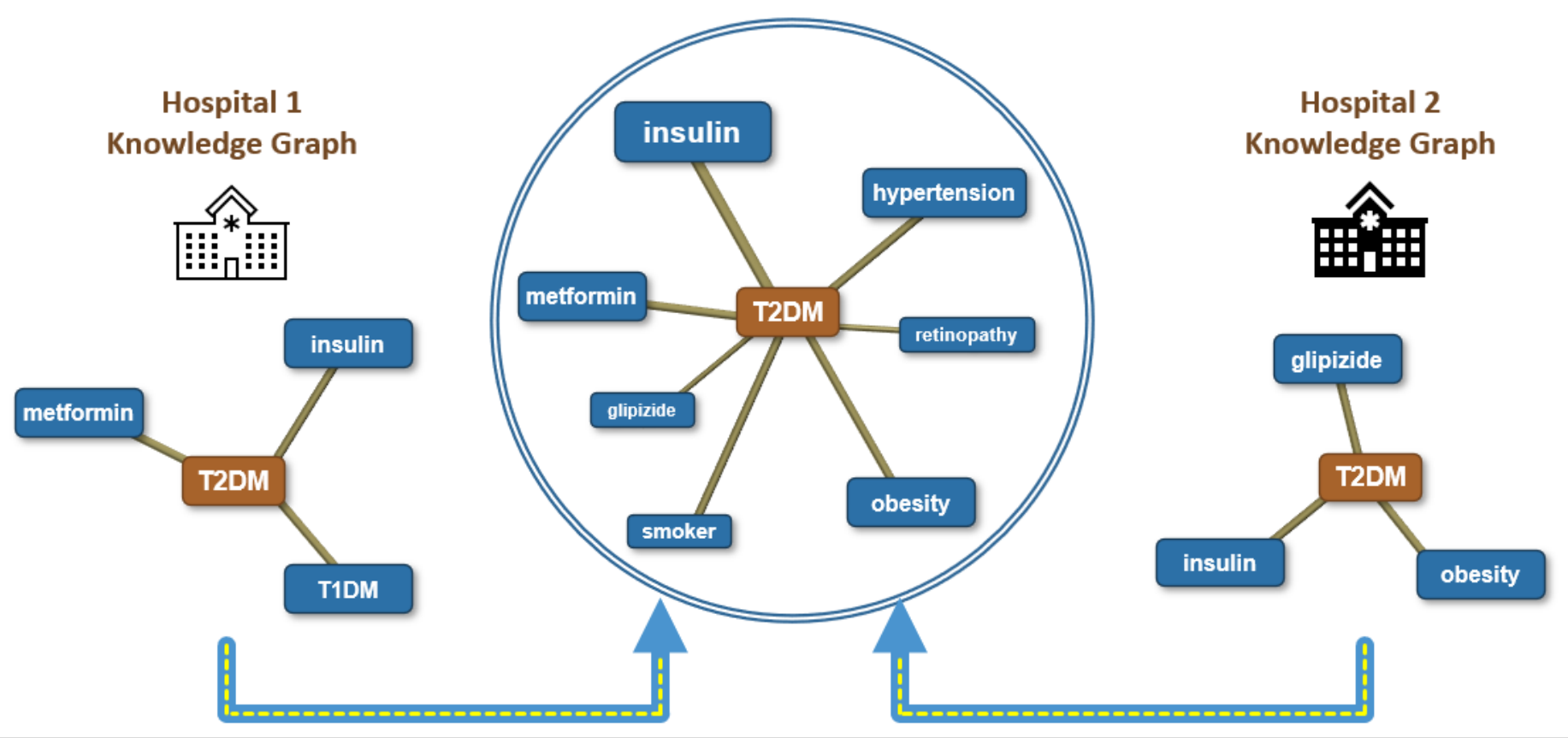
Implementation of Computed Phenotypes Across Sites

Site	Medical Conditions with Computed Phenotypes	Patients with 1+ Phenotypes	Total Patient-Phenotype Pairs
Mayo Clinic	1,604	2,566,963	53,500,000
Mass General Brigham	1,403	2,221,712	27,424,914
University of Florida	1,186	1,082,522	15,997,414
Beth Israel Lahey Health	983	619,596	6,952,004
University of Pittsburgh	69	1,577,078	4,593,142
UT Health Houston	69	764,504	1,582,305

No human effort is needed to train the models (1000+ phenotypes in a few days). These “silver-standard” phenotypes do not replace expert determined “gold-standard” phenotypes but can fill in the gaps for hundreds of medical conditions without an existing phenotype and provide accurate comorbidity profiles.

Federated learning for robust models without overfitting

Sites only share aggregate counts and summary statistics



Argos to Atlas Bridge

Copy Argos cohort definitions into Atlas

