Mount Sinai Data Warehouse Town Hall

Sharon Nirenberg, MD
Timothy Quinn, PhD
Scientific Computing & Data
Icahn School of Medicine at Mount Sinai
October 24, 2023

Icahn
School of
Medicine at
Mount
Sinai

Agenda

1. MSDW Team Updates
2. MSDW Operations Since Q2 2023
3. Project Updates
4. MSDW Roadmap

MSDW Team Updates

The MSDW Team

Patricia Kovatch
Professor & Dean for Scientific Computing

Sharon Nirenberg MD
Physician Informaticist

Naomi So MD
Physician Informaticist

Teja Ganta MD
Physician Informaticist

Farhan Mahmood
Director Scientific Computing

Raj Bose PhD
Director Research Engagement

Jing Yang PhD
Research Engagement Specialist

Jiani Xiang
Clinical Data Specialist

Priyal Mehta
Healthcare Data Analyst

Timothy Quinn PhD
Principal Data Architect

Shivaji Punukollu
Healthcare Data Engineer

Manoj Chekuri
Healthcare Data Engineer

Jacob Weiser
Healthcare Data Engineer

Darius Boopal
Healthcare Data Engineer

Rupan Hossain
Database Administrator

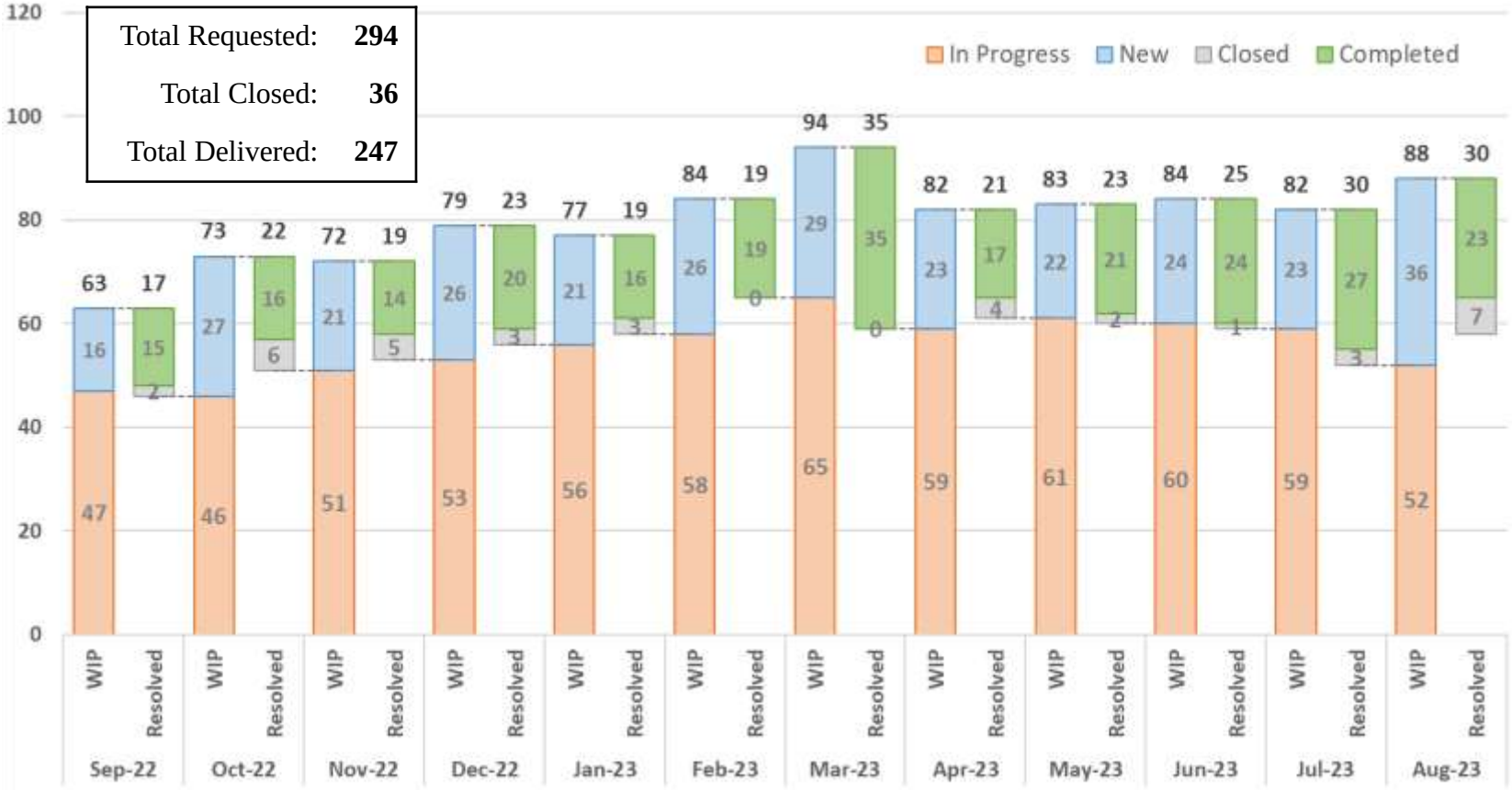
New Informaticist: Teja Ganta MD

- Physician Informaticist
  - 30% with Scientific Computing & Data
- Joint appointments
  - Division of Hematology & Medical Oncology
  - Tisch Cancer Institute
  - Office of the CMIO
- Joins us from ISMMS's fellowship programs
  - Hematology & Medical Oncology
  - Clinical Informatics

MSDW Operations Since Q2 2023

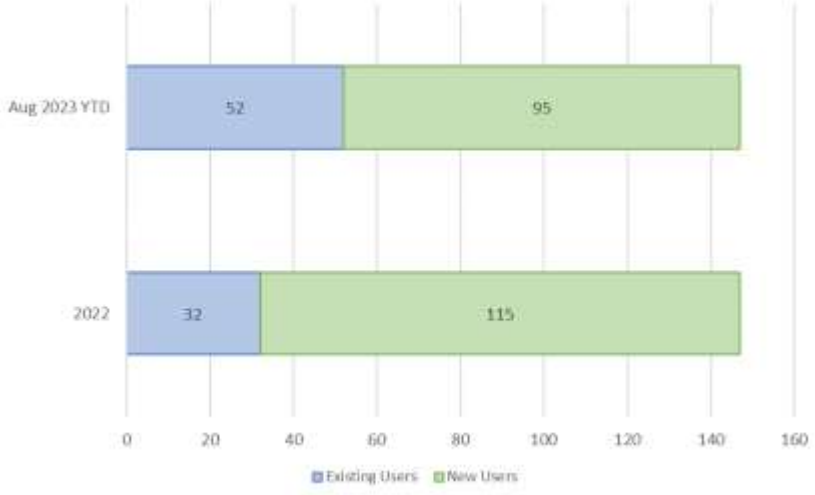
Custom Data Sets Requested & Delivered (Last 12 Months)

Total Requested:	**294**
Total Closed:	**36**
Total Delivered:	**247**



Data Sets Delivered to 147 Researchers in 2023 YTD

- In 2023 YTD, MSDW team has delivered **204** custom data sets to **147** distinct customers
- **35%** are repeat customers from 2022



Top 5 Requestors – Aug 2023 YTD

Name	Department	# of Data Requests
Amy Kontorovich	Medicine – Cardiology	10
Harm van Bakel	Genetics & Genomic Sciences	9
Minal Kale	Medicine – Gen. Internal Medicine	7
Nina Bickell	Population Health Science & Policy	4
Carol Horowitz	Population Health Science & Policy	3

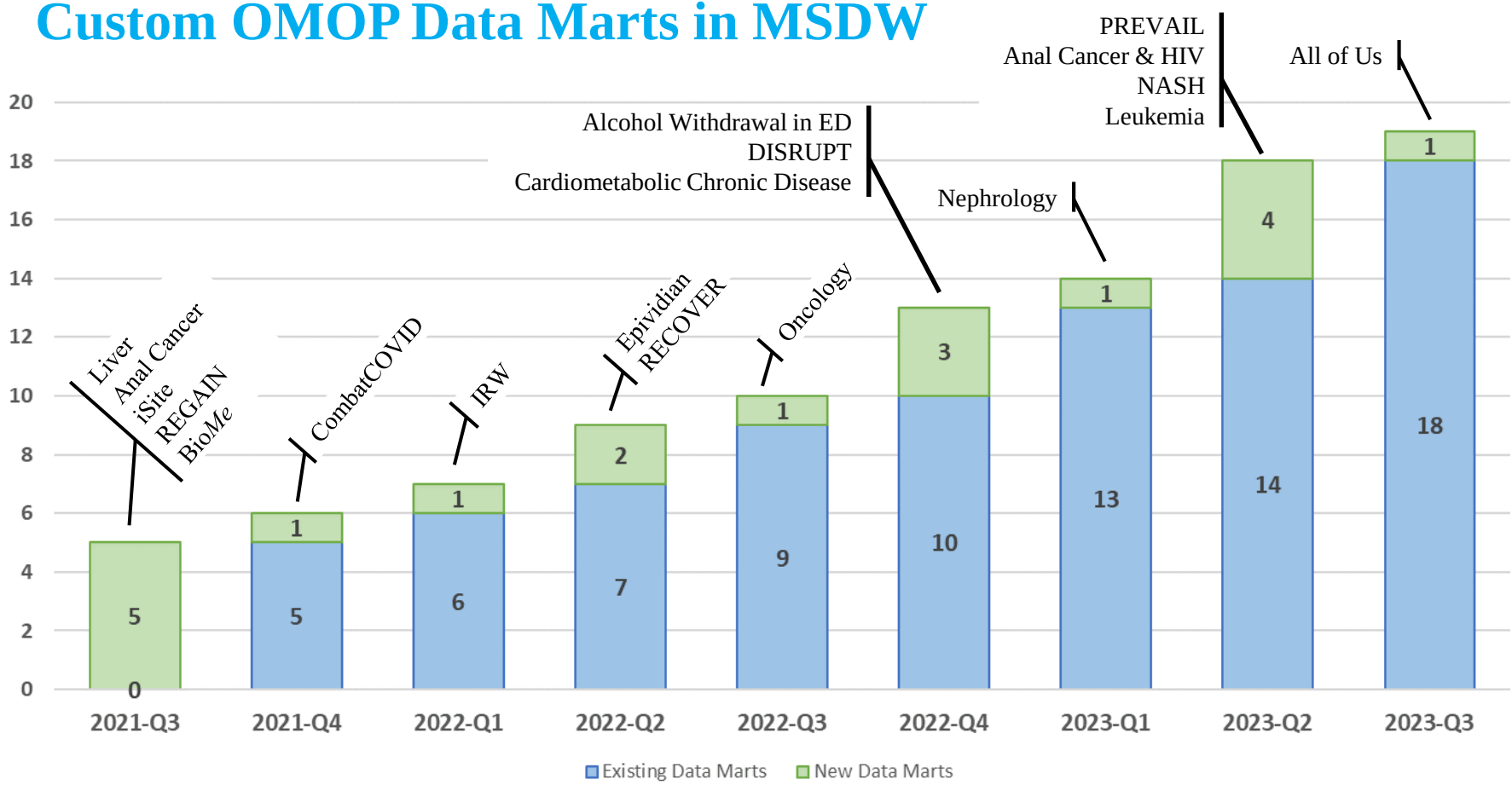
Top 5 Requesting Departments – Aug 2023 YTD

Department	# of Data Requests
Medicine - General Internal Medicine	17
Medicine - Cardiology	16
Genetics & Genomic Sciences	15
Population Health Science & Policy	9
Medicine - Gastroenterology	9

Custom Data Set Turnaround Times Improved in 2023

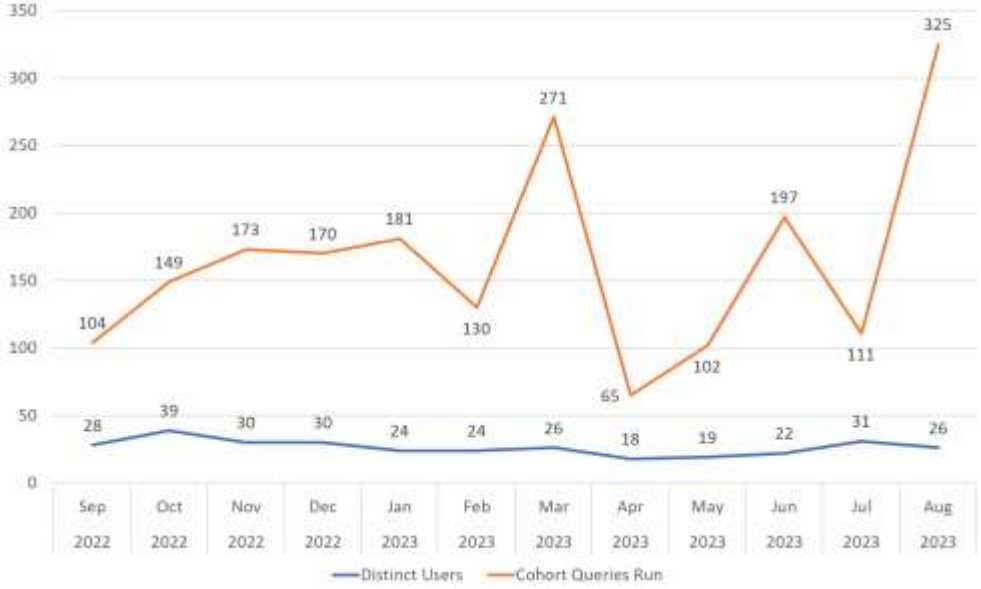
Note: **Lower values are better**

Custom OMOP Data Marts in MSDW

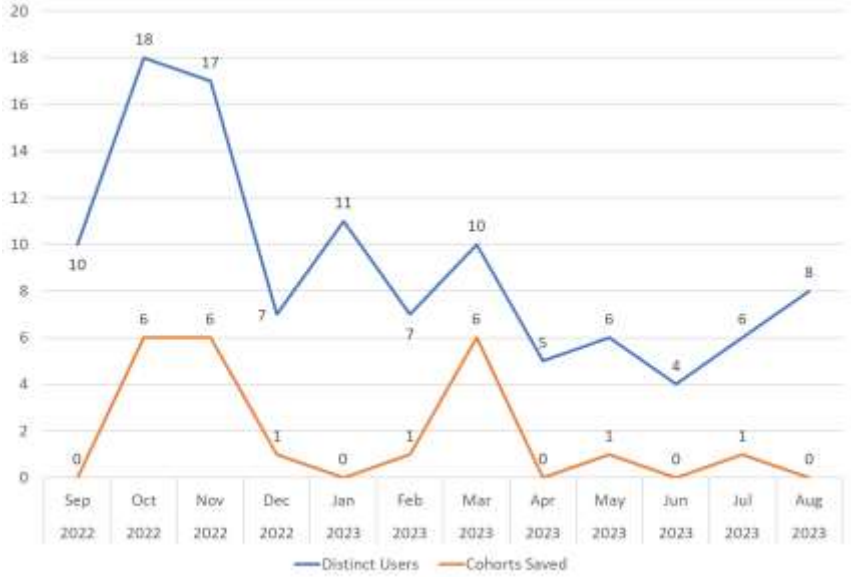


Utilization of Leaf and ATLAS (Sep 2022 to Aug 2023)

Leaf Utilization

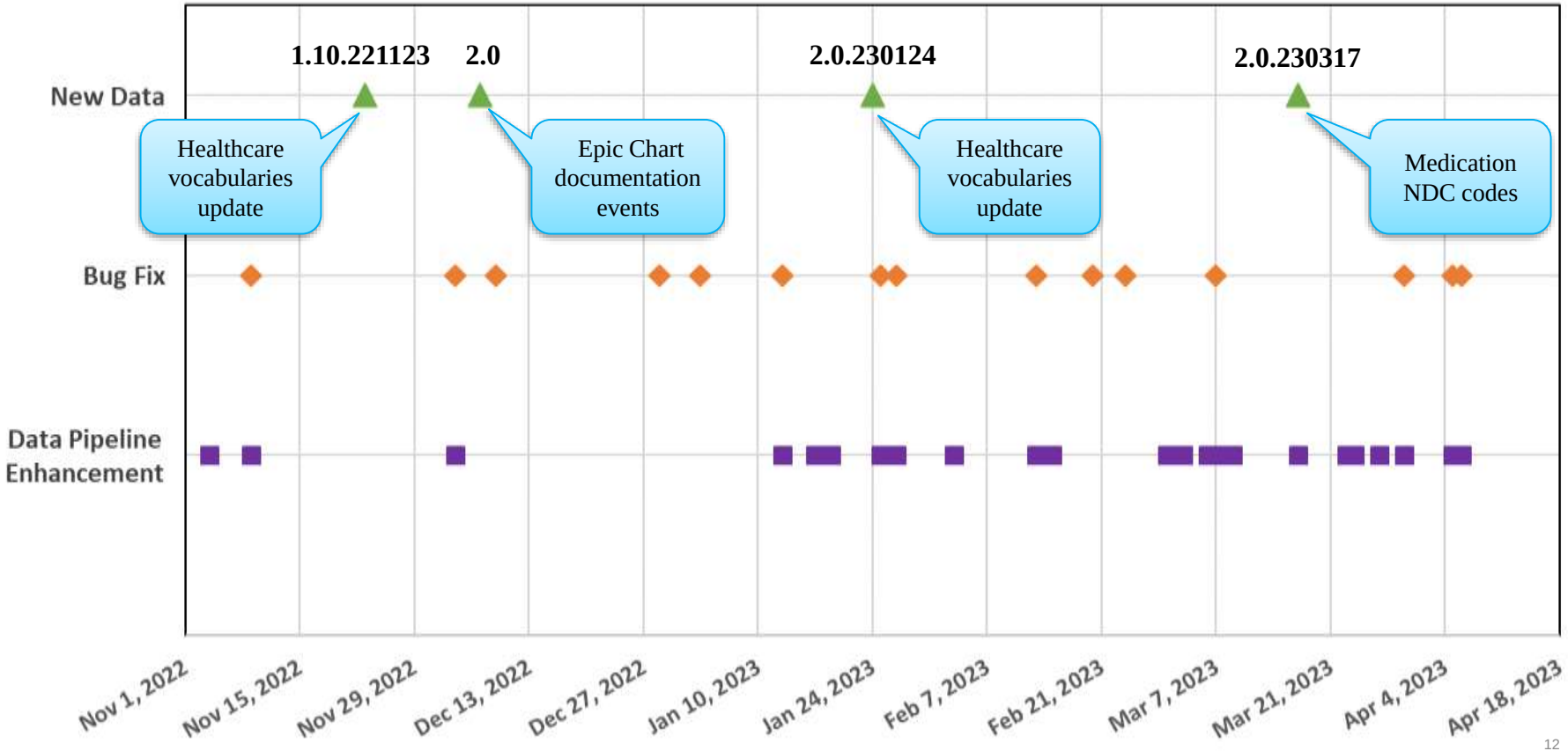


ATLAS Utilization

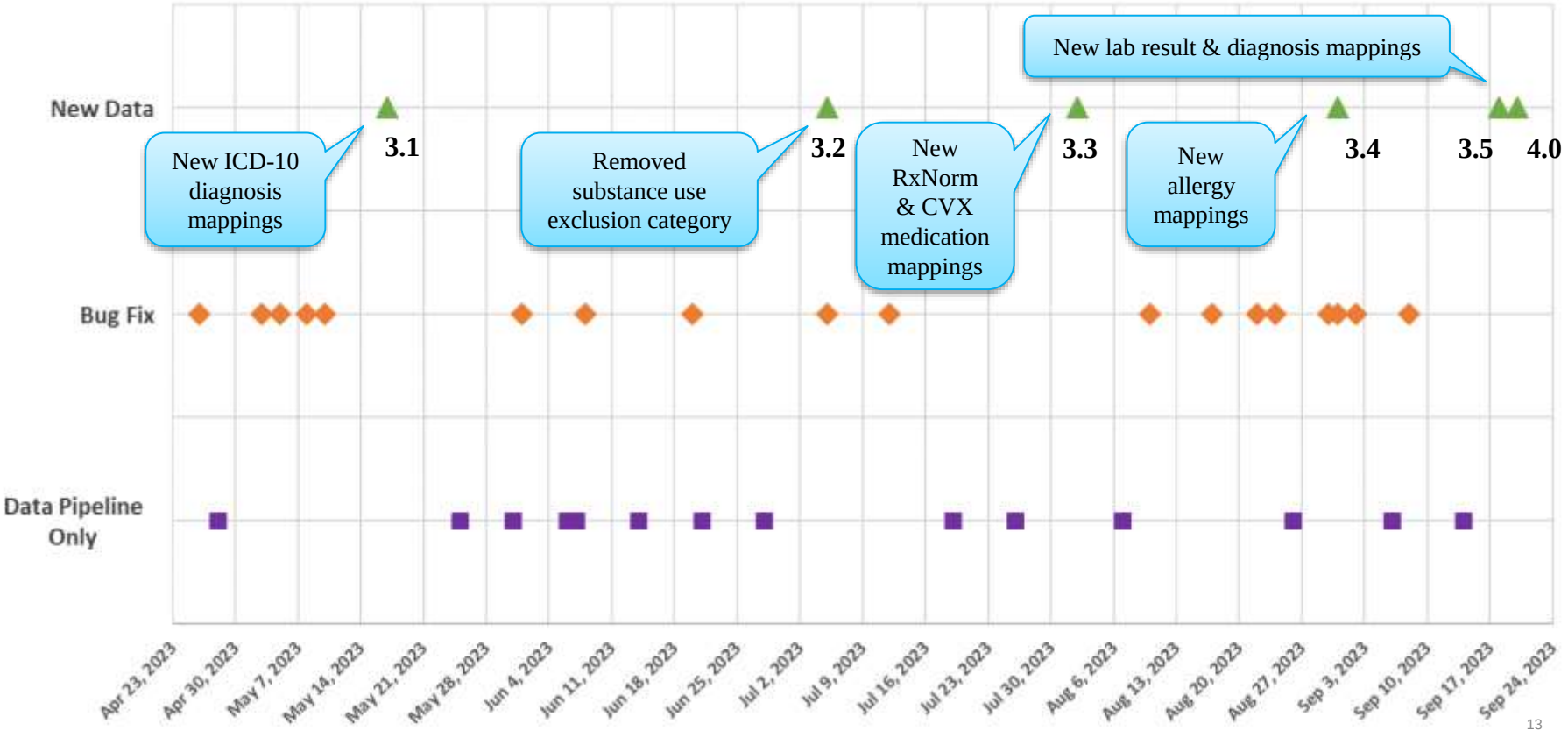


Leaf utilization has remained steady at **~26 users per month** over the last 12 months

33 OMOP Releases: Nov 1, 2022 to Apr 18, 2023 (~1.4 per week)



36 OMOP Releases: Apr 23, 2023 to Sep 20, 2023 (~1.6 per week)



Data Quality Improvements

- **Patient Ethnicity**
  - Combined Epic's old **ethnic group** and new **ethnic background** data elements
  - Increased fill rate from **0.26%** to **23.19%**

- **Patient Race**
  - Fill rate is **25.79%**
  - New "**Patient Declined/Prefer Not to Say**" category selected by ~20k patients
  - Correctly determine **~30k multi-racial patients** (Epic Caboodle's flag is wrong)
  - Filter out Epic's current and historical invalid categories (e.g., Hispanic/Latino, Unknown)

- **Classification of Encounters**
  - Correctly classify telephone and video **telehealth encounters** (now ~3.25% of outpatient encounters)
  - New category for **chart documentation events** (52% of Epic's encounter table)

- **Content Standardization / Concept mapping**

Increased & Improved Mappings from Epic to Standard Codes in OMOP

Clinical Architecture

11611 North Meridian St
Suite 500
Carmel, IN 46032

Leading provider of healthcare terminology mapping services

Symedical mapping software

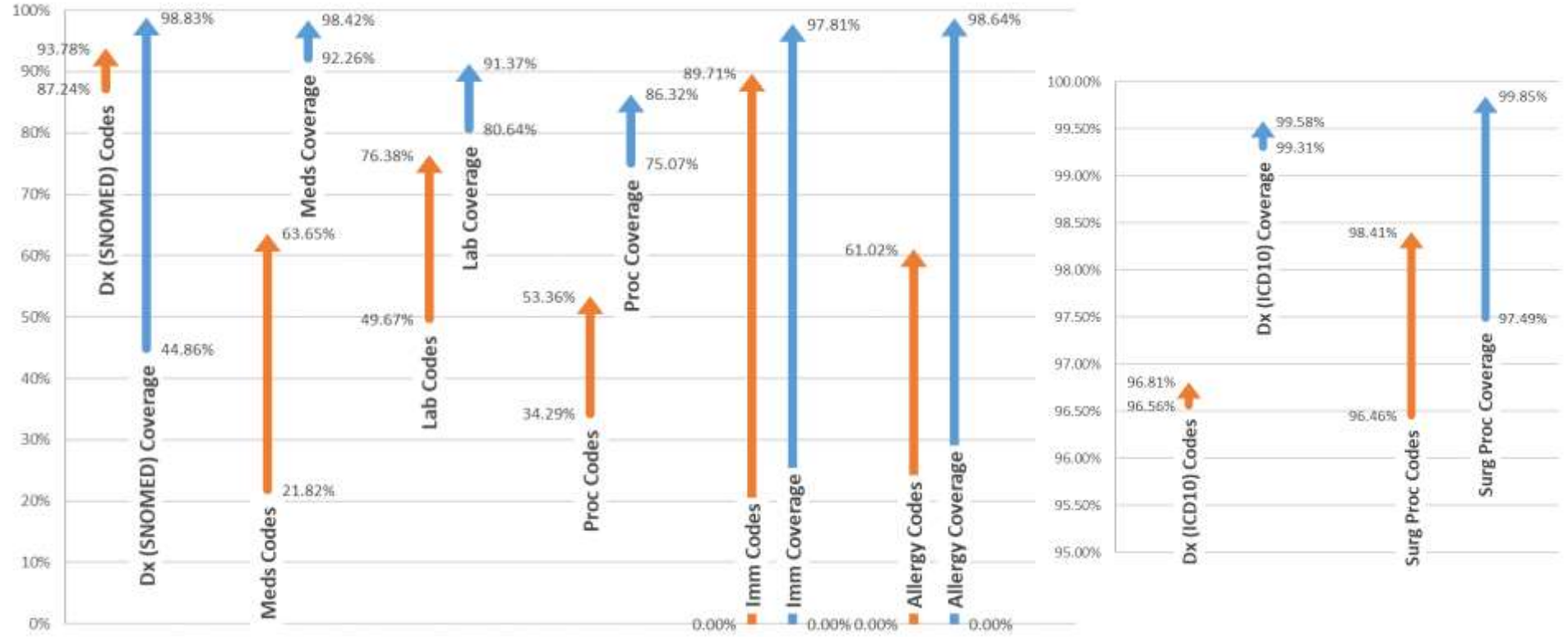
- Multiple matching algorithms
- Matching algorithms tailored to domain
- Each algorithm tuned for domain and customer data
- Underlying data model, dictionaries, thesauri for each domain

Project Dates: April 20, 2022 – Sept 6, 2023

Substantial Increase in Concept Mapping Coverage in All Domains

Epic Master File	OMOP Table	Mapped Codes	Mapped Code Increase	Mapping Coverage	Mapping Coverage Increase
EDG Diagnoses (ICD10)	concept_relationship	1,316,155	+51,468	**99.58%**	+0.27%
EDG Diagnoses (SNOMED)	condition_occurrence	1,206,363	+102,405	**98.83%**	+53.97%
ERX Medications	drug_exposure	61,455	+53,928	**98.42%**	+6.16%
LIM Immunizations	drug_exposure	380	+380	**97.81%**	+97.81%
LRR Lab Components	measurement	23,350	+8,847	**91.37%**	+10.72%
EAP Procedures	procedure_occurence	144,237	+112,947	**86.32%**	+11.25%
ORP Surg Procedures	procedure_occurence	15,822	+9,726	**99.85%**	+2.36%
ELG Allergies	observation	16,327	+16,327	**98.64%**	+98.64%

Substantial Increase in Concept Mapping Coverage in All Domains



Engaged 112 Researchers via Events in 2023 (Sept YTD)

Dates	Event Details	Attendees
Feb 24, 2023	Training on Leaf & ATLAS (cohort query tools)	45
May 2, 2023	MSDW Town Hall	18
Jan 1 – Sep 20, 2023	Digital Concierge	49
	TOTAL:	**112**

Project Updates

Epic for Research Updates

New Position funded by the CTSA

- Gabby Kroi - Epic Research Analyst (started in Aug)
- First Mount Sinai employee dedicated to Epic Research build

Bi-weekly Meeting led by Dr. Bruce Darrow

- Two goals:
  - Building technical infrastructure to support research community
  - Supporting individual researchers whose projects requires Epic integration

Training Session on current Epic Research features

- Collaborative effort between Epic and Scientific Computing & Data teams
- Scheduled for November 7, 2023
- Recording will be available later in PEAK and on Scientific Computing website

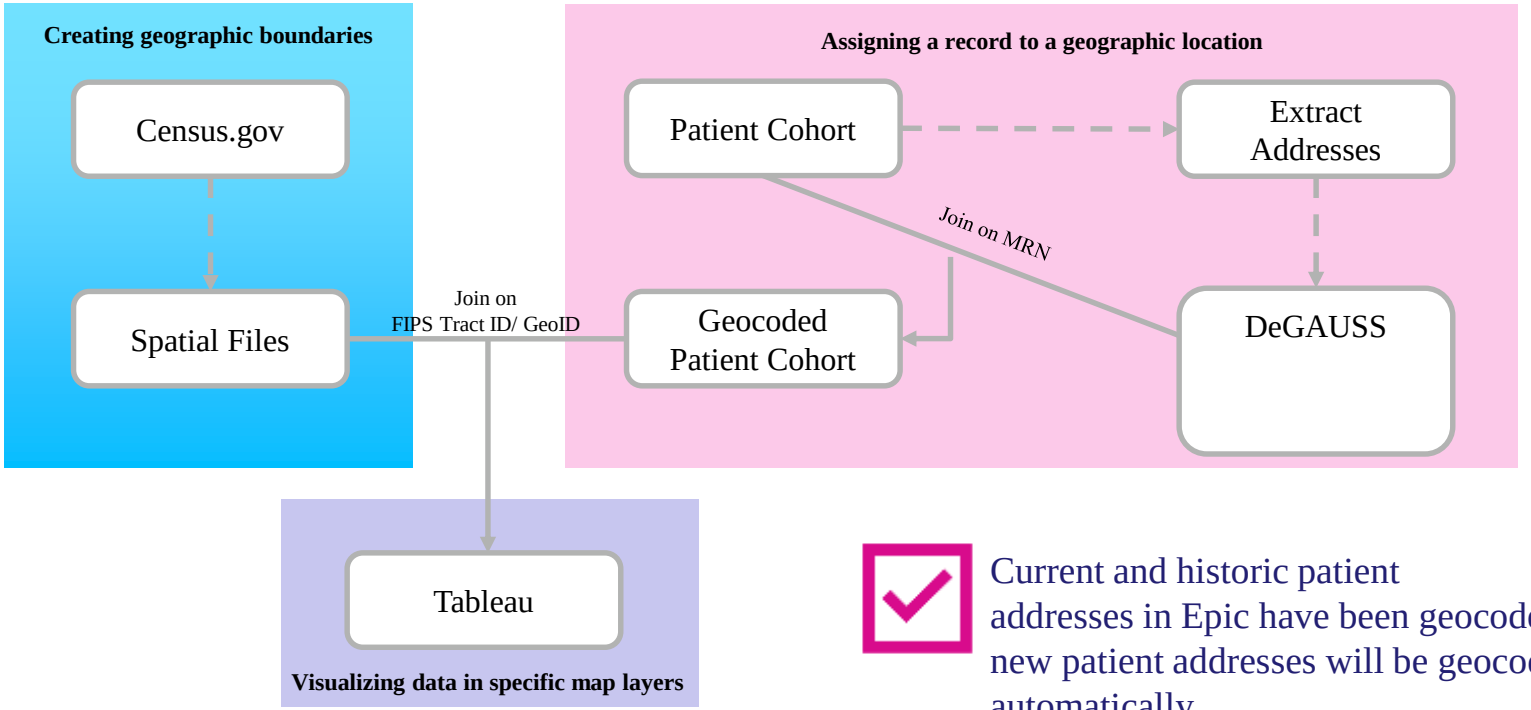
CTSA-Supported Initiative: Epic Research Module

2023 initiative to raise awareness of Epic's research-related capabilities via ORS and increased training

Epic Research Functionality	Currently Available at Mount Sinai
Automatically process research-specific claims	✓
Designate an encounter as a research-specific encounter	✓
Notify research staff of research patient upcoming appointment, ED or hospital admission, demise, etc.	✓
Enable research billing review	✓
Designate an order as research related	✓
Enable best practice advisories (BPAs) to identify potential research participants	✓
Patient opt in/out of research recruitment via MyMountSinai	✓
Send research project recruitment messages to patients via MyMountSinai	✓
Link research studies with study protocols	
Display current research studies in MyMountSinai	
Consent study participants via MyMountSinai	
Document and manage adverse events in Epic	

Current & Historic Patient Addresses in Epic have been Geocoded

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



Current and historic patient addresses in Epic have been geocoded & new patient addresses will be geocoded automatically

New Geocoding Expert at Mount Sinai

Andrew Maroko, PhD

- Associate Professor in Department of Population Health Science and Policy
- Hired by Dr. Carol Horowitz and Dr. Lynne Richardson
- Studies health disparities, inequities, exposures and accessibility by exploring the spatial variation and geographic associations of the environment
- Expert guiding the the MSDW team on the data sets and elements needed to make the geocoded addresses usable for Mount Sinai researchers

Geocoding and Social Determinants of Health

American Community Survey (ACS)

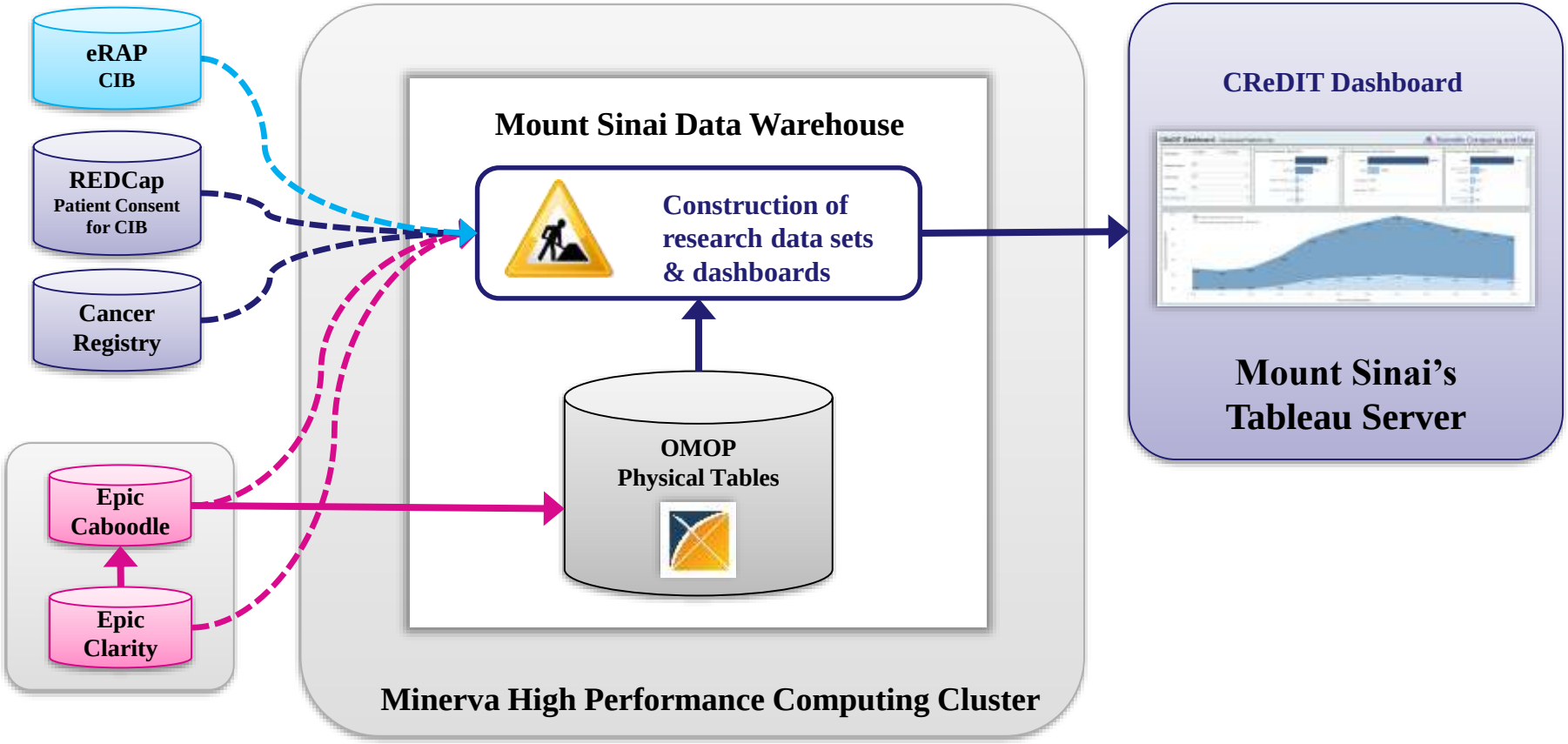
- Yearly survey of 3.5 million households across the US conducted by the United States Census Bureau
- Future work to link geocoded patient addresses in MSDW to key variables in the ACS

SDoH and Health Equity Domain	ACS Content
DEMOGRAPHIC	**Race/Ethnicity** **Citizenship** – Place of birth, ancestry, year of entry **Language** – Spoken at home, English proficiency, Linguistic isolation **Household Makeup** – Single-parent family, Multifamily household
SOCIAL	**Disability** – Type (cognitive, vision, hearing, other physical, etc.) **Educational Attainment** **Health Insurance**
ECONOMIC	**Income/Poverty Status** **Employment** – Status, Labor force participation **Public Programs** – SNAP, Income support **Transportation** – Commute to work, Vehicles available
HOUSING	**Type and Occupancy** – Owner/renter, Type, Time at address **Housing Costs and Conditions** – Monthly rent, Annual heating costs **Technology** – Internet connectivity Computers

Created CReDIT Dashboard (Phase 1)

- **C**omprehensive **Re**search **D**ata **I**ntegration **T**rust for Electronic Biorepository Phenotyping
- Dr. Matthew Galsky is the project stakeholder
- Dashboard linking patient level data from the cancer registry and the electronic health record with the Cancer Institute Biorepository
  - Included in updated MSDW IRB protocol
- Readily determine counts of patients with data in both CIB and the cancer registry
- Patient level data available with appropriate IRB approval
  - Electronic health record and cancer registry data provided by the MSDW team
  - Samples from CIB provided by Dr. Rachel Brody's team

CReDIT Dashboard Construction



MSDW Roadmap

MSDW Projects in Progress

	Project	Target Date	New Capabilities for Researchers
1.	**REST API Web Services** • Protected patients • Masked identifiers for de-ID IRW 2.0	2023-Q3	Easy access to list of protected patients
2.	**PowerPath** • Procedures, specimens, slides • De-ID slide image files in Data Ark	2024-Q1	Enable self-service cohort identification combining pathology metadata and EHR data, including links to slide images on Minerva
3.	**XNAT (IRW 2.0)** • DICOM parsing • Patient & procedure matching • Procedures, imaging series	2024-Q1	Enable self-service cohort identification combining imaging study metadata and EHR data
4.	**Expand Access to new Data Sources** • CNExT Cancer Registry • ASOB fetal ultrasound • Provation endoscopy • Cardiac Catheterization data • MOSAIQ radiation oncology • Dentrix dental	2024	Expand the scope of data elements available to include in custom data sets

Change in Mt Sinai Policy for De-identified HIV/AIDS Data

- **New York State's Article 27-F law applies to PHI data only**
  - Use of HIV/AIDS data requires IRB approval
  - The law makes no mention of de-identified data

- **New Guidance from the Office for Human Research Protections (OHRP)**
  - At the Federal level, OHRP has published new guidance that data de-identified in a HIPAA-compliant manner is **Non-Human Subjects Research (NHSR)** and, therefore, IRB review is decided by institutional (i.e., Mount Sinai) policy
  - https://www.hhs.gov/ohrp/coded-private-information-or-biospecimens-used-research.html

- **Mount Sinai's Policy Change**
  - Mount Sinai is adopting the OHRP Guidance: **HIPAA-compliant de-identified data is NHSR**
  - Mount Sinai does not require IRB approval for research use of HIPAA-compliant de-identified data, including HIV/AIDS data

- **MSDW Next Steps**
  - MSDW team is working to restore the HIV/AIDS data to the de-identified OMOP database and data marts

New OMOP Projects starting in 2024

1. **2nd generation of OMOP tables**
   - Include Article 27-F (HIV/AIDS) data in de-identified OMOP
   - Support for HIPAA-compliant Limited Data Set (LDS) OMOP data marts
   - Support for OMOP version 5.4
2. **Integrate CNExT Cancer Registry data into OMOP**
3. **Integrate MSX billing data into OMOP**

Integrate CNExT Cancer Registry into OMOP

- Why is this important?
  - Researchers can create their own patient cohorts combining cancer registry data with EHR data
  - High-quality data manually abstracted from patients' charts by clinical subject-matter experts
- Cancer stage of disease
  - Clinical staging
  - Pathologic staging
- Tumor histology
- Researchers want these data in OMOP and searchable in Leaf, ATLAS, and TriNetX
- Limitations
  - Lag in chart abstraction

Integrate Billing Data from MSX into OMOP

- Why is this important?
  - MSDW's OMOP database is loaded primarily from the Epic EHR, which does not include billing data
- CPT4 procedure codes from billing data
  - Not in SlicerDicer or anywhere else in Epic
- Billing diagnosis codes (ICD-10-CM)
  - Do not necessarily overlap completely with diagnoses in clinical documentation
- Patient insurance status
  - Not in Epic (yet)
- Researchers want these data in OMOP and searchable in Leaf, ATLAS, and TriNetX

New Initiatives to Engage Researchers

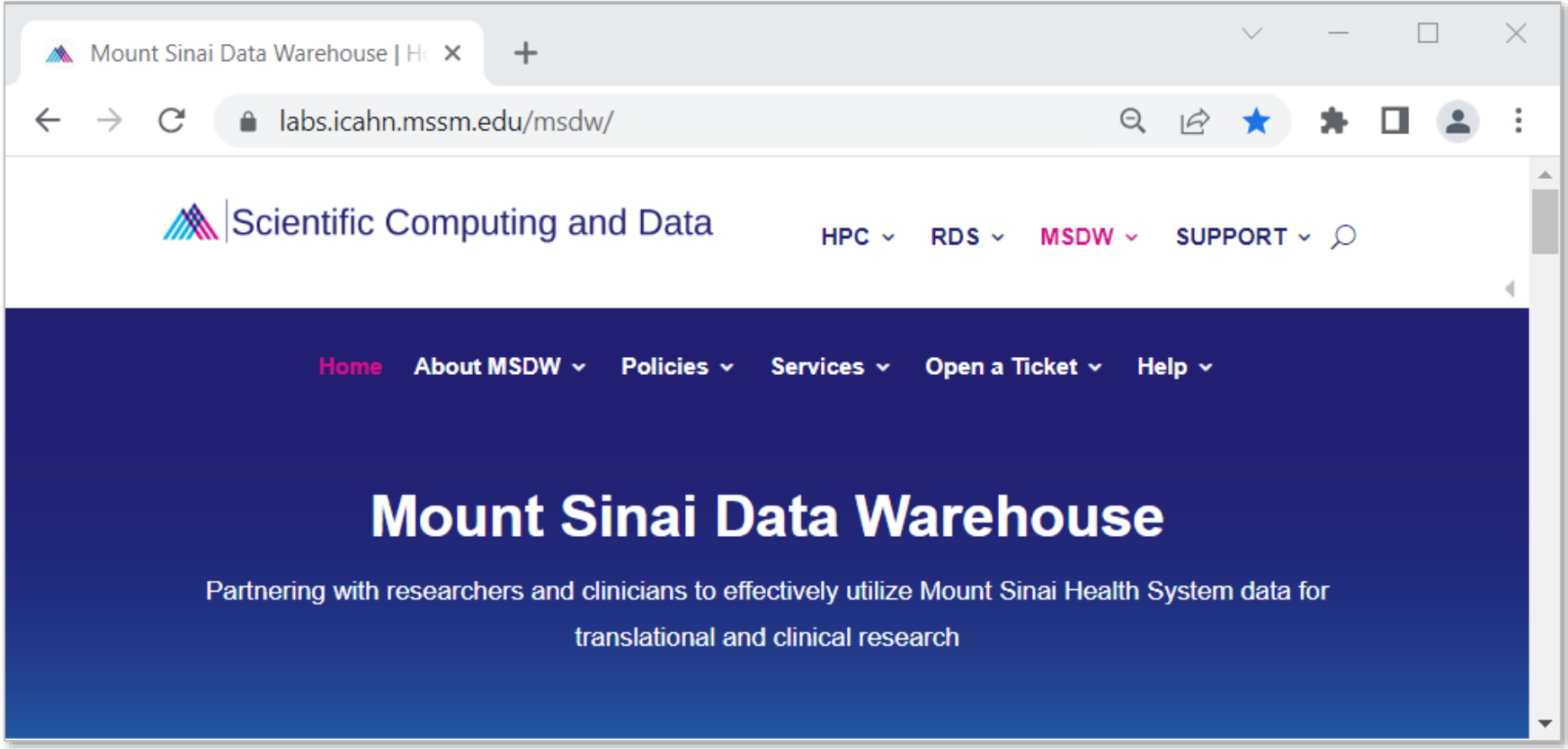
Lunch & Learn Events (in-person)

- Co-sponsors
  - Scientific Computing & Data
  - Brian Nickerson, Senior Associate Dean for Masters Programs
- Audience
  - Masters and PhD students
  - Postdoctoral researchers
  - CTSA-funded assistant professors
- Content
  - Real examples of how new researchers have used our services to achieve funding and publications
  - Presenters will include Scientific Computing and Data (SCD) customers, informatics fellows

Video Training Modules

- Short, recorded video sessions
- Each focusing on a different SCD service
- Promoted via various Mount Sinai channels: PEAK, Mount Sinai Daily, Research listserv, MSDW website, etc.

Our Website: msdw.mountsinai.org



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/

Thank You!