

Mount Sinai Data Warehouse Town Hall

Patricia Kovatch

Farhan Mahmood

Sharon Nirenberg, MD

Timothy Quinn, PhD

Scientific Computing

Icahn School of Medicine at Mount Sinai

November 17, 2022



Agenda

- ▶ MSDW Overview
- ▶ MSDW Operations in 2022
- ▶ Data Integrations: IRW and PowerPath
- ▶ Concept Mapping for OMOP
- ▶ New Services: Interactive Dashboards & Geocoding
- ▶ MSDW Roadmap

MSDW Overview

The MSDW Team



Patricia Kovatch
Professor & Dean for
Scientific Computing



Sharon Nirenberg MD
Physician Informaticist



Praveen Medabalmi MD
Physician Informaticist



Farhan Mahmood
Director Scientific
Computing



Timothy Quinn PhD
Principal Data Architect



Shivaji Punukollu
Healthcare Data Engineer



Jacob Weiser
Healthcare Data Engineer



Darius Boopal
Healthcare Data Engineer



Priyal Mehta
Healthcare Data Analyst



Rupan Hossain
Database Administrator



Manoj Chekuri
Healthcare Data Engineer



AJ Caberto
Clinical Data Specialist



Jiani Xiang
Clinical Data Specialist

MSDW Relaunched in September 2021

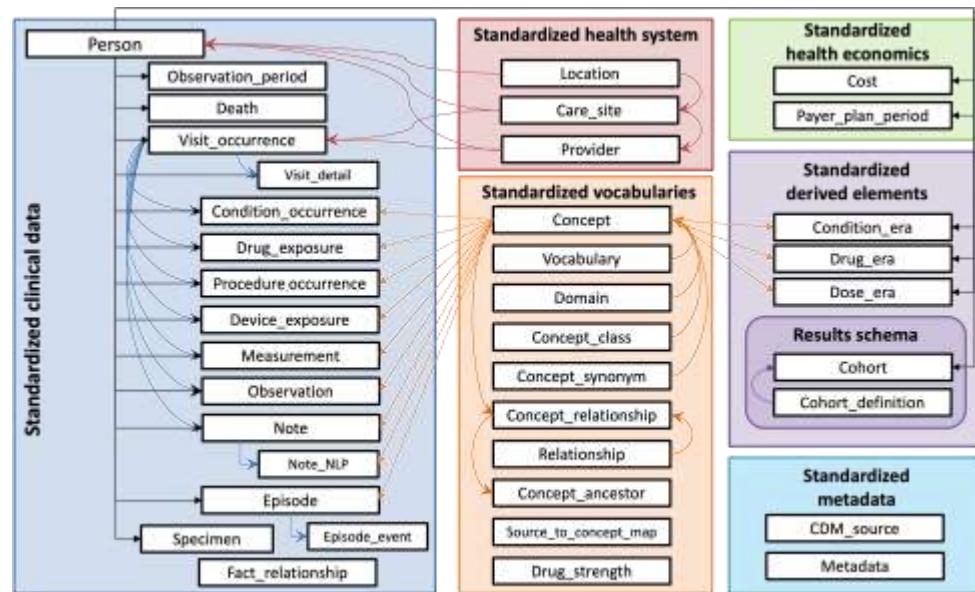


MSDW Evolution	Benefits
Epic as Primary Data Source	<ul style="list-style-type: none">• Higher data quality• Transparent data lineage• Access to more data elements
OMOP Common Data Model	<ul style="list-style-type: none">• The <i>de facto</i> standard data model for research• Standardized structure & content• Vibrant research community & tools• Easier data sharing with research networks
Restore Daily Refreshes	<ul style="list-style-type: none">• Faster access to data• Full history of all changes
MSDW on Minerva	<ul style="list-style-type: none">• MSDW co-located with Data Ark and other research data sets

OMOP Common Data Model (v5.4)



- ▶ Observational Medical Outcomes Partnership (OMOP)
 - Public-private partnership
 - 5-year program for surveillance of medical products
 - Claims & EHR data
 - Chaired by FDA during 2009-2013
 - Transitioned to OHDSI consortium in 2013
- ▶ Common Data Model (CDM)
 - Standardizes both structure & content for interoperability
 - Now used by many regional, national, and international research programs & networks

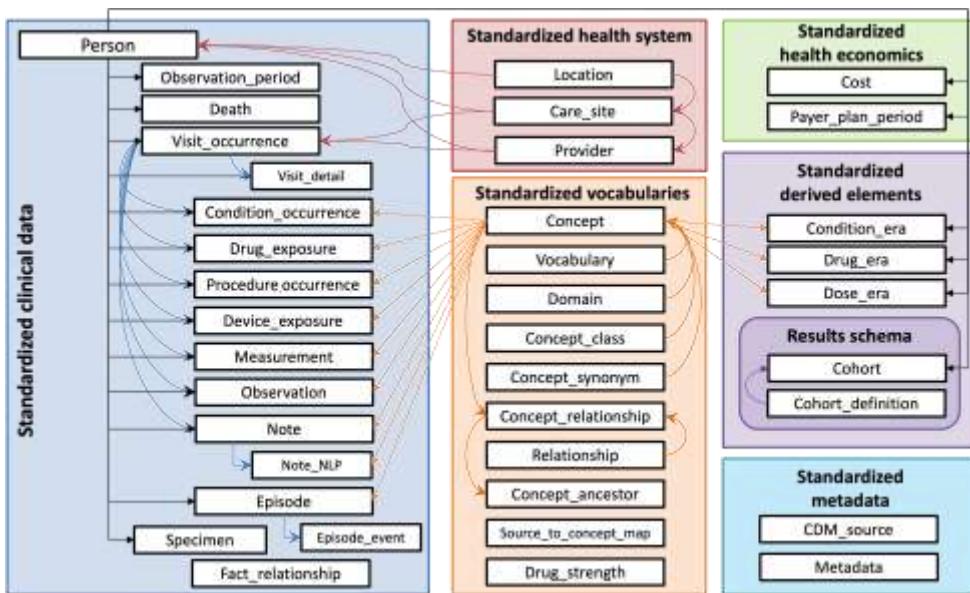


<https://ohdsi.github.io/CommonDataModel/>

OMOP Common Data Model Requirements



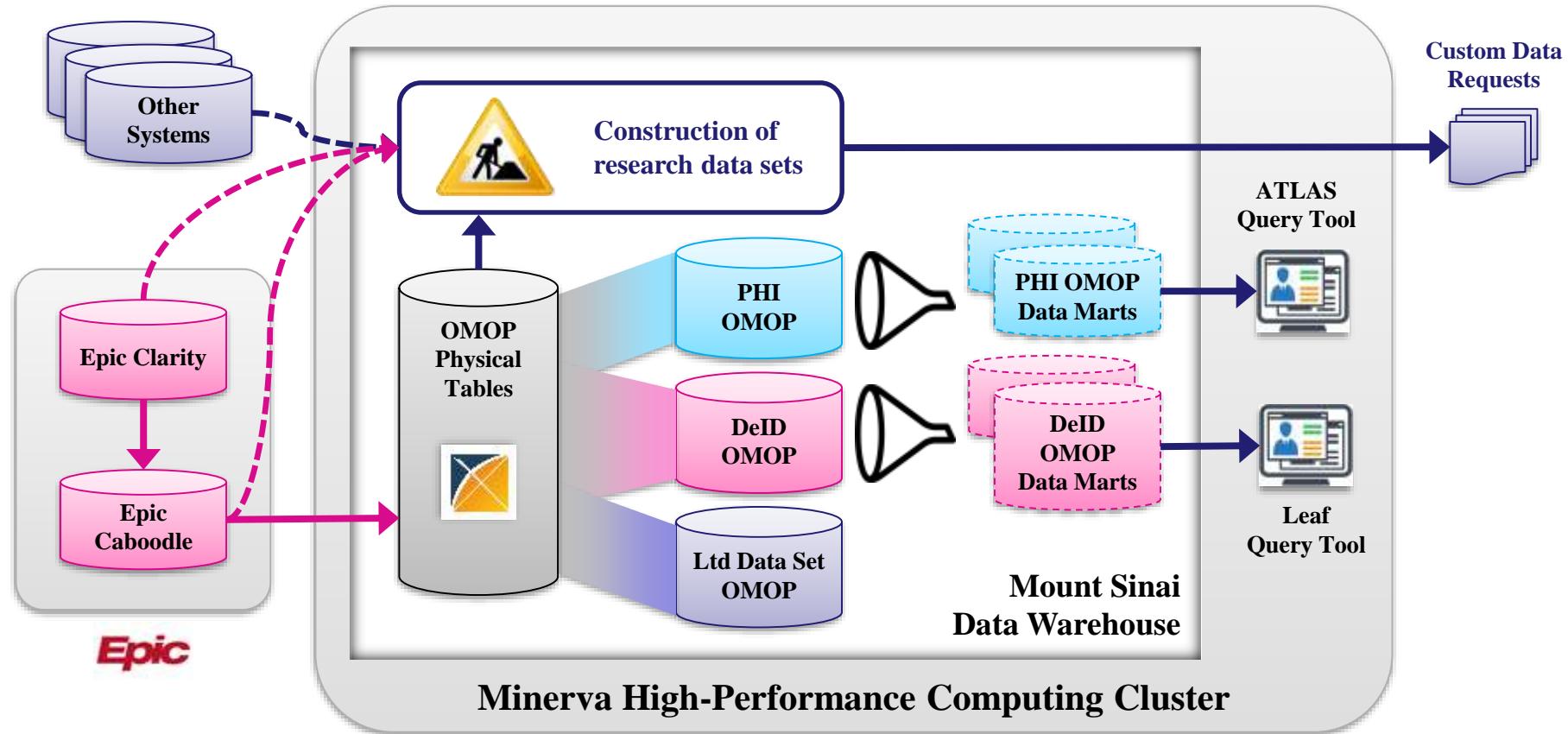
1. Standardize **data structure** via common format



2. Standardize **data content** via mapping EHR codes to standard healthcare vocabularies

OMOP Domain	Standard Vocabularies	Non-standard Vocabularies
Condition	SNOMED-CT	ICD-10-CM, ICD-9-CM
Drug	RxNorm, CVX	ATC, NDC, Multum
Measurement	LOINC	SNOMED-CT, Nebraska Lexicon
Procedure	CPT4, HCPCS, ICD-10-PCS	ICD-9-Proc
Observation	SNOMED-CT, LOINC	ICD-10-CM, ICD-9-CM
Race, Ethnicity	OMOP Race, OMOP Ethnicity	SNOMED-CT, Nebraska Lexicon
Provider (Specialty)	NUCC, Medicare Specialty	SNOMED-CT, Nebraska Lexicon
Route	SNOMED-CT	Nebraska Lexicon
Unit	UCUM	SNOMED-CT, Nebraska Lexicon

Mount Sinai Data Warehouse on Minerva HPC Cluster



Minerva High-Performance Computing (HPC) Cluster

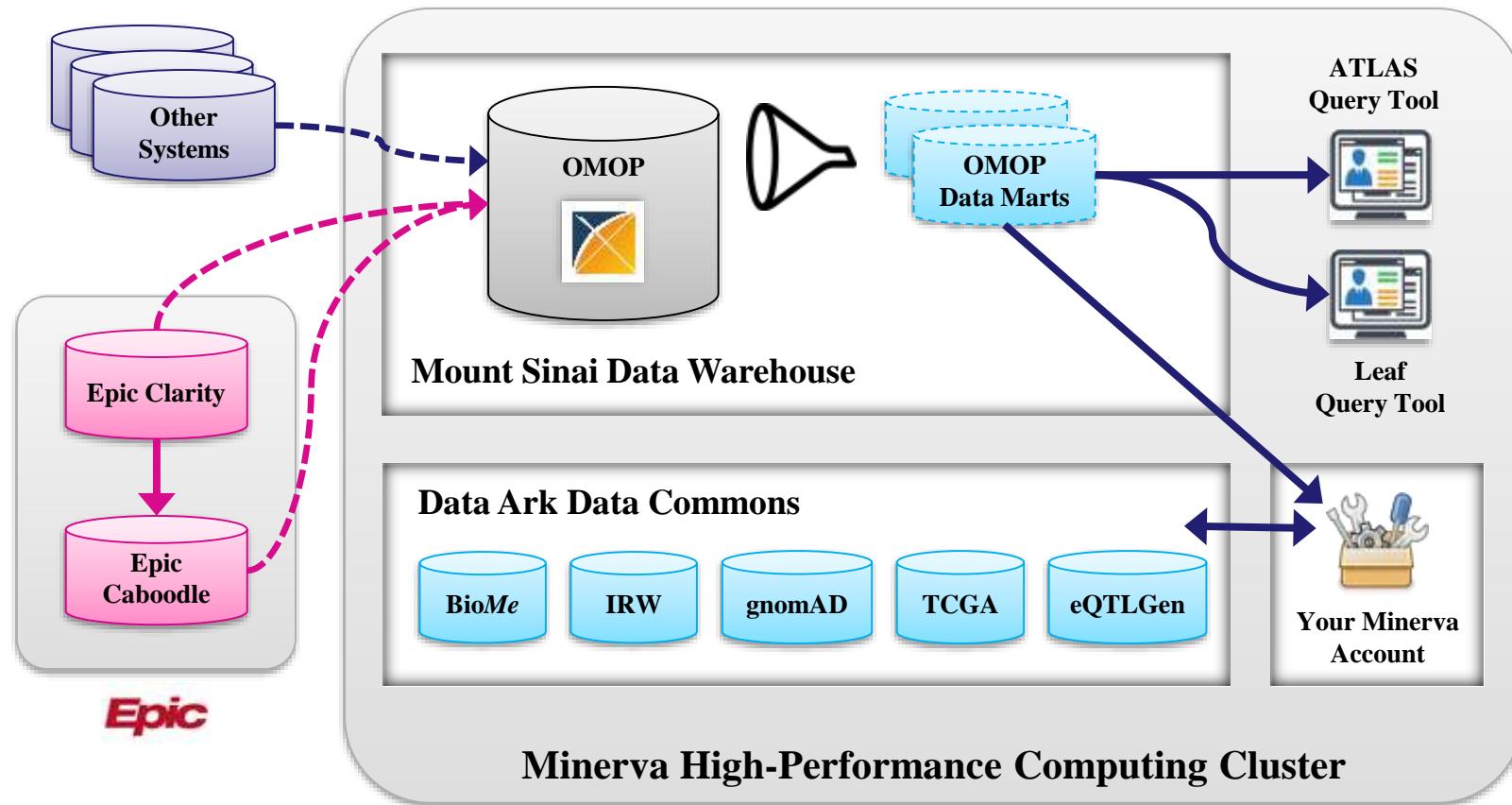


Servers hosting
Mount Sinai Data
Warehouse (MSDW)
“virtual machines”

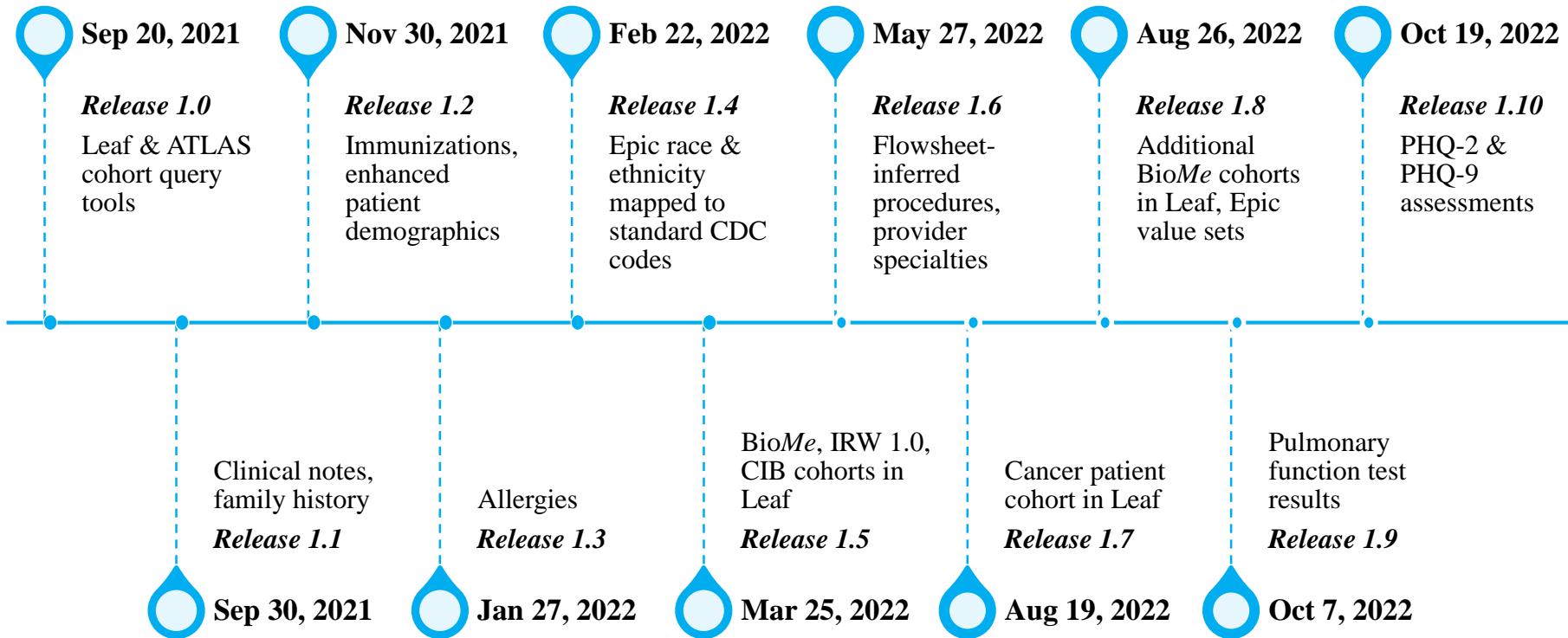


Hess Center for Science and Medicine
1470 Madison Ave, New York, NY 10029

Computational & Data Science Ecosystem



Release Timeline for MSDW



MSDW First Release 1.0 – September 20, 2021

Record Types	OMOP Tables	Record Types	OMOP Tables
Patient Master File	person death	Inpatient Hospitalization Inpatient Hospitalization from ED Visit Hospital Outpatient Visit ED Visit Outpatient Visit Telehealth Visit Mobile Unit Encounter	visit_occurrence
Provider Master File	provider		
Org Hierarchy Master File	care_site		
Patient Home Address Provider Office Address Facility/Clinic Address	location	Encounter Diagnosis Problem List Hospital Problem	condition_occurrence
Diagnosis Master File Medication Master File Procedure Master File Lab Test Master File Flowsheet Metrics Master File Reference Data Files	concept concept_class vocabulary domain	Lab Test Results Vital Signs	measurement
Concept Mappings	concept_relationship	General Procedures Surgical Procedures	procedure_occurrence
Medication Orders Medication Dispenses Medication Administration Record	drug_exposure	Patient Demographics Past Medical History Surgical History Family History Social History	observation

MSDW Releases

Release	Release Date	Record Types	OMOP Tables
1.1	Sep 30, 2021	Clinical Notes Family History	note observation
1.2	Nov 30, 2021	Immunizations Patient Demographics	drug_exposure observation
1.3	Jan 27, 2022	Allergies	observation
1.4	Feb 22, 2022	CDC Race & Ethnicity Mount Sinai Combined Race & Ethnicity Variable	observation
1.5	Mar 25, 2022	BioMe Biobank Patient Cohort	observation
1.6	May 27, 2022	Flowsheet-inferred Procedures Provider Specialties	procedure_occurrence provider_attribute_xtn
1.7	Aug 19, 2022	Cancer Patient Cohort	observation
1.8	Aug 26, 2022	BioMe Biobank Patient Sub-cohorts Epic Value Sets	observation concept_relationship
1.9	Oct 7, 2022	Flowsheet Measurements	measurement
1.10	Oct 19, 2022	Flowsheet Observations	observation

MSDW Operations in 2022

MSDW Statistics- Tickets in 2022

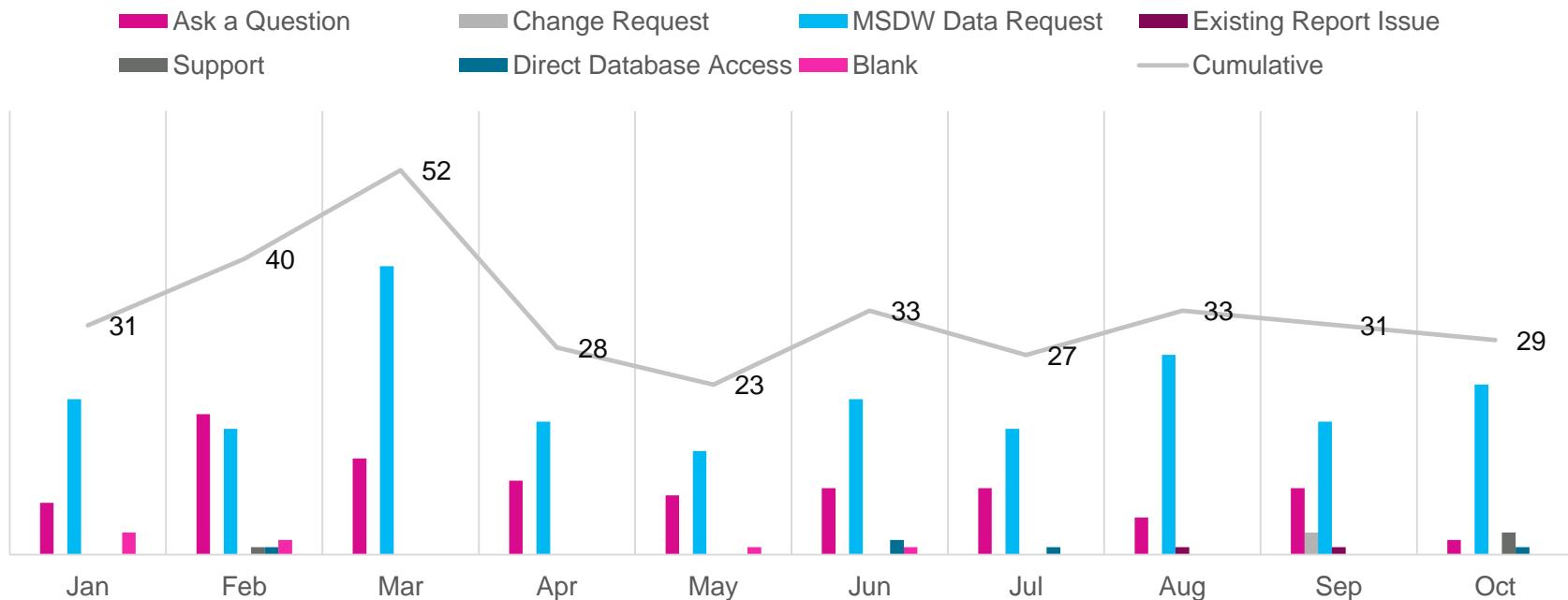
- ▶ **Total MSDW Tickets: 327**

- January 1, 2022 to October 25, 2022

Ticket Type Breakdown:	
Ticket Type	Quantity
MSDW Data Request	215
Ask A Question	90
MSDW Direct Database Access	4
Change Request	3
Existing Report Issue	2
Other Questions	1
Blank	8

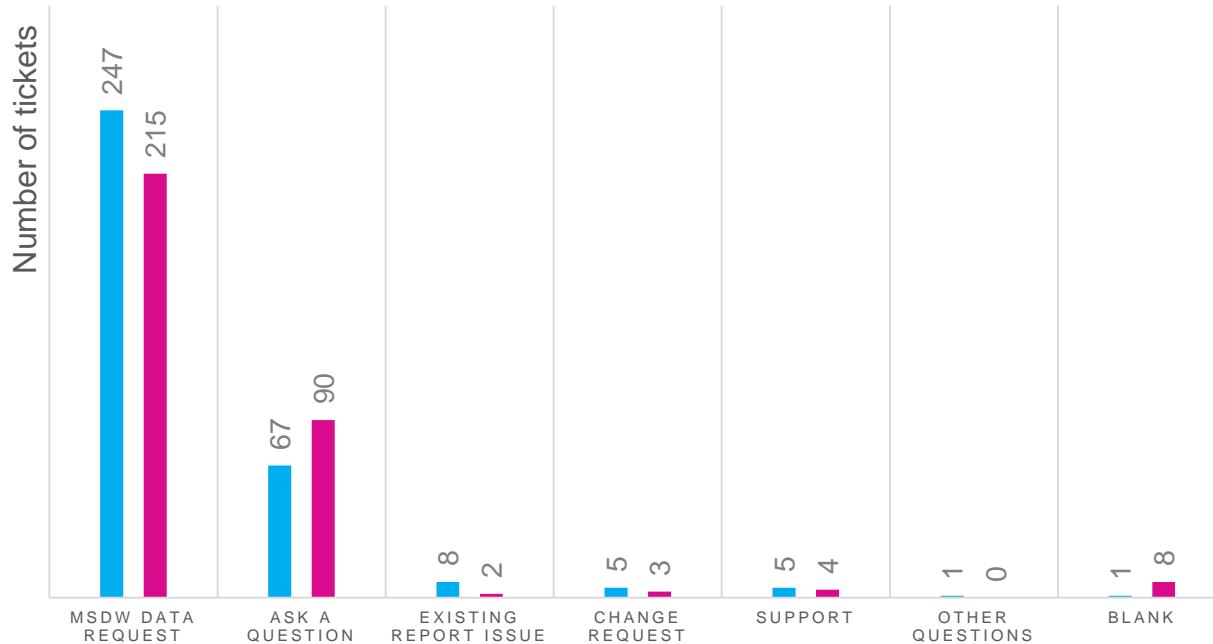
MSDW Statistics- Tickets in 2022 By Month

MSDW Tickets by Month and Type



MSDW Statistics- Tickets in 2021 and 2022

MSDW Ticket Volume by Request Type

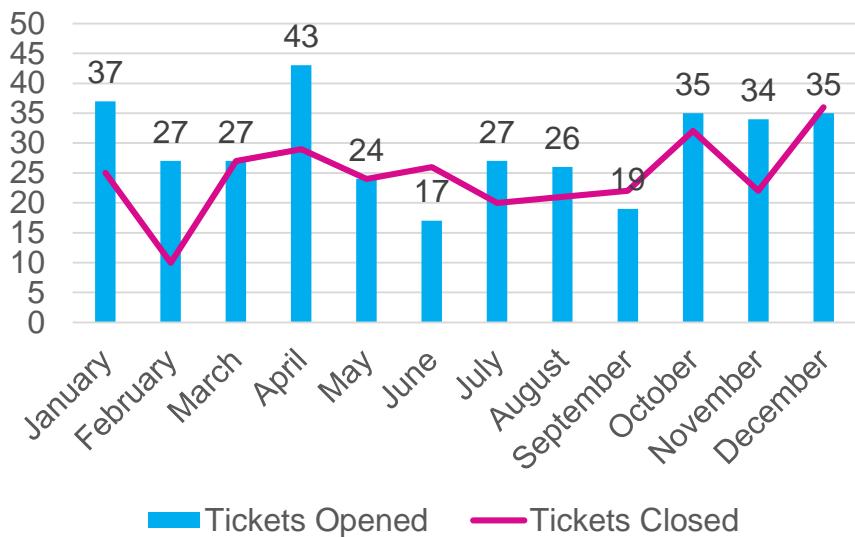


2021 Tickets: 335
2022 Tickets: 327

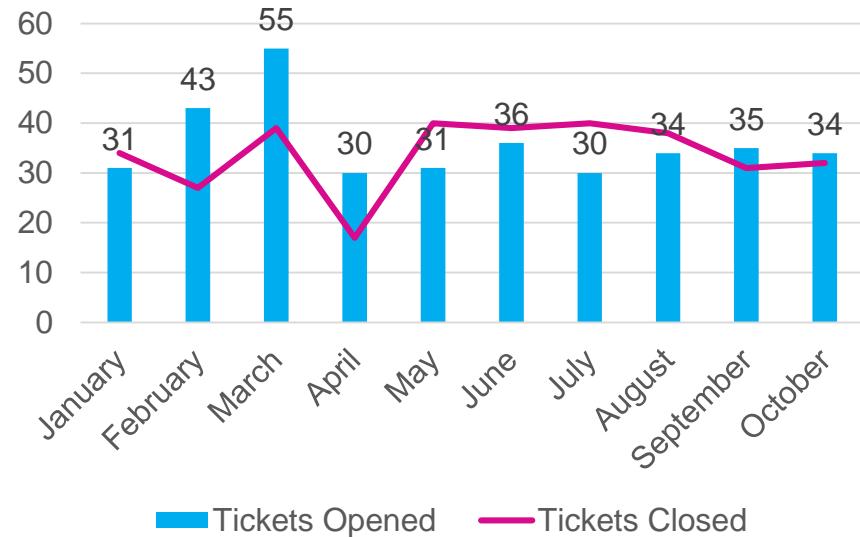
2021 Date Range: January 1, 2021 to December 31, 2021
2022 Date Range: January 1, 2022 to October 25, 2022

MSDW Statistics- Ticket Volume: Open and Resolved

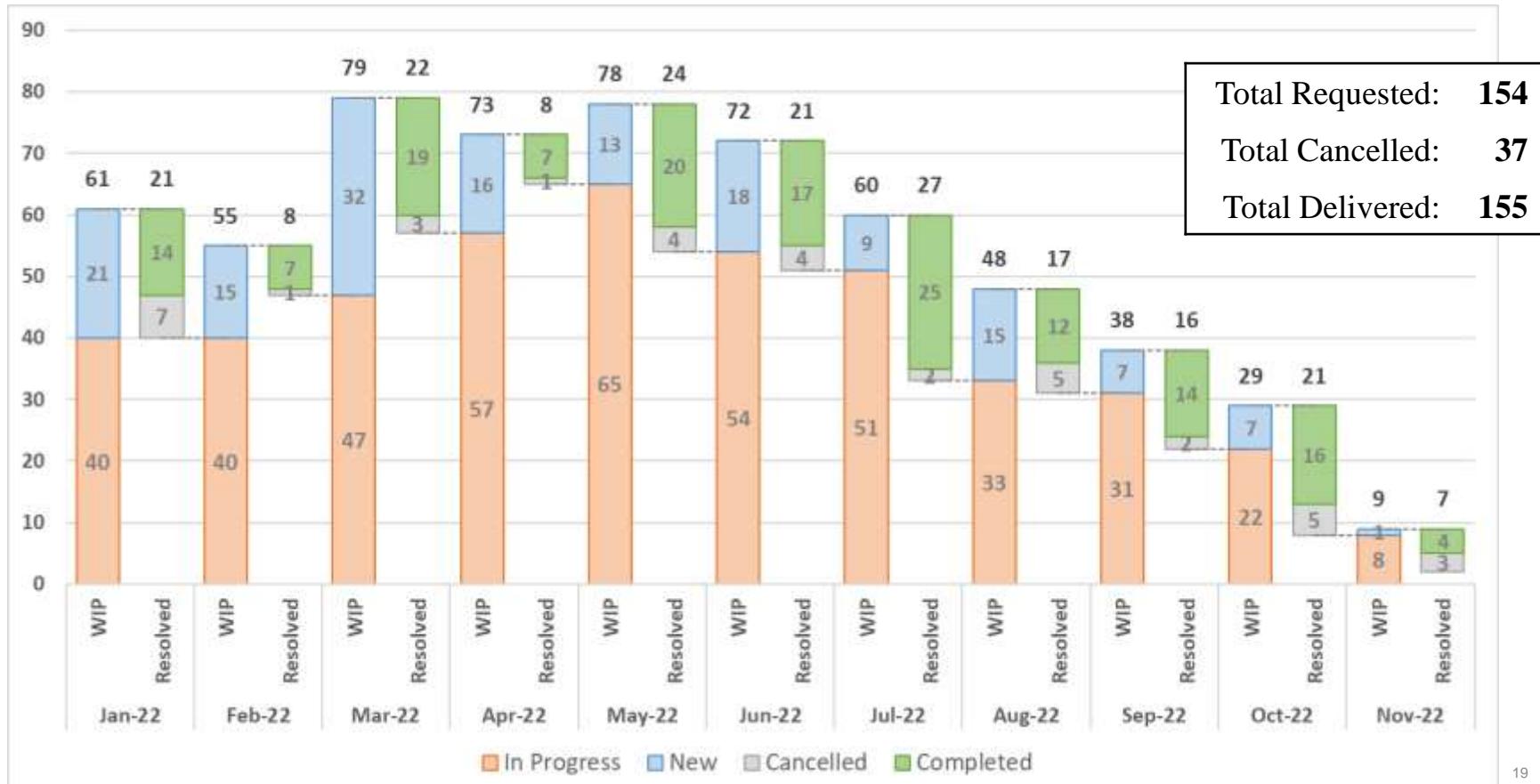
2021 MSDW Tickets Opened and Closed by Month



2022 MSDW Tickets Opened and Closed by Month



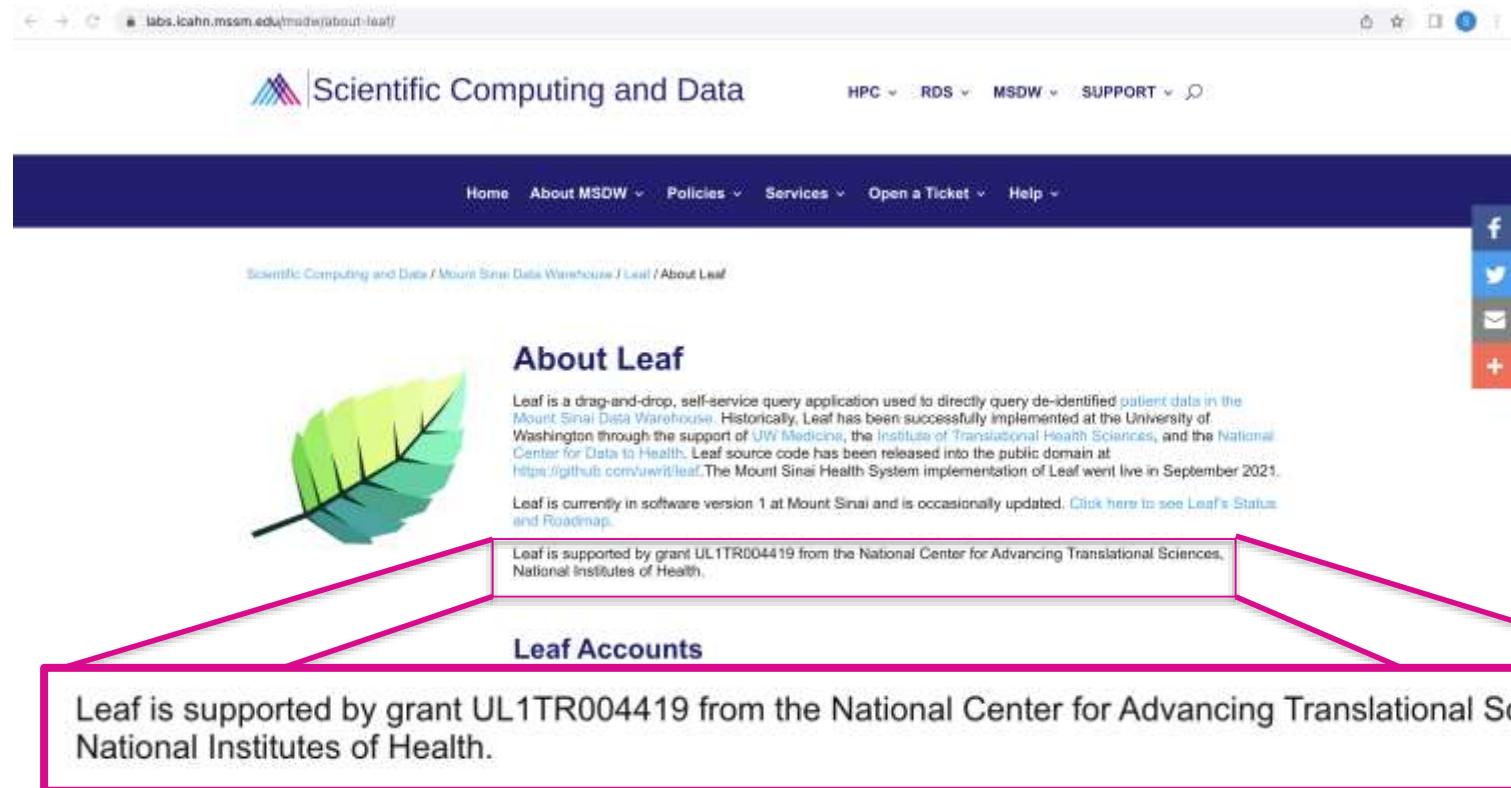
Custom Data Sets Requested & Delivered in 2022 (YTD)



Utilization of Leaf and ATLAS is Increasing

	Leaf	ATLAS
Unique users in last 12 months	120	88
Unique users in last 6 months	99	48
Cohorts created in last 12 months	1060	37
Cohorts created in last 6 months	908	16

Encourage MSDW Users to Acknowledge CTSA



Scientific Computing and Data

Home About MSDW Policies Services Open a Ticket Help

Scientific Computing and Data / Mount Sinai Data Warehouse / Leaf / About Leaf

About Leaf

Leaf is a drag-and-drop, self-service query application used to directly query de-identified patient data in the Mount Sinai Data Warehouse. Historically, Leaf has been successfully implemented at the University of Washington through the support of UW Medicine, the Institute of Translational Health Sciences, and the National Center for Data to Health. Leaf source code has been released into the public domain at <https://github.com/ctsa/Leaf>. The Mount Sinai Health System implementation of Leaf went live in September 2021.

Leaf is currently in software version 1 at Mount Sinai and is occasionally updated. [Click here to see Leaf's Status and Roadmap](#).

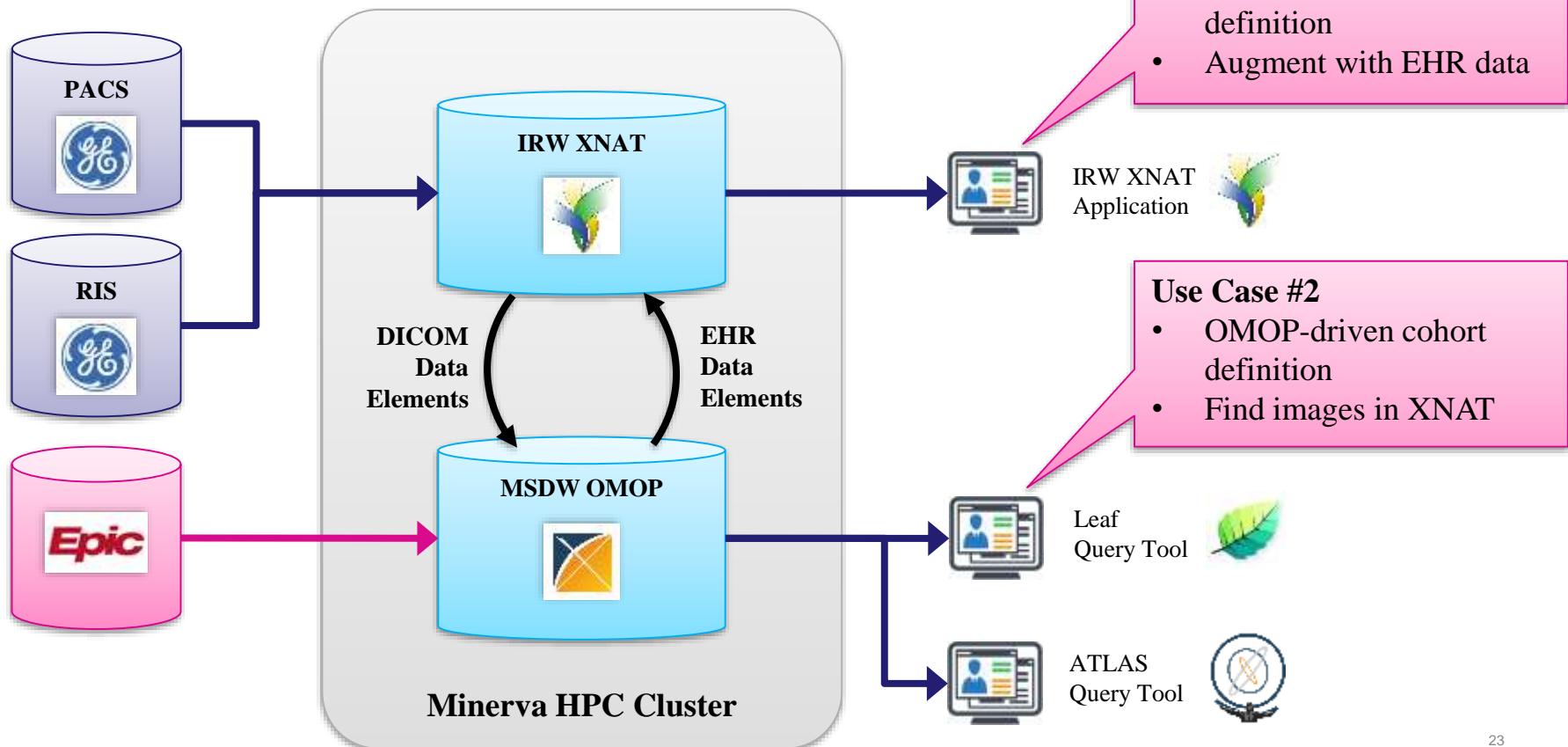
Leaf is supported by grant UL1TR004419 from the National Center for Advancing Translational Sciences, National Institutes of Health.

Leaf Accounts

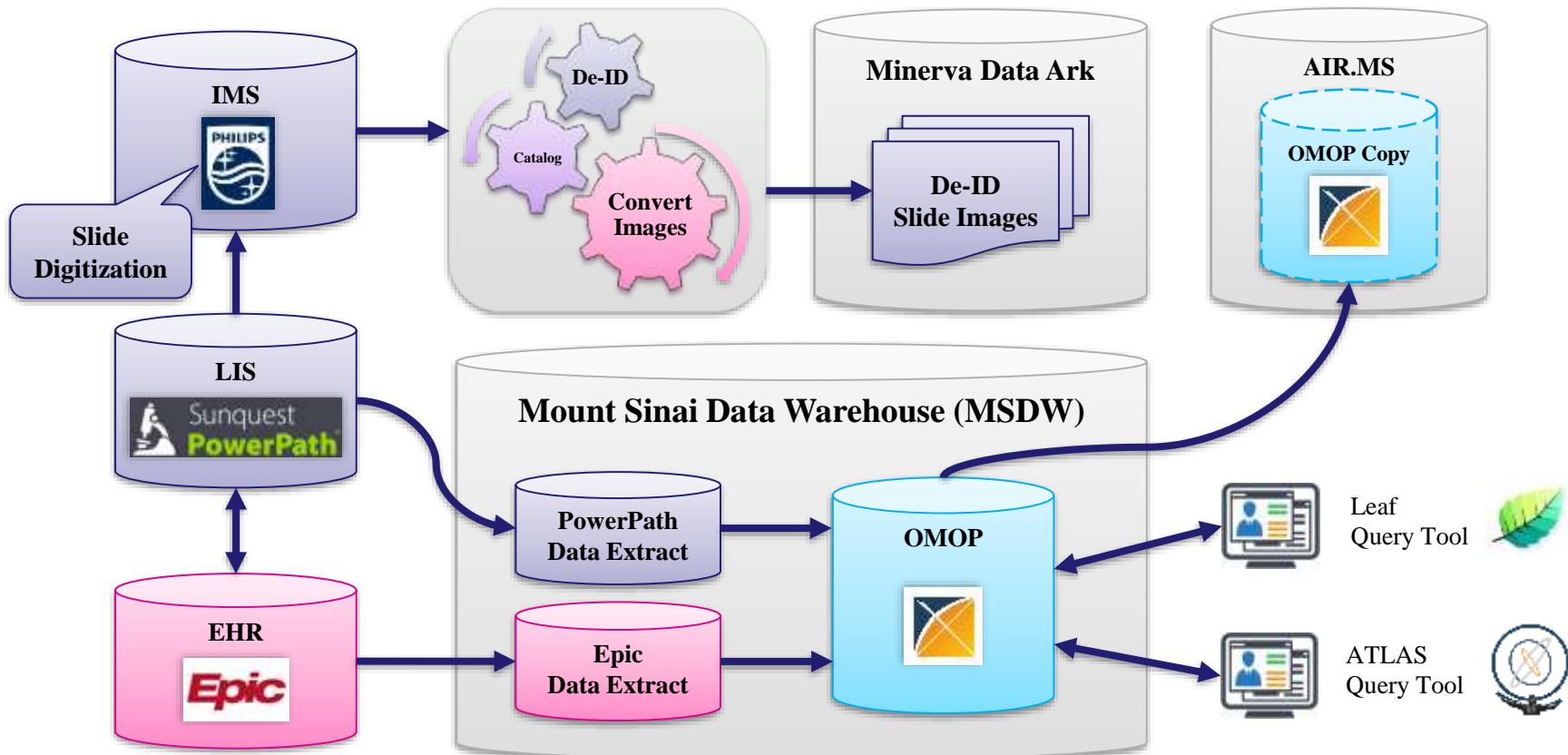
Leaf is supported by grant UL1TR004419 from the National Center for Advancing Translational Sciences, National Institutes of Health.

Integrations: Imaging Research Warehouse (IRW) and PowerPath for Digital Pathology

Integration between IRW 2.0 and MSDW



Digital Pathology Data Integration



Concept Mapping for OMOP

Terminology Mapping Services Vendor



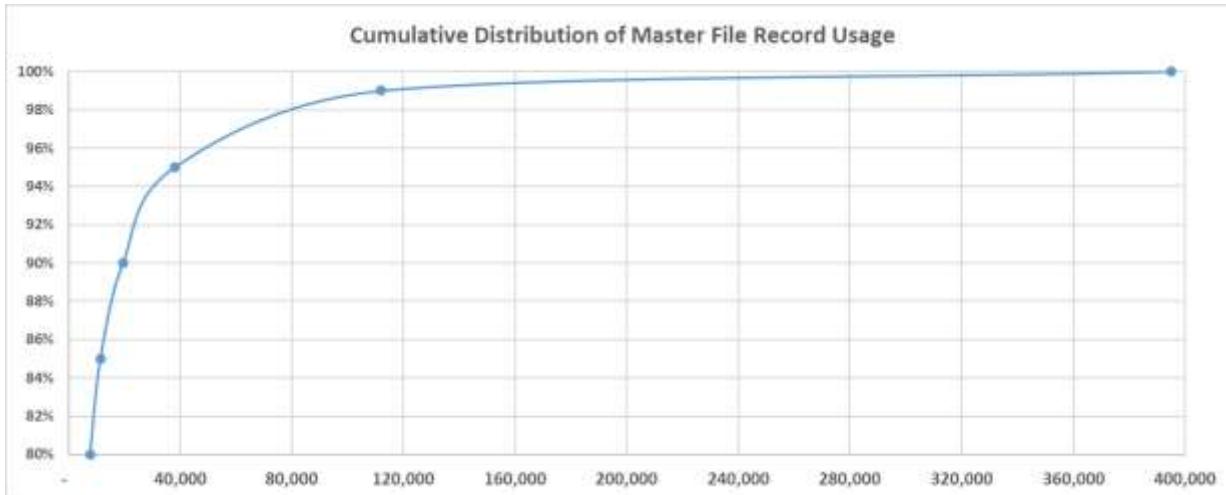
11611 North Meridian St
Suite 500
Carmel, IN 46032

Leading provider of healthcare terminology mapping services

Symedical mapping software

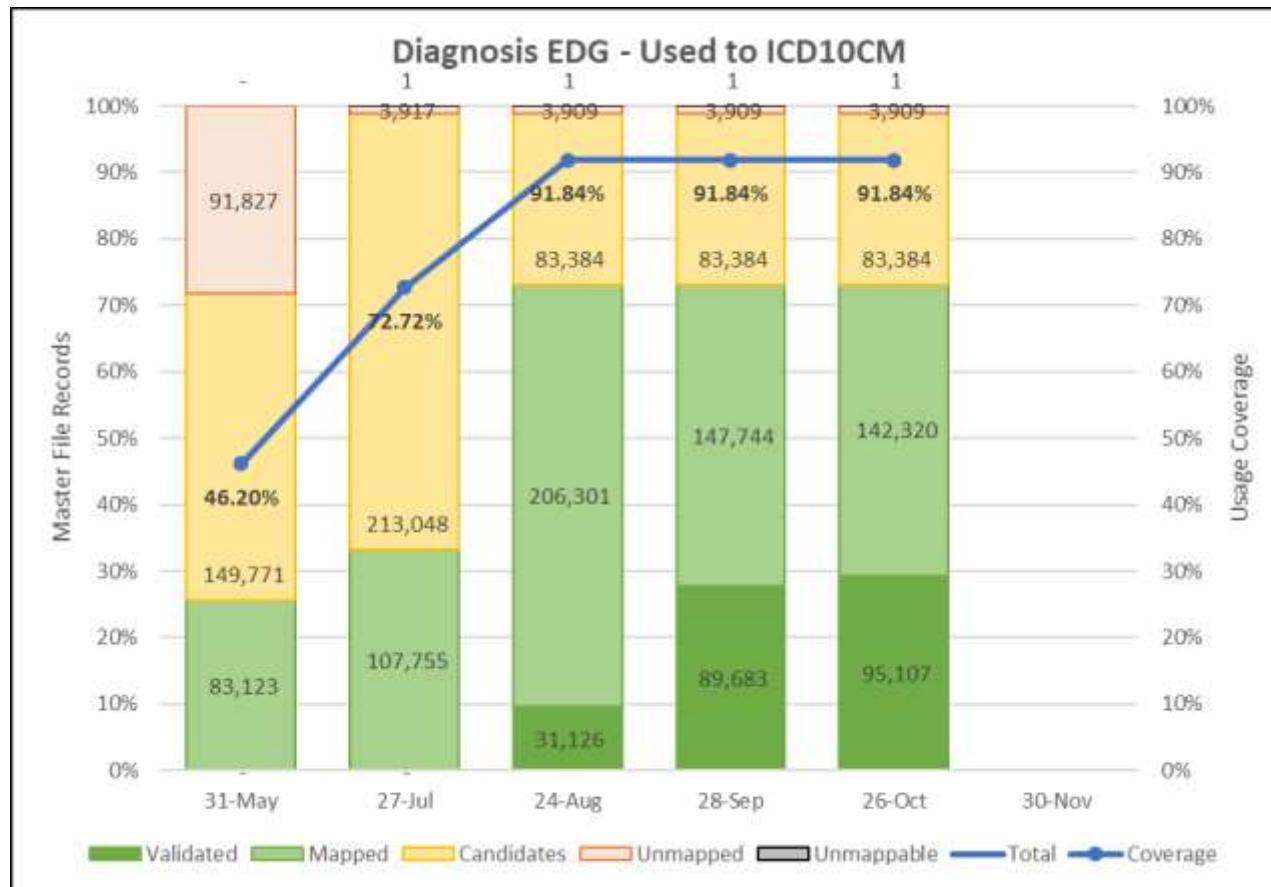
- ▶ Multiple matching algorithms
- ▶ Matching algorithms tailored to domain
- ▶ Each algorithm tuned for domain and customer data
- ▶ Underlying data model, dictionaries, thesauri for each domain

Most Epic Codes Used Infrequently (“Long Tail”)



Priority	Master File	Target Vocabulary	Rows for 80%	Rows for 85%	Rows for 90%	Rows for 95%	Rows for 99%	Used Rows	Total Rows
1	Diagnosis EDG	SNOMED-CT	6,000	9,000	16,000	32,000	96,000	324,721	1,566,543
2	Diagnosis EDG	ICD10CM	6,000	9,000	16,000	32,000	96,000	324,721	1,566,543
3	Lab Component LRR	LOINC / SNOMED	200	300	400	600	1,800	21,866	33,891
4	Medication ERX	RxNorm	800	1,000	1,400	2,400	6,200	53,376	134,037
5	Procedure EAP	CPT / HCPCS	200	400	600	1,000	3,000	26,159	154,086
6	Surgical Procedure ORP	CPT / HCPCS	400	600	800	1,400	2,600	4,712	31,486
7	Immunization LIM	CVX / RxNorm	30	40	50	80	130	333	399
8	Allergy ELG	SNOMED-CT / RxNorm	150	200	350	700	2,300	8,249	38,914
			TOTAL:	7,780	11,540	19,600	38,180	112,030	439,416
				80%	85%	90%	95%	99%	100%

Mapping Example: Epic Diagnosis (EDG) to ICD-10-CM



New Services: Interactive Dashboards & Geocoding

Interactive Dashboards and Data Visualization

Use Cases

- Clinical trial recruitment
 - Identifying potentially eligible patients
 - Monitoring recruitment activity
- Geographic distribution of a cohort
- Survival curves
- Visualization of basic demographic

Technology

- Features and Functionality
 - Dashboards created in Tableau and PowerBI
 - Formatting of visualizations to tell a coherent story
 - Interactive functions including on demand filtering
- Data Sources
 - Mount Sinai Data Warehouse
 - Epic Caboodle database
 - Epic Clarity database
 - REDCap
 - Cancer registry
 - MSX
 - Additional sources as requested

Trial Recruitment Dashboard

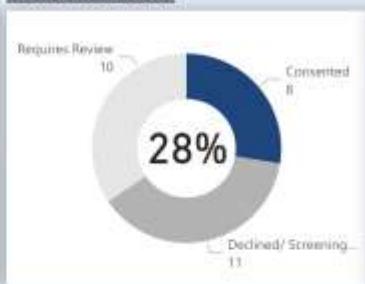
Daily Recruitment Report:

Scientific Computing and Data

Patient Volume

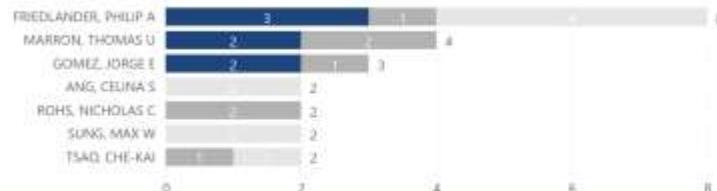


Consent Status



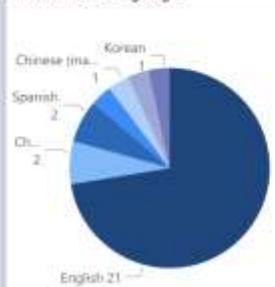
Ordering Physicians

● Consented ● Declined/ Screening Failure ● Requires Review



Demographic Breakdown

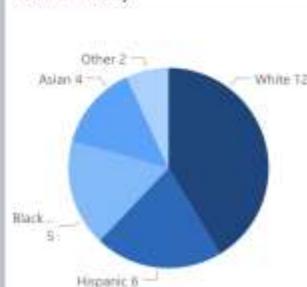
Preferred Language



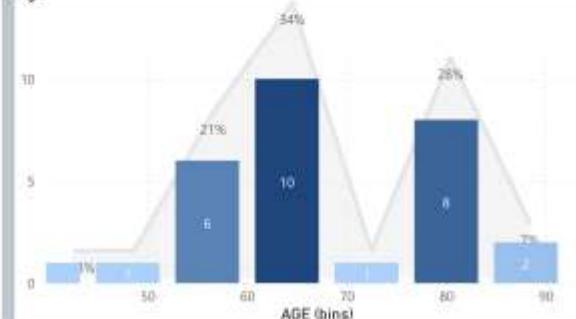
Gender



Race/Ethnicity

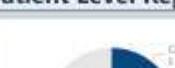


Age



Trial Recruitment Dashboard cont.

Patient Level Report



21%

Completed: 10
Remaining: 30

Upcoming Cycle

Cycle	Count
Cycle 1	10
Cycle 2	5
Cycle 3	10
Cycle 4	5

Preferred Language



Language	Count
English	15
Spanish	3
Chinese	2
Other	1

Gender



Gender	Count
Female	15
Male	4
Other	1

Race/Ethnicity



Race/Ethnicity	Count
White	8
Hispanic	7
Other	3

Scientific Computing and Data

Cycle Name	Ordering Provider	Tx Plan Start Dt/Tm	MRN	First Name	Last Name	Sex	DOB	Race	Ethnicity	Family
Cycle 1	FREELANDER, PHILIP A	2023-09-01 08:00:00	1234567890	PHILIP	FREELANDER	FEMALE	1985-01-01	WHITE	NON-HISPANIC	HS
Cycle 2	FREELANDER, PHILIP A	2023-09-01 08:00:00	1234567890	PHILIP	FREELANDER	FEMALE	1985-01-01	WHITE	NON-HISPANIC	HS
Cycle 1	FREELANDER, PHILIP A	2023-09-01 08:00:00	1234567890	PHILIP	FREELANDER	FEMALE	1985-01-01	WHITE	NON-HISPANIC	HS
✓ Cycle 1	MARRIOL, THOMAS U	2023-09-01 08:00:00	1234567890	THOMAS	MARRIOL	FEMALE	1985-01-01	WHITE	NON-HISPANIC	HS
✓ Cycle 2	MARRIOL, THOMAS U	2023-09-01 08:00:00	1234567890	THOMAS	MARRIOL	FEMALE	1985-01-01	WHITE	NON-HISPANIC	HS
Cycle 3	MARRIOL, THOMAS U	2023-09-01 08:00:00	1234567890	THOMAS	MARRIOL	FEMALE	1985-01-01	WHITE	NON-HISPANIC	HS
✓ Cycle 1	FREELANDER, PHILIP A	2023-09-01 08:00:00	1234567890	PHILIP	FREELANDER	MALE	1985-01-01	WHITE	NON-HISPANIC OR LATINO	WP
✓ Cycle 2	FREELANDER, PHILIP A	2023-09-01 08:00:00	1234567890	PHILIP	FREELANDER	MALE	1985-01-01	WHITE	NON-HISPANIC OR LATINO	WP
✓ Cycle 3	FREELANDER, PHILIP A	2023-09-01 08:00:00	1234567890	PHILIP	FREELANDER	MALE	1985-01-01	WHITE	NON-HISPANIC OR LATINO	WP
✓ Cycle 1	TSANG CHE-FAW	2023-09-01 08:00:00	1234567890	CHE-FAW	TSANG	MALE	1985-01-01	WHITE	NON-HISPANIC	HS
✓ Cycle 2	TSANG CHE-FAW	2023-09-01 08:00:00	1234567890	CHE-FAW	TSANG	MALE	1985-01-01	WHITE	NON-HISPANIC	HS
✓ Cycle 3	TSANG CHE-FAW	2023-09-01 08:00:00	1234567890	CHE-FAW	TSANG	MALE	1985-01-01	WHITE	NON-HISPANIC	HS
✓ Cycle 1	YELLOWWAMP, RAJAWANTH R	2023-09-01 08:00:00	1234567890	RAJAWANTH	YELLOWWAMP	MALE	1985-01-01	BLACK OR AFRICAN-AMERICAN	NON-HISPANIC OR LATINO	BL
✓ Cycle 2	YELLOWWAMP, RAJAWANTH R	2023-09-01 08:00:00	1234567890	RAJAWANTH	YELLOWWAMP	MALE	1985-01-01	BLACK OR AFRICAN-AMERICAN	NON-HISPANIC OR LATINO	BL

Role of MSDW in Clinical Trial Recruitment

Using electronic health record data to provide focused lists of eligible patients to clinical research coordinators

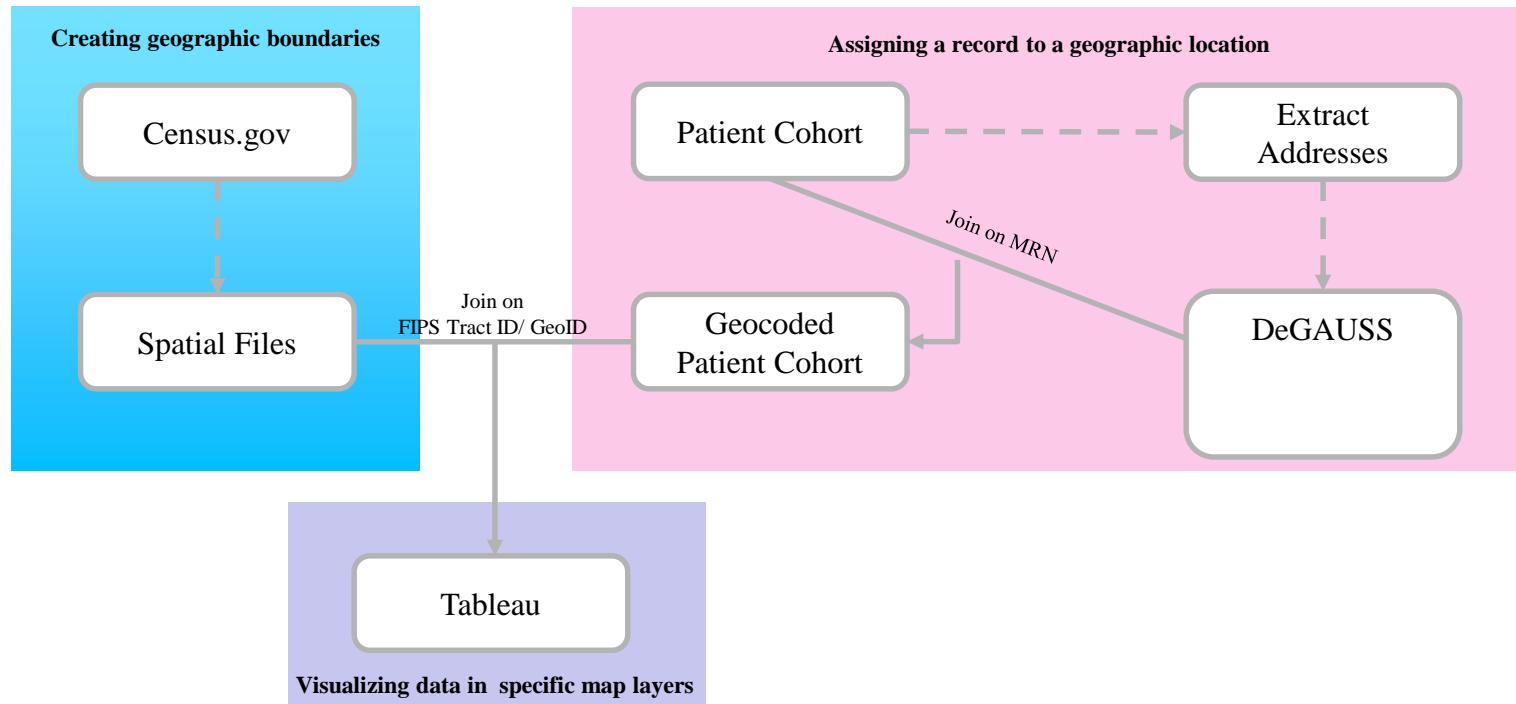
1. Define patient cohort **inclusion/exclusion criteria** for studies
 - Operationalized with Epic EHR data
 - Obtain as narrow a list as possible
2. Identify new patients **potentially eligible** to enroll in each clinical trial
 - Provide automated, regular reports
 - Facilitate easy chart review by clinical research coordinators
3. CRCs review charts of short-listed patients to **ensure enrollment eligibility**

Prostate Cancer Survivor Dashboard

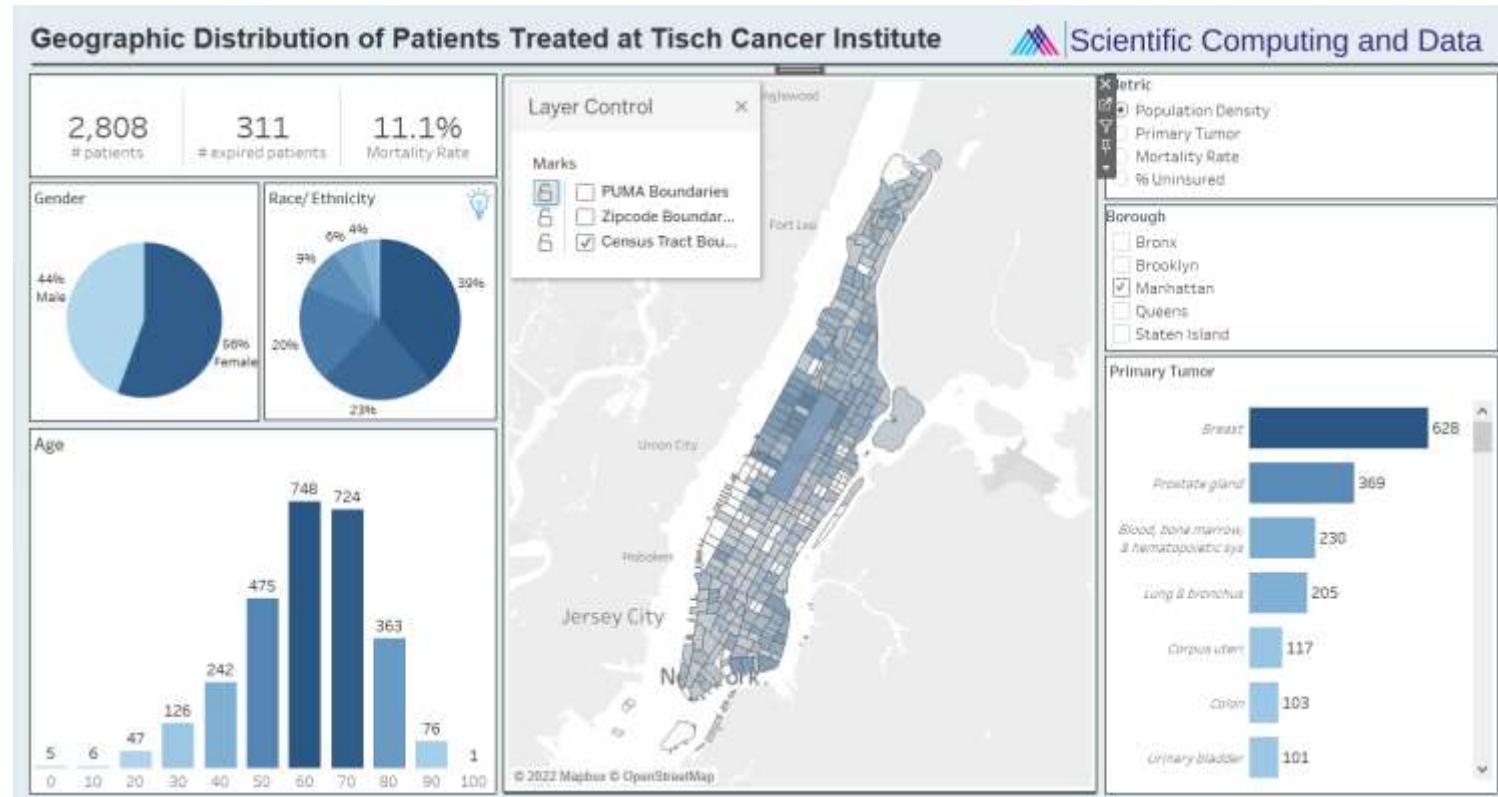


Geocoding Workflow with DeGAUSS

DeGAUSS is an open-source application, executable on a local machine, for geocoding



Example of a Dashboard Using Geocoding



MSDW Roadmap

MSDW Development Roadmap for 2023

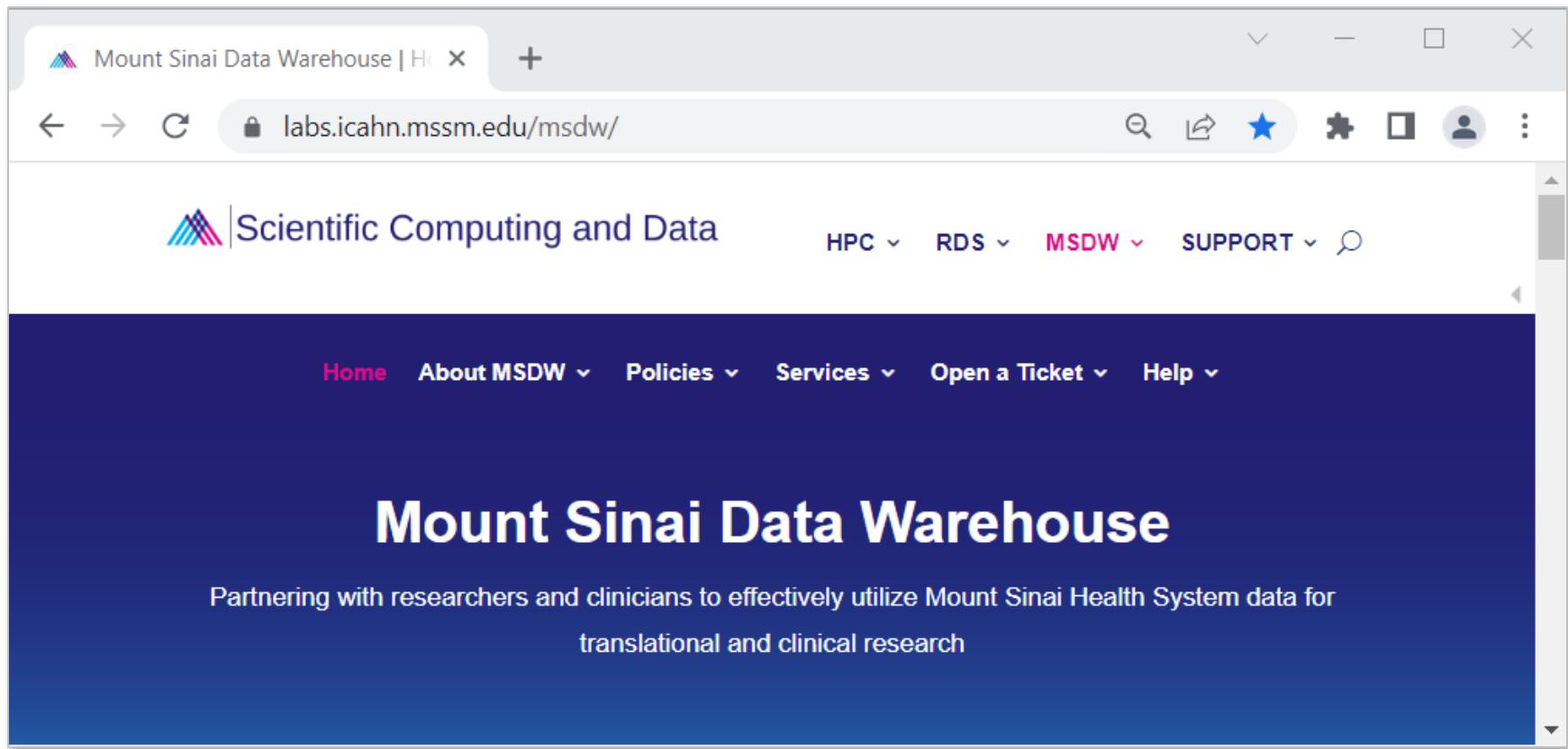
Quarter	Integrate New Data into MSDW	Upgrade MSDW Infrastructure
Q1 2023	<ul style="list-style-type: none">• Impressions from radiology reports• Imaging study data elements from IRW• Pathology specimen and slide data from PowerPath	<ul style="list-style-type: none">• Automated data quality checks
Q2 2023	<ul style="list-style-type: none">• Historical MRNs<ul style="list-style-type: none">• Improved patient record linking• Links to digitized pathology slide images on Minerva	<ul style="list-style-type: none">• Support for limited data sets in OMOP data marts & TriNetX• Support for OMOP version 5.4
Q3 2023	<ul style="list-style-type: none">• Admission, Discharge & Transfer (ADT) Transactions<ul style="list-style-type: none">• When patients are in the ICU, etc.	<ul style="list-style-type: none">• Framework for Epic upgrades<ul style="list-style-type: none">• Next upgrade on Sun, Nov 5 2023
Q4 2023	<ul style="list-style-type: none">• Expand imaging study data elements from IRW	

Potential Features for MSDW Roadmap

Feature	Data Sources
Cross-reference links to other data sets	BioMe Biobank GE MUSE ECG Data
Hospital & professional billing (OMOP cost table)	MSX
OMOP Genomics extension tables (G-CDM)	BioMe Biobank Epic Genomics Module
Sexual Orientation & Gender Identity (SOGI)	Epic SOGI SmartForm
Labor & Delivery	Epic
Antimicrobial Sensitivity	Epic
Ultrasound metadata	ASOB
Geocoding of patient addresses	deGAUSS
Cancer diagnosis, tumor characteristics, stage of disease, treatment, etc.	Cancer Registry
Date of death from external source(s)	Social Security Administration Death Master File (DMF)

Appendix

Our Website: msdw.mountsinai.org



The screenshot shows a web browser window displaying the Mount Sinai Data Warehouse (MSDW) website. The URL in the address bar is labs.icahn.mssm.edu/msdw/. The page header includes the Mount Sinai logo, the text "Scientific Computing and Data", and navigation links for "HPC", "RDS", "MSDW" (which is highlighted in blue), and "SUPPORT". Below the header is a dark blue navigation bar with links for "Home", "About MSDW", "Policies", "Services", "Open a Ticket", and "Help". The main content area features a large, bold, white text "Mount Sinai Data Warehouse" on a blue gradient background. Below this, a white text box contains the tagline: "Partnering with researchers and clinicians to effectively utilize Mount Sinai Health System data for translational and clinical research".

MSDW Website Resources (a subset)

Website Resource	URL Link
MSDW Homepage	https://labs.icahn.mssm.edu/msdw/
Open a Ticket for Assistance	https://labs.icahn.mssm.edu/msdw/open-a-ticket/
Data Contents & Record Statistics	https://labs.icahn.mssm.edu/msdw/data-sources/
Protected Patient Categories & Data Exclusions	https://labs.icahn.mssm.edu/msdw/about-us/protected-patient-categories/
MSDW Release Notes	https://labs.icahn.mssm.edu/msdw/msdw-release-notes/
Presentations	https://labs.icahn.mssm.edu/msdw/presentations/
Training & Tutorials	https://labs.icahn.mssm.edu/msdw/training/
FAQs	https://labs.icahn.mssm.edu/msdw/faqs/