Leaf and ATLAS Query Tools

Patricia Kovatch, Dean
Sharon Nirenberg, MD, Lead Physician Informaticist
Naomi So MD, Physician Informaticist
Scientific Computing and Data
Icahn School of Medicine at Mount Sinai

March 5, 2025

Icahn
School of
Medicine at
Mount
Sinai

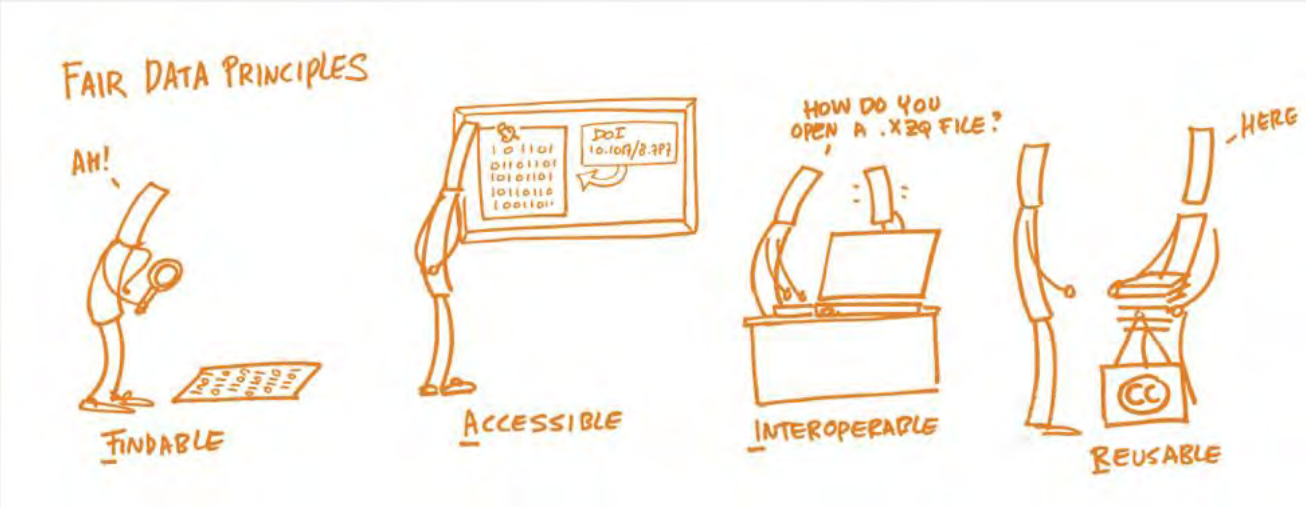
Agenda

1. The Mount Sinai Data Warehouse
2. Introduction to Leaf & ATLAS Cohort Query Tools
3. Leaf
4. ATLAS
5. MSDW Custom Data Set Request

Mount Sinai Data Warehouse

Scientific Computing FAIR Principles for Data

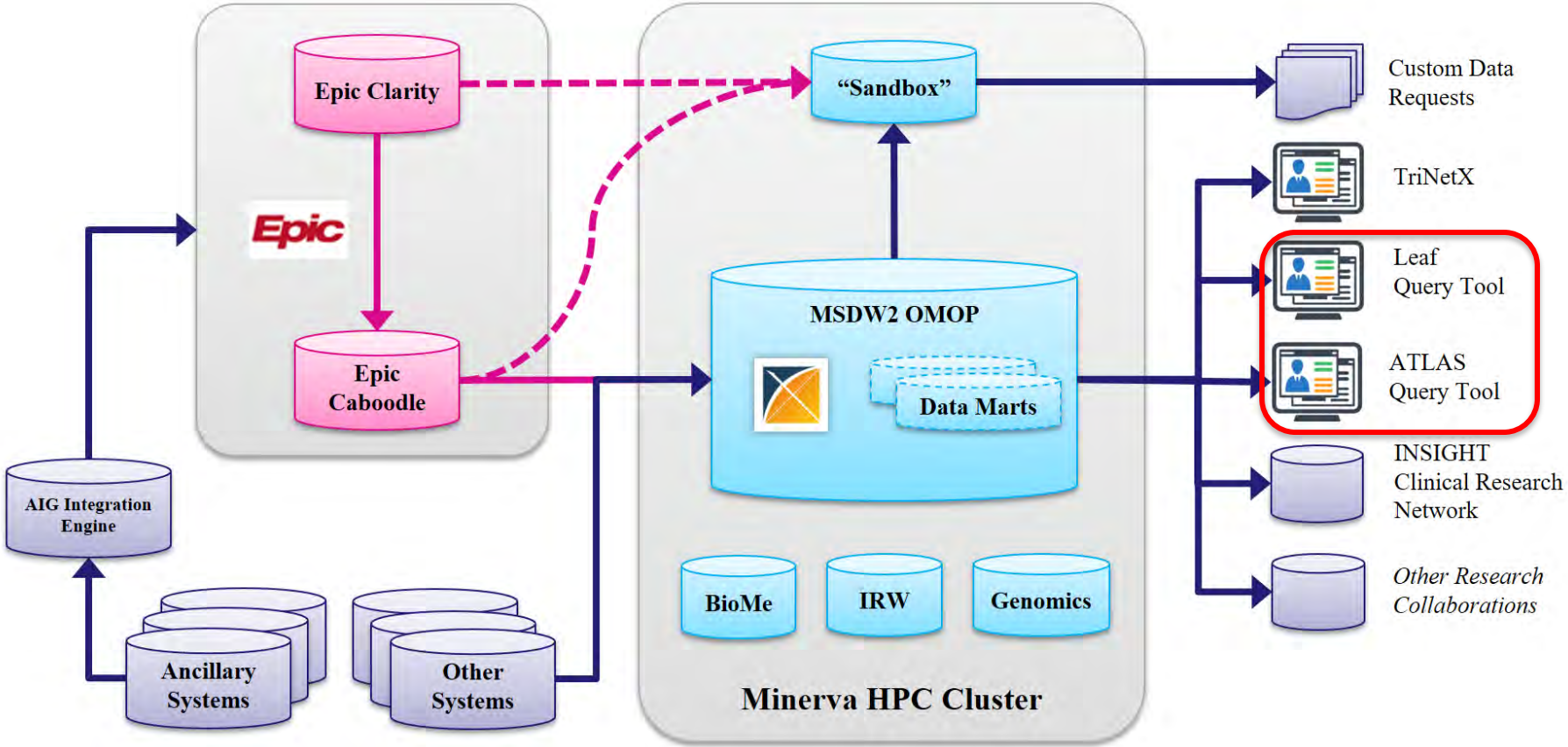
Findable | **A**ccessible | **I**nteroperable | **R**eusable / Reproducible



Source: NIH's Big Data to Knowledge (BD2K) Initiative (https://commonfund.nih.gov/bd2k)

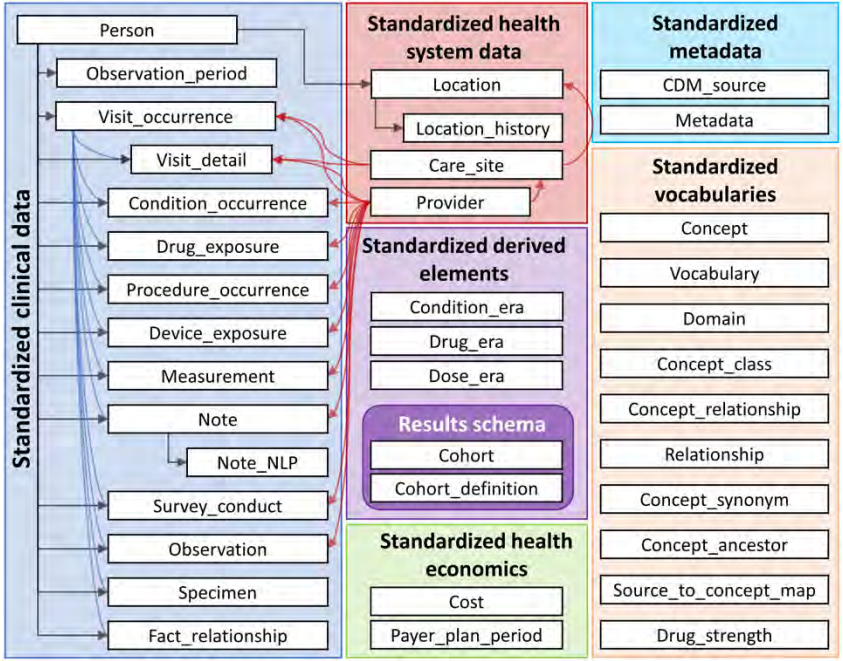
Image Source: https://book.fosteropenscience.eu/

Mount Sinai Data Warehouse Ecosystem



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

https://ohdsi.github.io/CommonDataModel/cdm60.html#Clinical_Data_Tables

MSDW Data Contents (*examples as of July 2024*)

OMOP Table	Record Type	Distinct Patients	Record Count
person	Patient Demographics	11,907,714	11,907,714
death	Patient Date of Death	51,578	51,578
visit_occurrence	Chart Documentation Event	5,916,518	101,871,186
visit_occurrence	ED Visit	1,264,805	3,123,540
visit_occurrence	Hospital Outpatient Visit	976,262	2,784,853
visit_occurrence	Inpatient Hospitalization	649,474	972,623
visit_occurrence	Inpatient Hospitalization from ED Visit	312,887	602,899
visit_occurrence	Mobile Unit Encounter	87,221	145,630
visit_occurrence	Outpatient Visit	4,477,479	85,285,020
visit_occurrence	Telehealth Visit	715,881	3,178,310
visit_occurrence	Urgent Care Visit	7,079	7,604
condition_occurrence	Billing Diagnosis	2,617,770	60,472,990
condition_occurrence	Encounter Diagnosis	4,237,020	119,783,046
condition_occurrence	Hospital Problem	922,623	3,593,212
condition_occurrence	Problem List	2,444,954	13,455,224
measurement	Flowsheet Measurement	1,820,482	226,459,093
measurement	Lab Component Result	4,136,901	1,065,127,536
measurement	Vital Signs	3,766,375	649,758,358
drug_exposure	Immunization Administration	1,439,713	8,789,505
drug_exposure	Immunization from Medication Order	217,360	280,661
drug_exposure	Immunization from Procedure Order	793,602	3,599,282

See MSDW website for the complete list: https://labs.icahn.mssm.edu/msdw/data-sources/

Introduction: Leaf & ATLAS Cohort Query Tools

Self-Service Cohort Query Tools

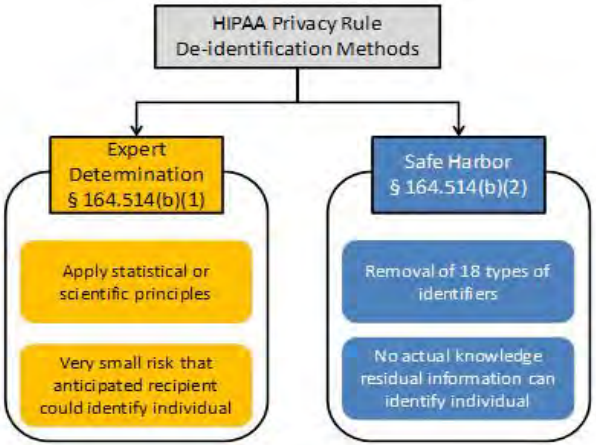
	Leaf 	Atlas 
Development	Nic Dobbins, Univ. of Washington, plus collabs., including at ISMMS	OHDSI community: www.ohdsi.org
License	Free and Open-Source Software (FOSS)	
Tradeoff	Easier, quicker, less powerful	Harder, laborious, more powerful
Data available	De-identified	De-identified
Capabilities	• Simple Boolean logic • Predefined stats & visualizations • Can download lists of patients (with masked IDs)	• Sophisticated logic • Customized stats & visualizations • Save your work and reuse parts • Run entire statistical analyses • **No data downloads**

See more details at https://labs.icahn.mssm.edu/msdw/services/

What is PHI? What is De-identification?

"**PHI** (Protected Health Information) is information (demographic, financial, social, clinical) relating to an individual's past, present, or future health history, treatment, or payment for health care services that is held or transmitted by a CE or its BA that identifies the individual or ***for which there is a reason to believe it can be used to identify the individual***."

De-identification is the process by which PHI is rendered not individually identifiable. The HIPAA Privacy Rule establishes two methods to de-identify PHI:



Types of Identifiers

- Name
- Street Address, city, county, zip code (the first three digits of the zip code may be used if there are more than 20,000 people in the zip code)
- All element of dates (except year), including dates of birth, admission, discharge or death
- All ages over 89
- All telephone/fax numbers
- Fax number
- E-mail addresses
- Social Security Number (SSN)
- Medical Record Number (MRN)
- Health plan beneficiary number
- Account numbers (health plan IDs, credit card, bank, invoice #s)
- Certificate/License numbers
- Vehicle identifiers, including license plate numbers
- Device identification and/or serial number
- Uniform Resource Locator (URL)
- Internet Protocol (IP) address
- Biometric identifiers (finger, voiceprints, etc)
- Full face photographic images and other comparable images
- Any other unique identifying number, characteristic, or code

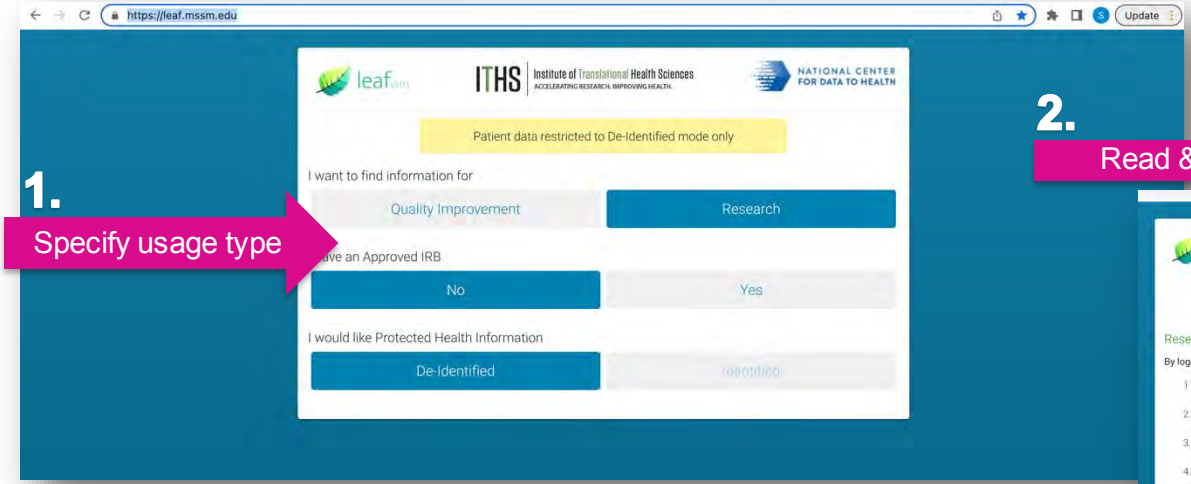
Leaf Query Tool

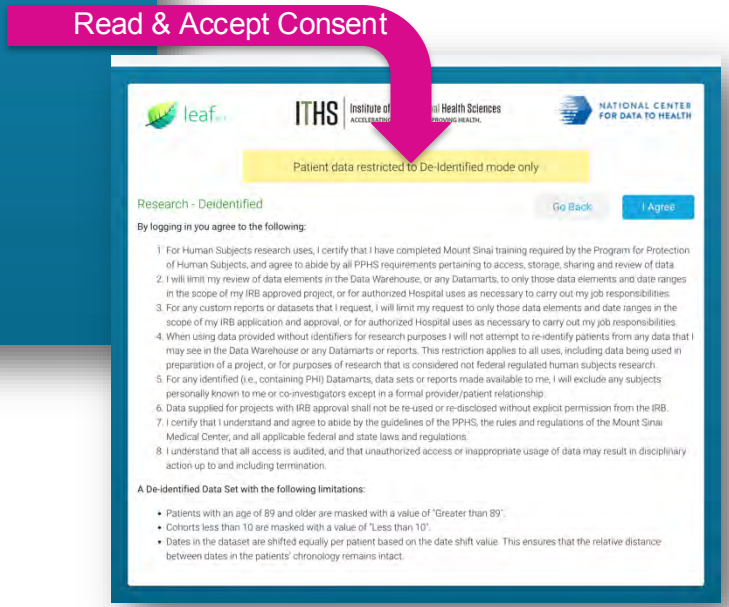
Features of the Leaf Application

- Open-source, model-agnostic and data-driven web application for cohort discovery
- Simple drag-and-drop user interface
- Simple Boolean logic-based searches
- View pre-defined basic stats and visualizations on your cohort
- Save queries for later

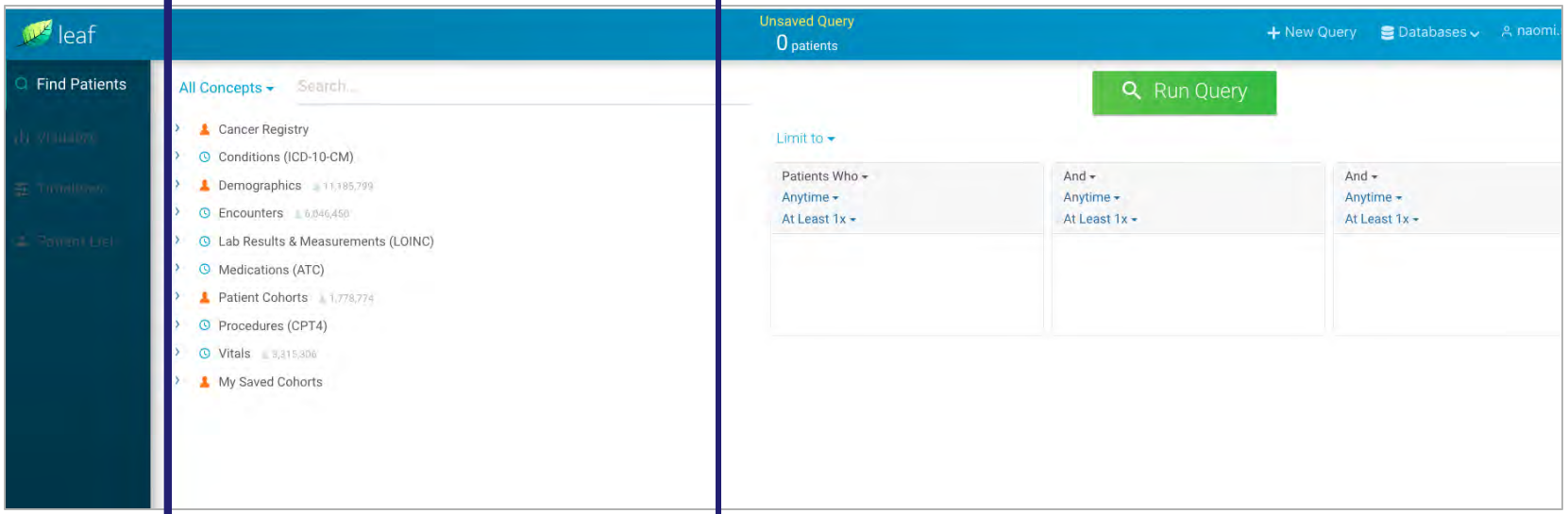
Accessing Leaf

- All Mount Sinai Faculty, staff or students can access Leaf at https://leaf.mssm.edu
- Requires VPN access and use of your Mount Sinai Login credentials

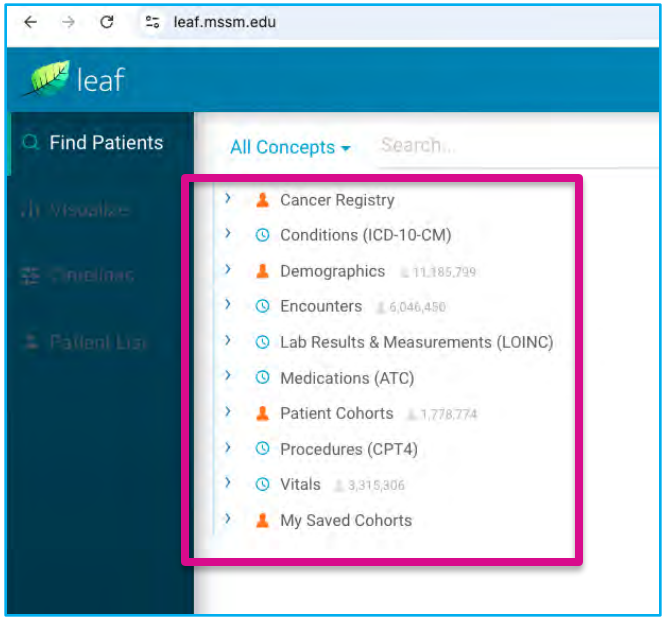




Leaf Landing Page

Left Side Navigation	Concept Search Bar	Build and Run Query
		
Toggle screens	Search by text or code	Use select concepts and Boolean logic to build query

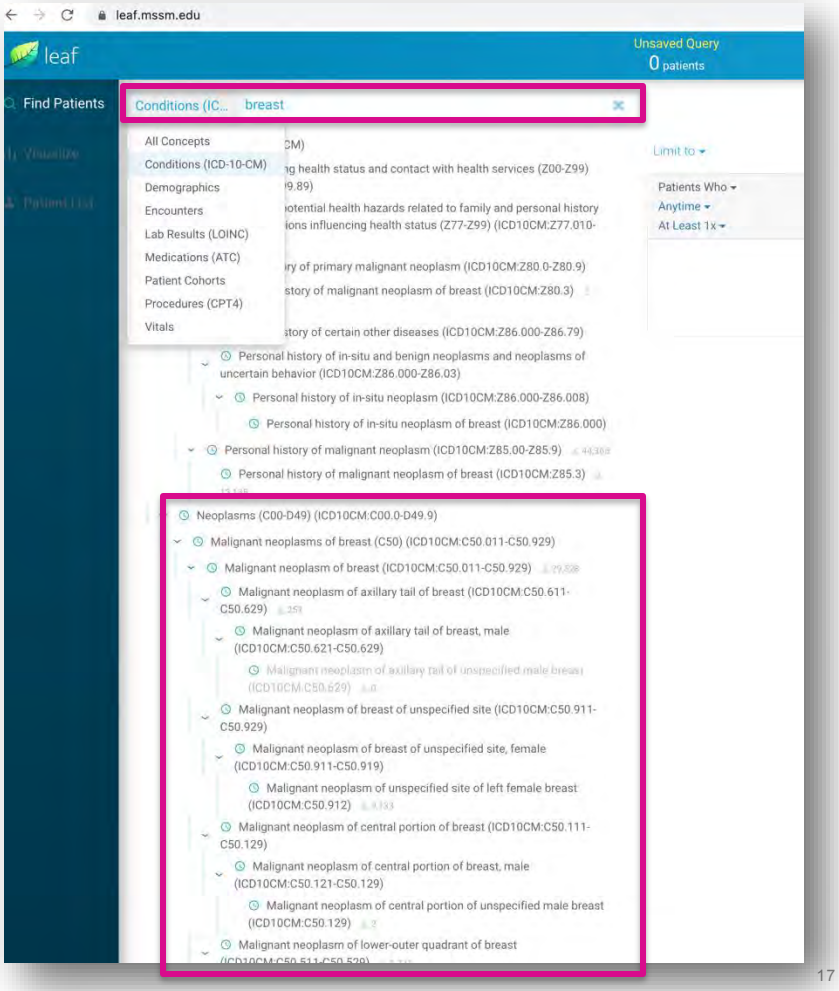
Searchable Data Domains



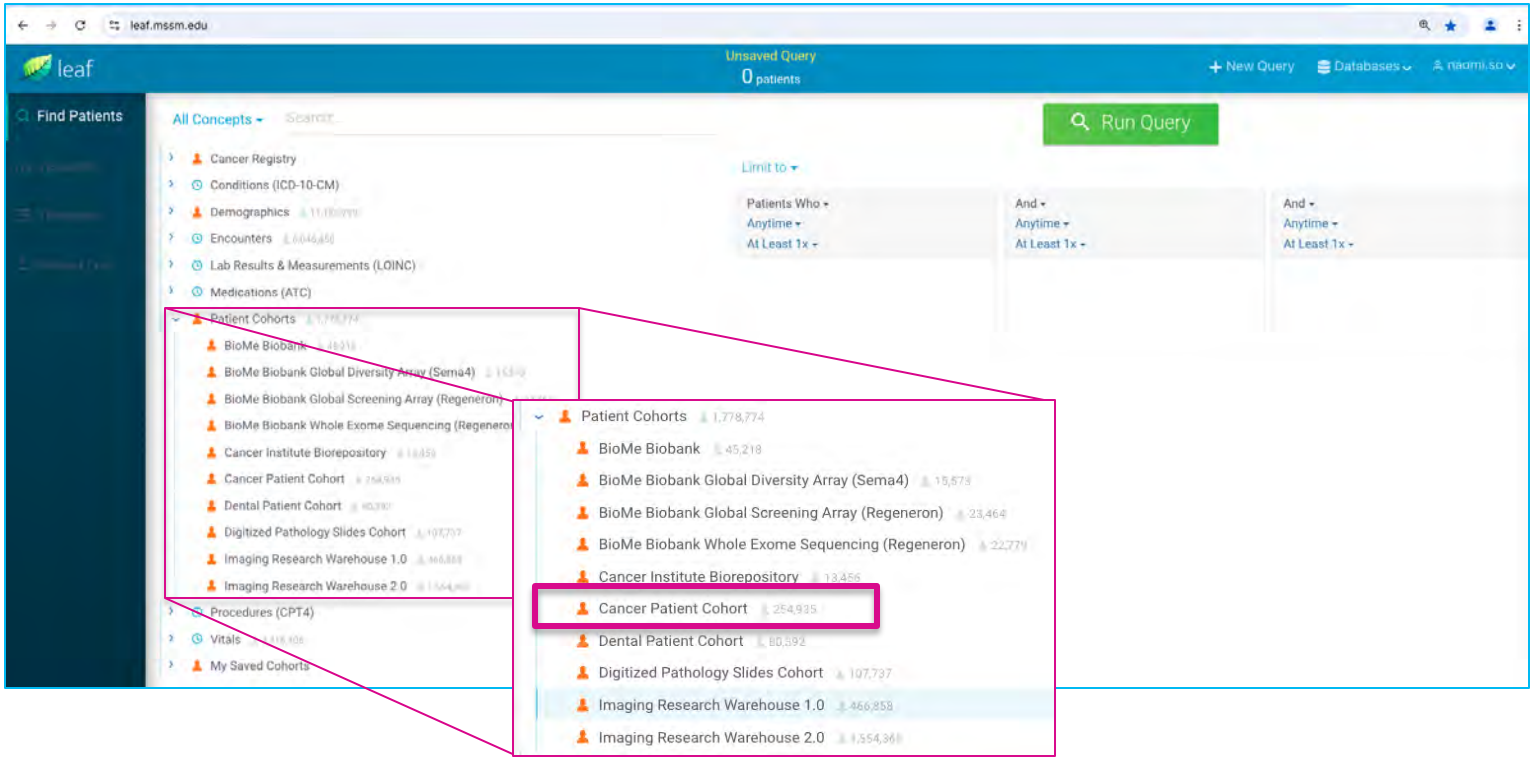
Domains	Vocab	Content	Time
Conditions	ICD-10-CM	Descriptive diagnoses and codes	Encounter-based
Demographics		Age, Gender, Race, Ethnicity, Vital Status	Time-invariant
Encounters		ED visit, Inpatient, Ambulatory, Telehealth	Encounter-based
Lab Results	LOINC	Lab Orders	Encounter-based
Medications	ATC	Medications Orders and Administrations	Encounter-based
Procedures	CPT-4	Procedures	Encounter-based
Vitals	LOINC	BMI, O2 sat, Pulse, Respiratory Rate, etc.	Encounter-based

Identifying Concepts

- Two ways to search for concepts
  - Free Text search
  - Expand concept trees using left-hand arrows
- Each concept is denoted by a population quantity to the right
- To select a concept, click on it and drag it to the query box
  - The concept and all the dependent nodes will be included



Institutional Patient Cohorts are Searchable in Leaf

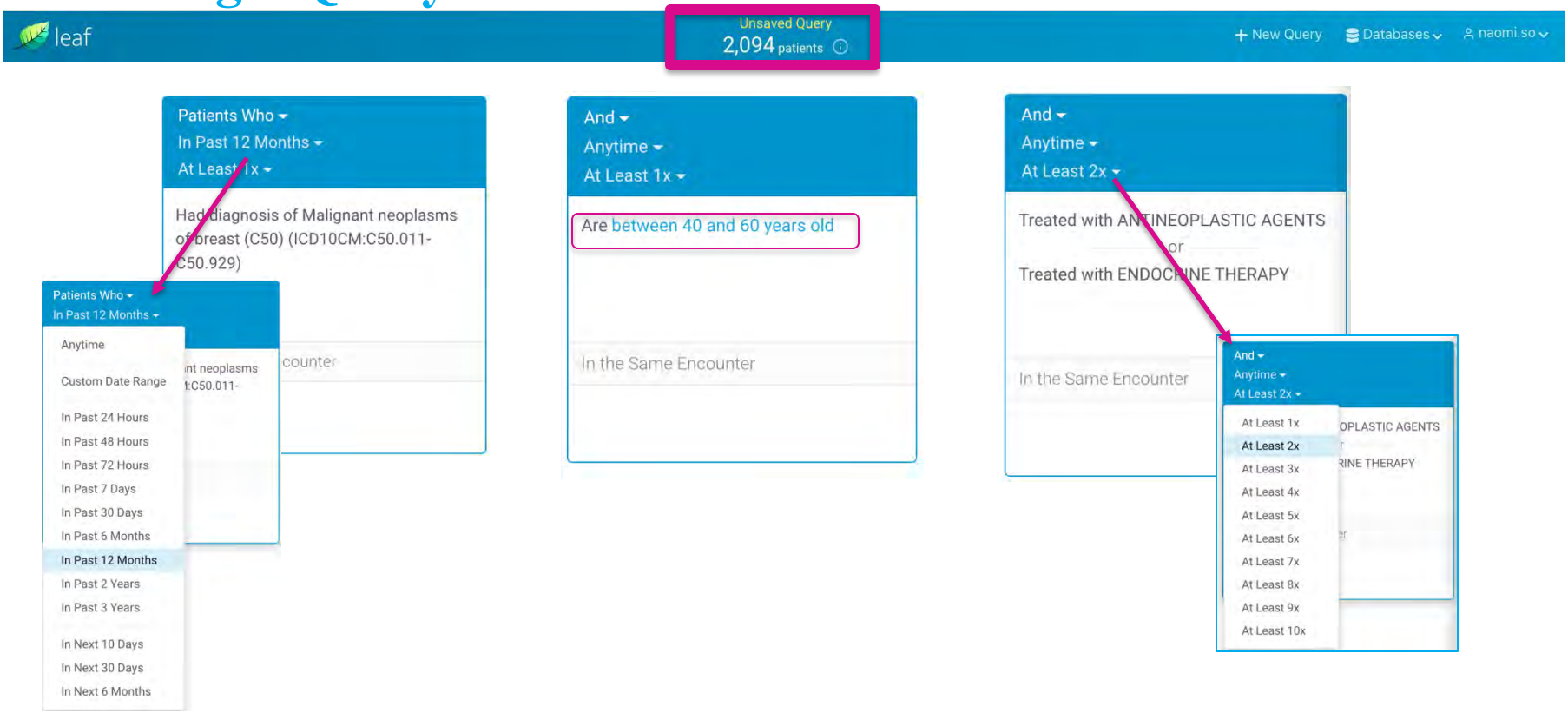


Use Leaf to query the Cancer Patient, BioMe or IRW Cohorts

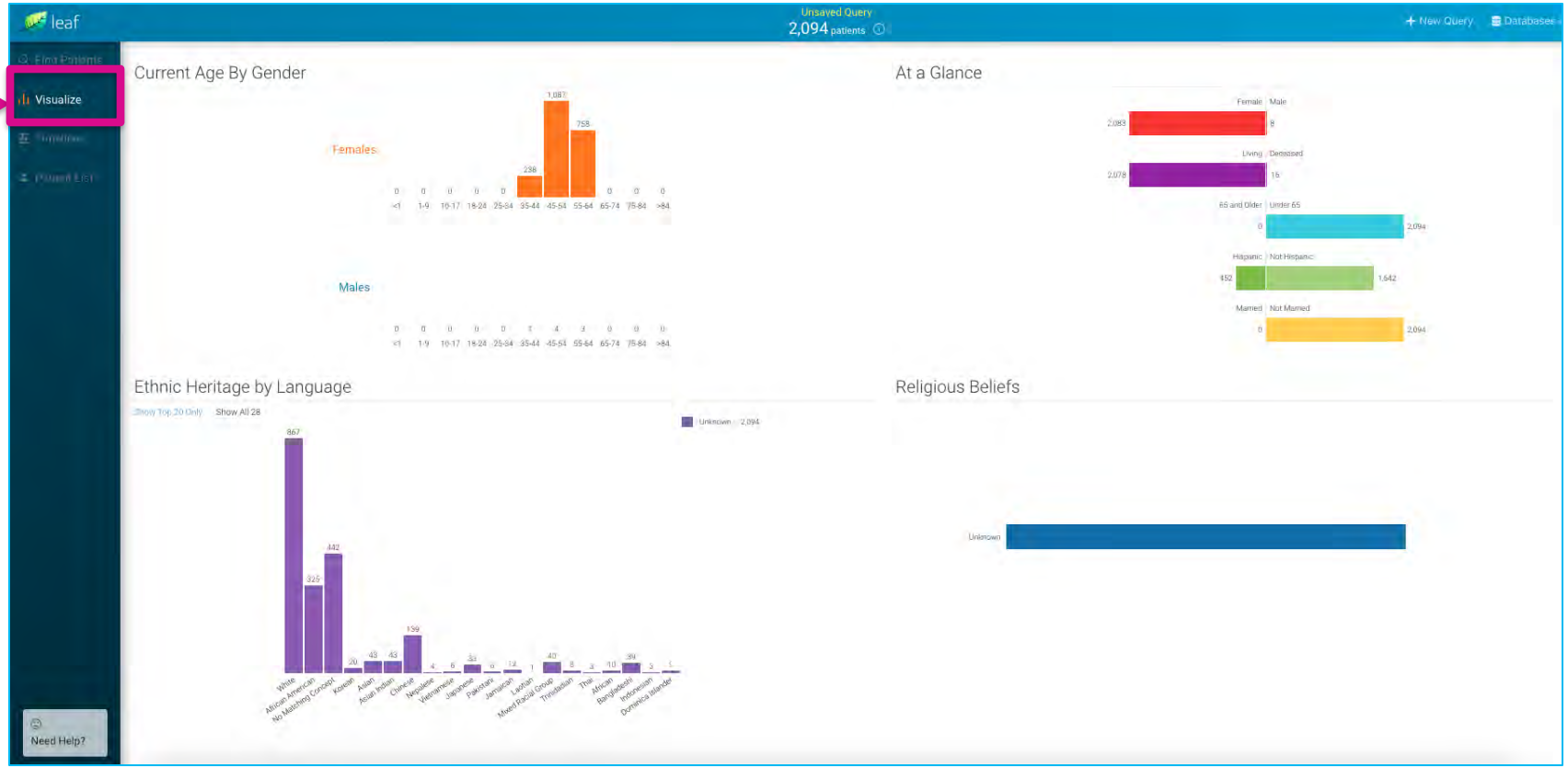
Leaf – Patient Cohorts

Patient Cohorts on Leaf	Description
BioMe Biobank	Patients who submitted tissue samples to Mount Sinai's BioMe Biobank
BioMe Biobank Global Diversity Array - Sem4	Patients who submitted tissue samples to Mount Sinai's BioMe Biobank and have had their DNA analyzed with Illumina's Global Diversity Array by Sema4
BioMe Biobank Global Screening Array – Regeneron	Patients who submitted tissue samples to Mount Sinai's BioMe Biobank and have had their DNA analyzed with Illumina's Infinium Global Screening Array by Regeneron
BioMe Biobank whole Exome Sequencing – Regeneron	Patients who submitted tissue samples to Mount Sinai's BioMe Biobank with whole exome sequence (WES) data generated by Regeneron
Cancer Institute Biorepository	
Cancer Patient Cohort	Patients who have been diagnosed with cancer, refreshed on a monthly basis around the 15th of every month
Imaging Research Warehouse 1.0	Patients who have image data in version 1.0 of the Imaging Research Warehouse (IRW)
Imaging Research Warehouse 2.0	Patients who have image data in version 2.0 of the Imaging Research Warehouse (IRW)

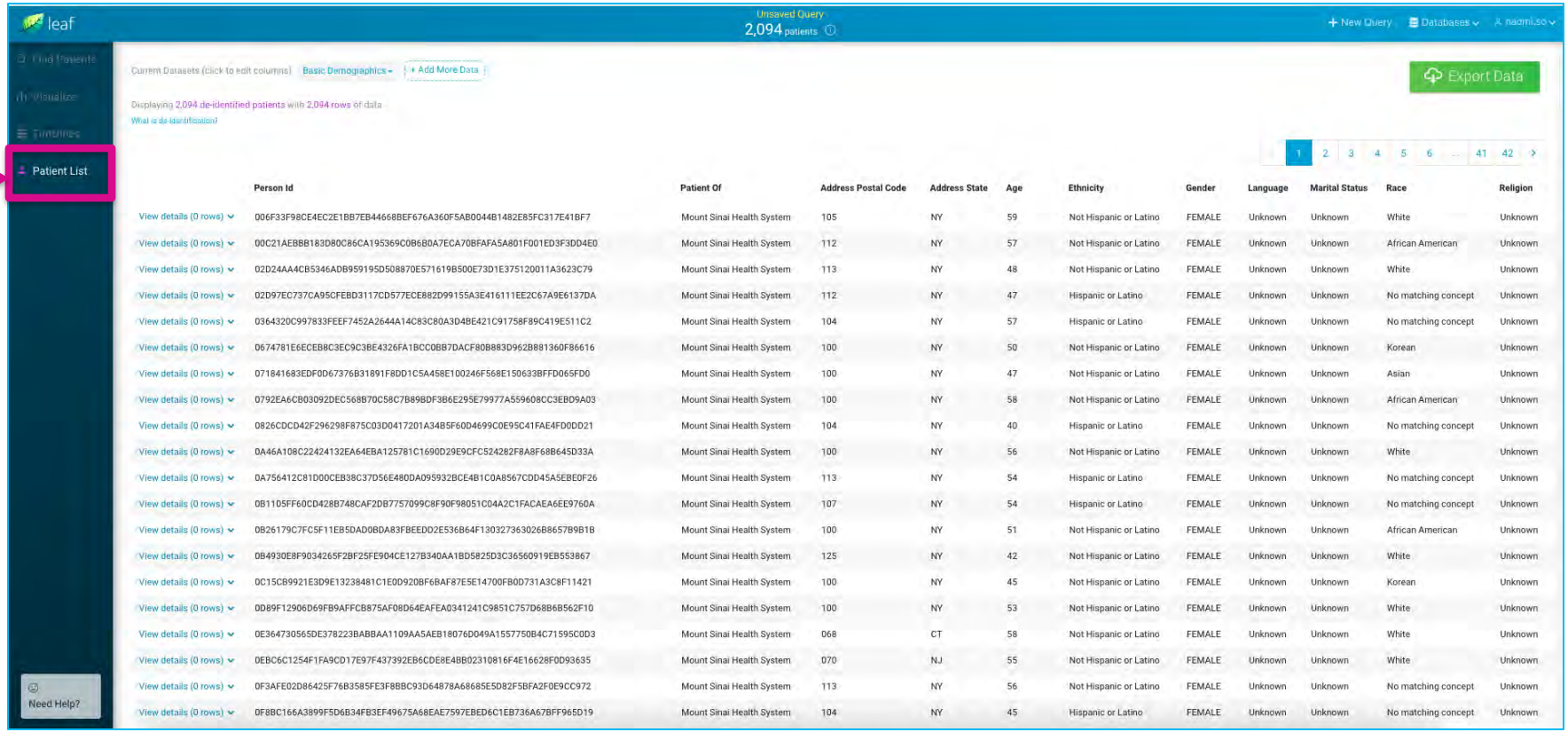
Building a Query



Basic Cohort Demographics



Patient List



Leaf - Timeline

Used to explore temporal relationships of additional clinical events (aka concepts) to your defined patient cohort.

Sample Query:

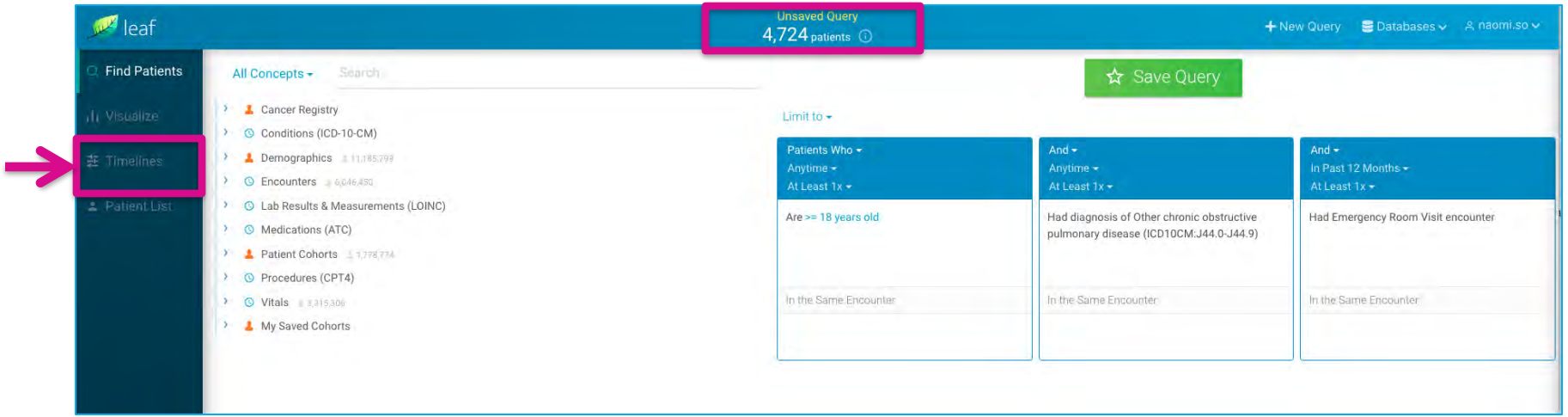
How many patients >=18 y.o. with a diagnosis of COPD (Chronic Obstructive Pulmonary Disease) had an ED visit in the past 12 months?

Secondly, what percentage of these patients had any of the following clinical events after their ED visit?

- **An inpatient visit**
- **Diagnosis of Lung Cancer**

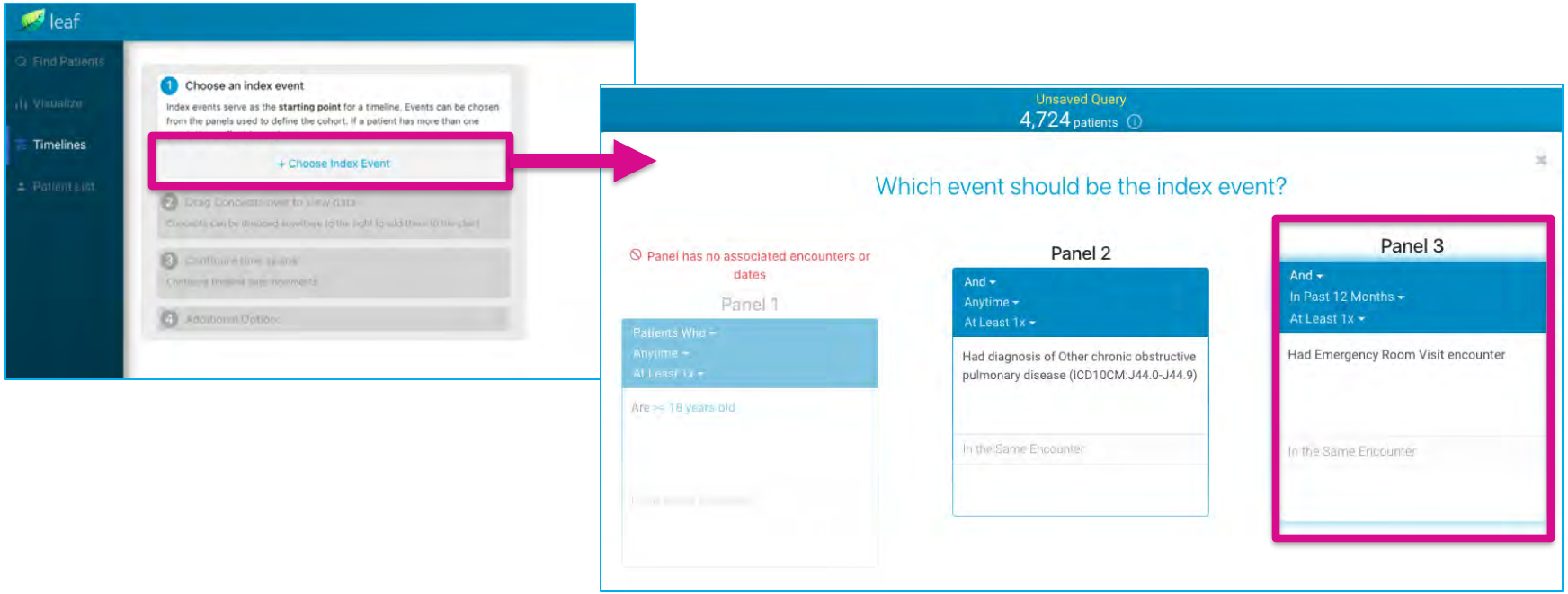
Leaf - Timeline

1. Build and Run query to identify patient cohort
2. Click on **Timelines** from the left-hand menu



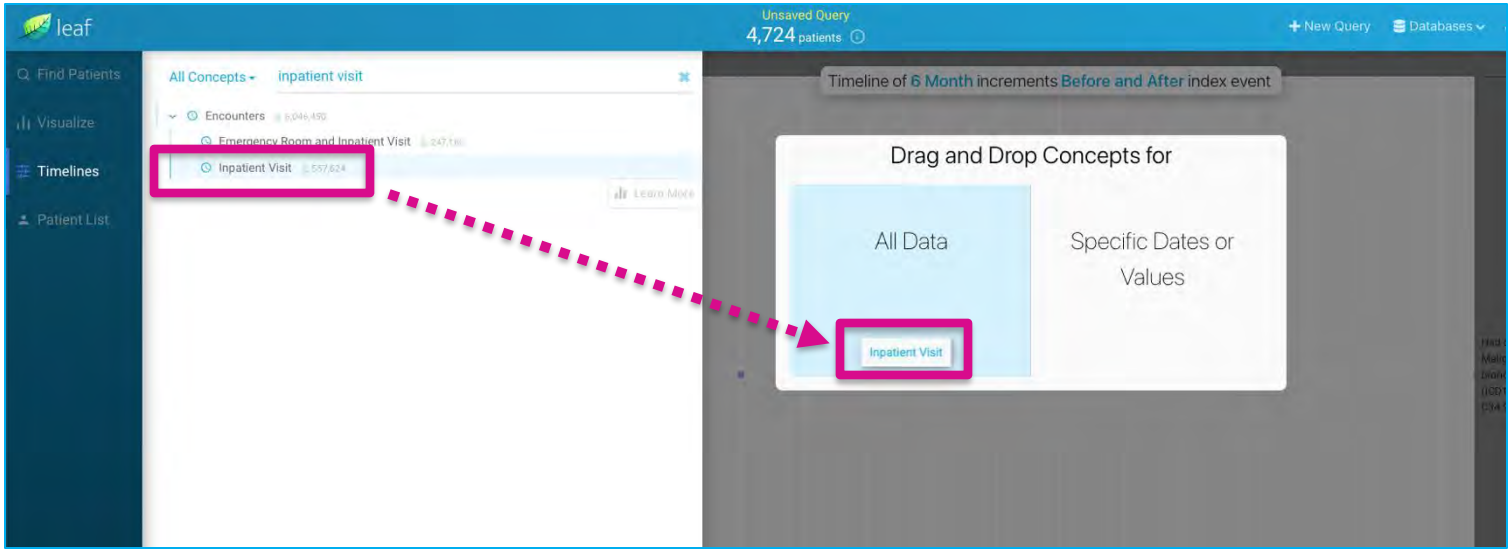
Leaf - Timeline

Identify an **Index Event** for your cohort. This is the starting point for your timeline and allows you to view other clinical events (aka concepts) that happened before and/or after, at defined time intervals.



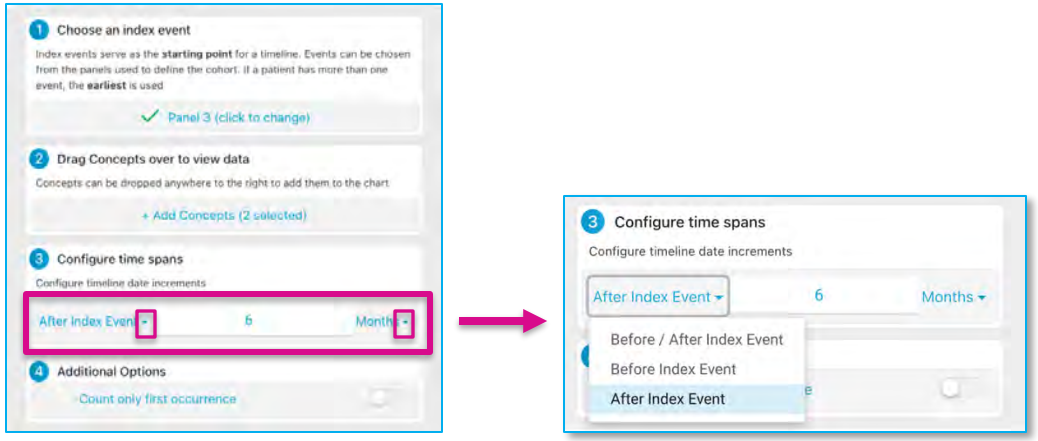
Leaf - Timeline

Add concepts of interest to your timeline by dragging and dropping from the ***All Concepts*** menu on the left to the ***Drag and Drop Concepts for*** window on the right.

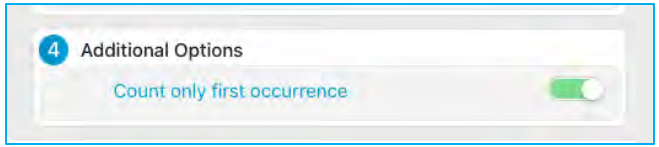


Leaf - Timeline

Under **Configure Time Spans**, adjust timeline intervals..

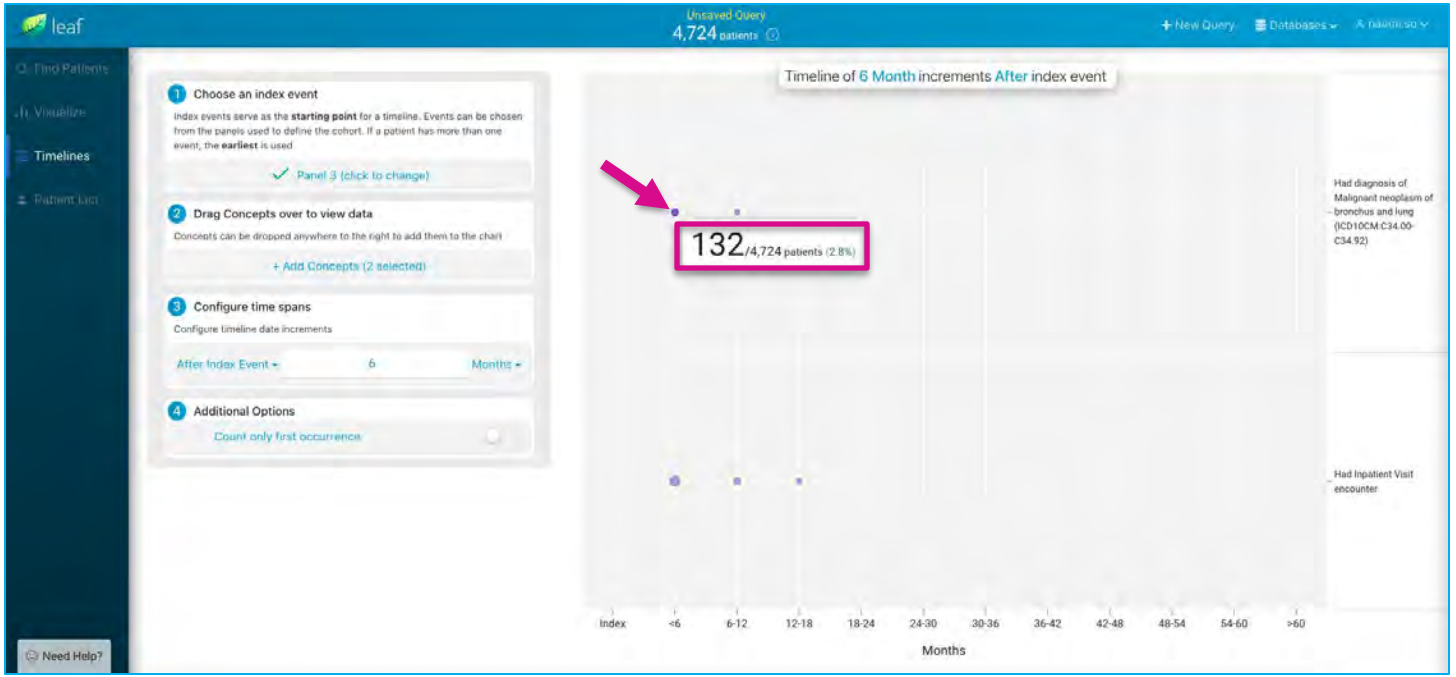


Under **Additional Options**, click on ***Count only first occurrence*** to *only* take into consideration the first time that each event took place (this applies to all added concepts in your timeline)



Leaf's Timeline

Hover over the circles in your timeline to view how many patients from your initial cohort fall within that category (as defined by the concept and time interval)



To remove a clinical event (aka concept) from your timeline, hover over it and click on the **X**

Live Demo - Leaf

ATLAS Query Tool

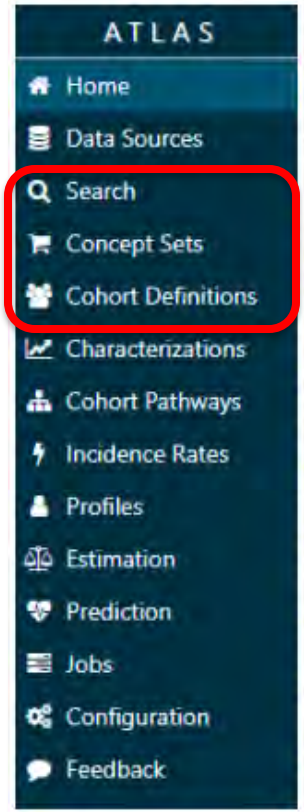
ATLAS

- A web-based application - design and execute observational analyses to generate real world evidence from patient level clinical data
- **Common Data Model** - A convention for representing healthcare data that allows portability of analysis
- **Concept** - A term (with a code) defined in a medical terminology, all clinical events in the OMOP CDM are expressed as concepts
- **Concept set** is an expression representing a list of concepts that can be used as a reusable component in various analyses
- **Cohort** is a set of persons who satisfy one or more inclusion criteria for a duration of time

ATLAS Access

- All Mount Sinai Faculty, staff or student can access ATLAS at https://atlas.msdw.mountsinai.org
- Requires VPN access and Mount Sinai School Credentials to log in
- Mount Sinai users with a Hospital account may navigate to SailPoint and request a Mount Sinai School account.
- You will be required to read and accept the SNOMED INTERNATIONAL SNOMED CT LICENSE AGREEMENT
- Sign in using your school credentials through the button on the top right corner of the interface

ATLAS - Interface



Home: Permalink redirects you to the Atlas landing page.

Data Sources: Provides capability to review standardized reporting for each of the data sources configured for your Atlas environment. Here, review available populations and data sets. From select drop-down menus, select from any available observational database(s). Subsequently, select from any of the corresponding standardized reports available within the previously selected source.

Search: Enables you to search the OMOP standardized vocabularies, and understand and apply concepts within those vocabularies.

Concept Sets: Enables you to create your own set of codes that will be used throughout the standardized analyses. These sets can be saved and reused in all your analyses.

Cohort Definitions: Provides ability to construct a set of persons who satisfy one or more criteria for a duration of time, and these cohorts can serve as a basis of inputs for all subsequent analyses.

Characterizations: Allows you to look at one or more of your defined cohorts and summarizes characteristics about those patient populations in an analytic capability.

Cohort Pathways: Reviews the sequence of clinical events that that occur within one or more populations.

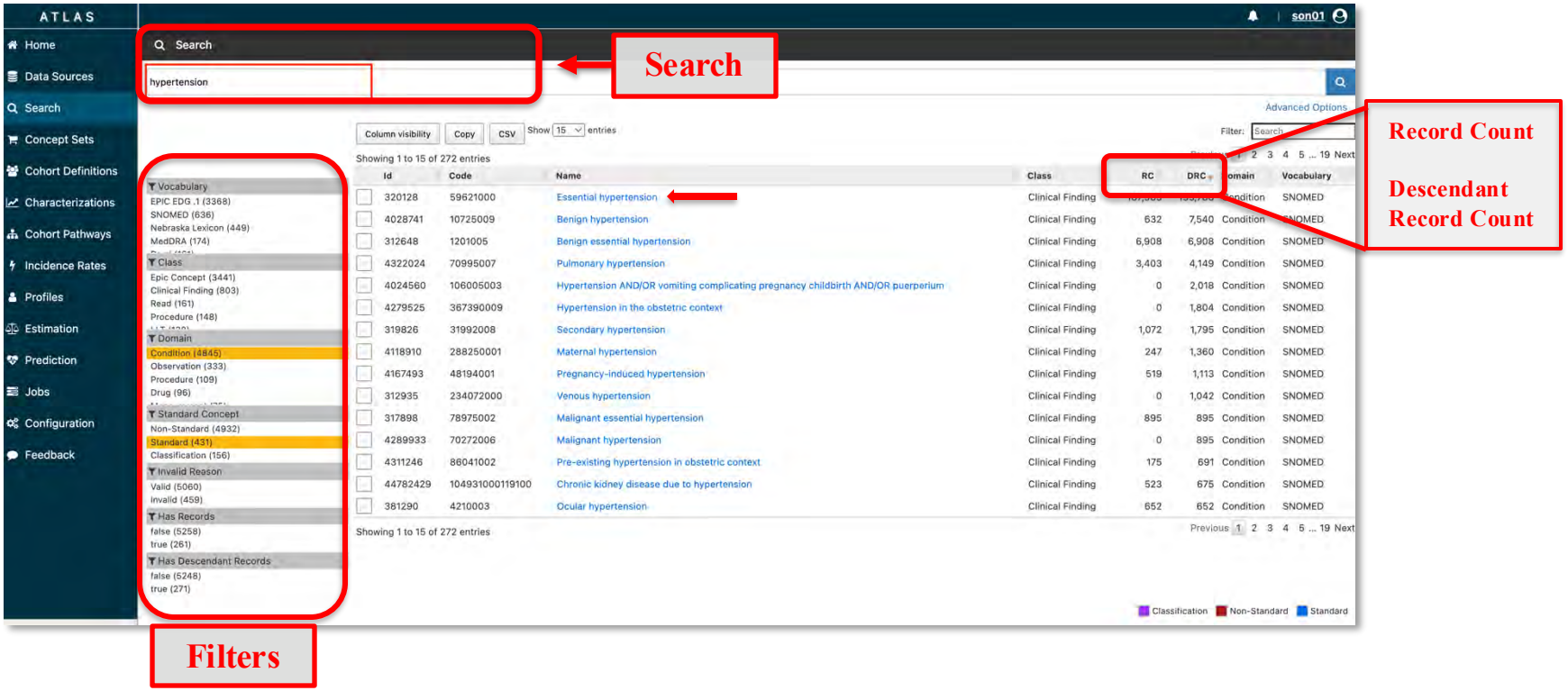
Incidence Rates: Provides the ability to estimate the incidence of outcomes within target populations of interest.

Profiles: Explores an individual patient's longitudinal observational data to summarize an individual's situation.

Estimation: Conducts population-level effect estimation studies using a comparative cohort design. Comparisons between one or more target and comparator cohorts can be explored for a series of outcomes.

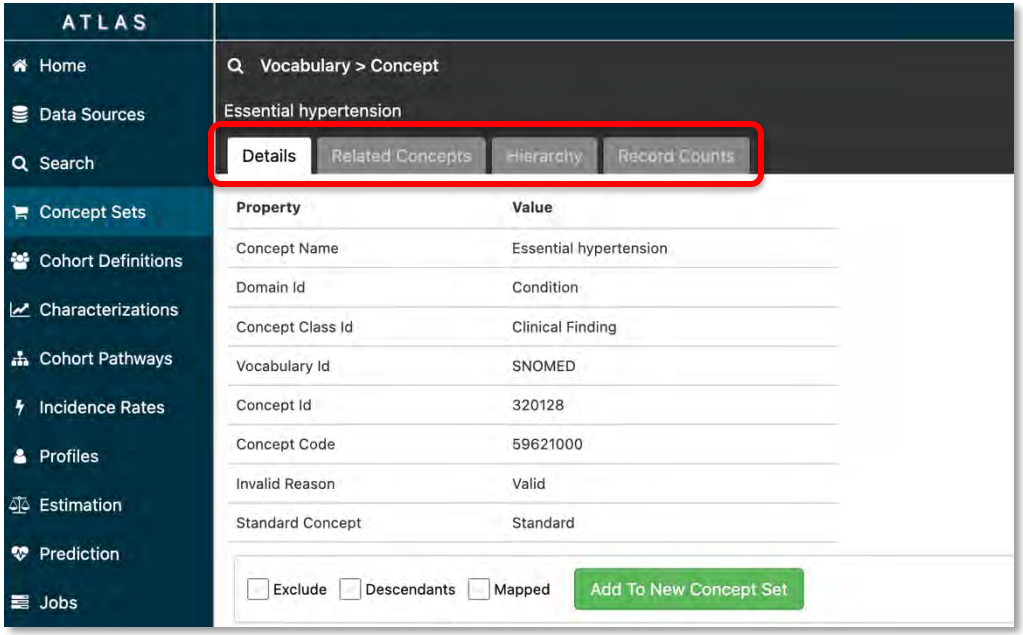
ATLAS - Search

- Enables you to search the OMOP standardized vocabularies, and understand and apply concepts within those vocabularies



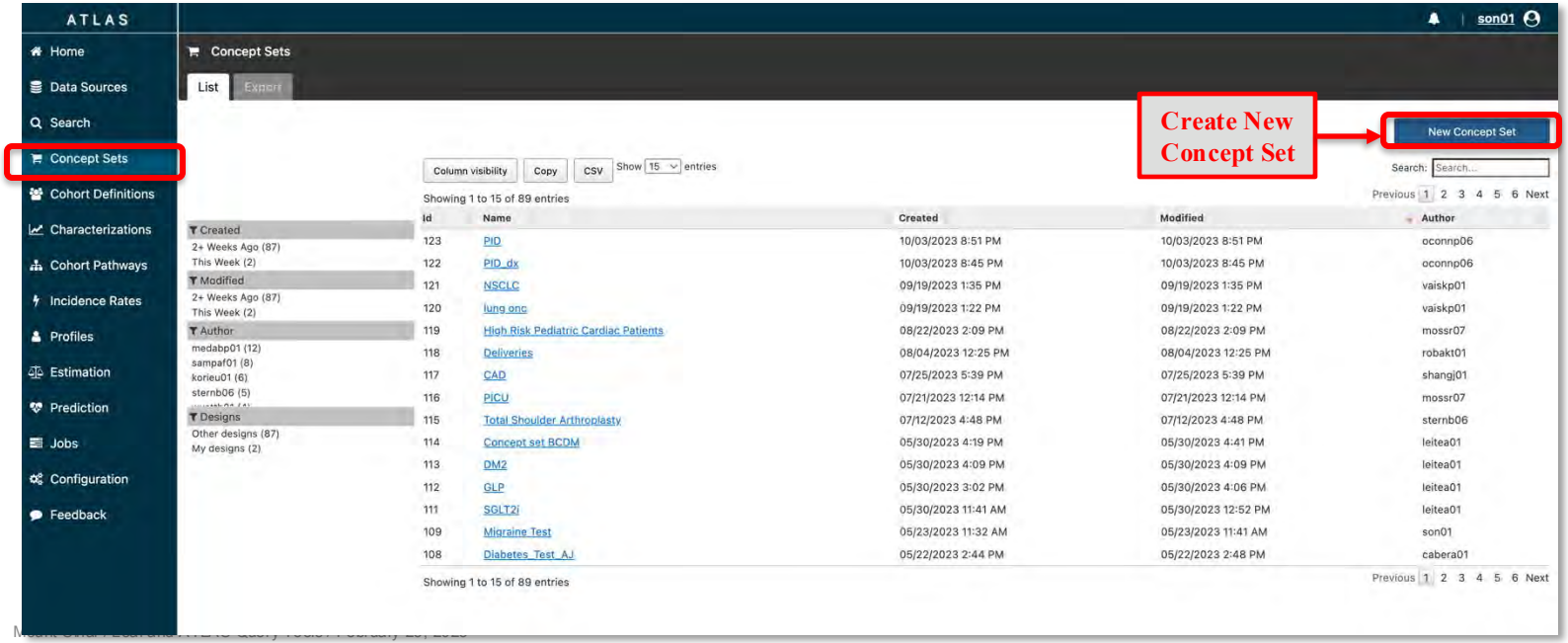
ATLAS – Search

- Clicking on a term will open a more detailed view within the vocabularies with the following tabs:
  - **Details** presents Vocabulary ID, Concept ID, Concept Code, and other property values connected to the record
  - **Related Concepts** provides other vocabulary for similar terms that may specify or broaden the search
  - **Hierarchies** indicates parents and children of the concept within the OMOP vocabulary
  - **Record Counts** displays the source of the records as well as the quantity



ATLAS – Concept Sets

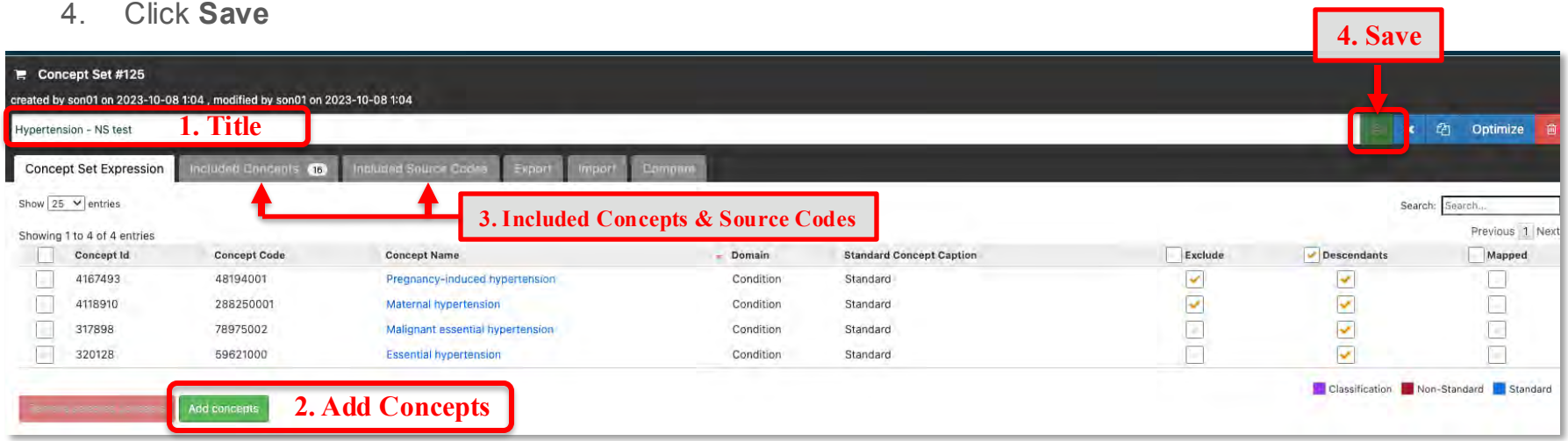
- Building blocks of ATLAS queries
- Expression that allow for identifying sets of concepts that can be grouped together and used as a reusable component in various analyses
- Can contain any set of concepts across any of the domains within the OMOP standardized vocabulary. These can be customized so that different terms can be expressed in one item.



ATLAS – Create New Concept Set

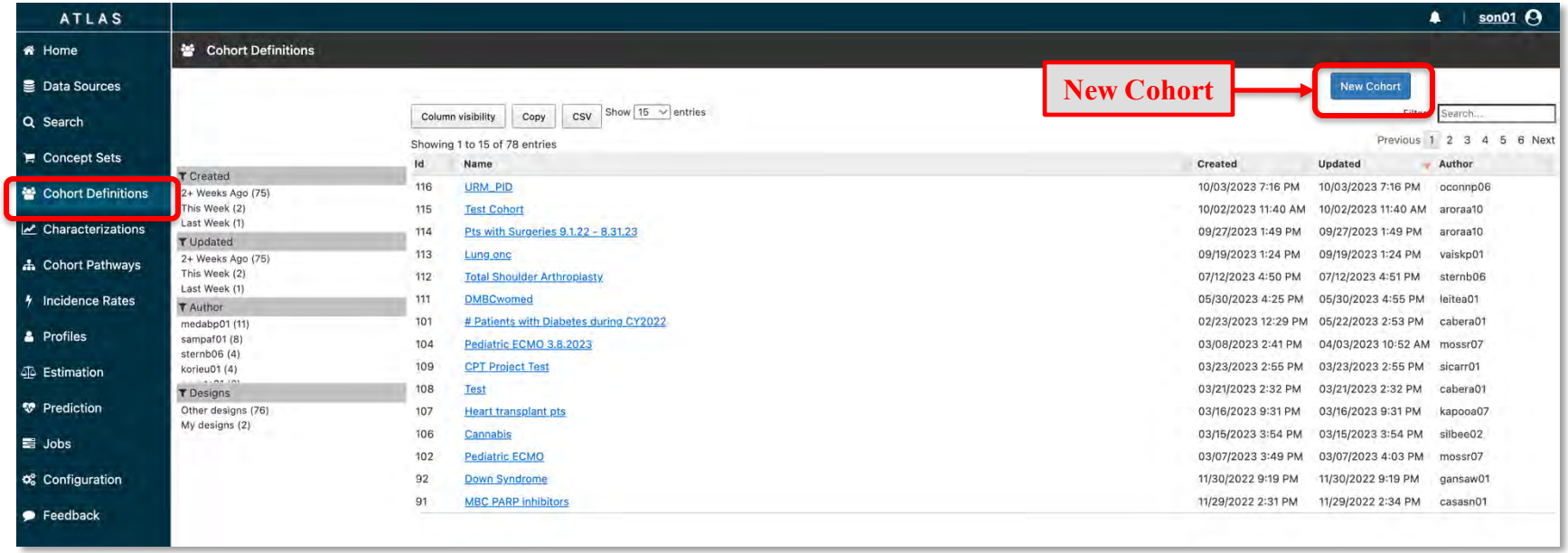
New Concept Set

1. Title your Concept Set (i.e. *Hypertension – NS Test*)
2. **Add concepts** → Search for concepts of interest (i.e. *essential hypertension*)
   - Select concepts to include or exclude, along with any of their associated Descendants
3. View **Included Concepts** and **Included Source Codes** under respective tabs
4. Click **Save**



ATLAS – Cohort Definitions

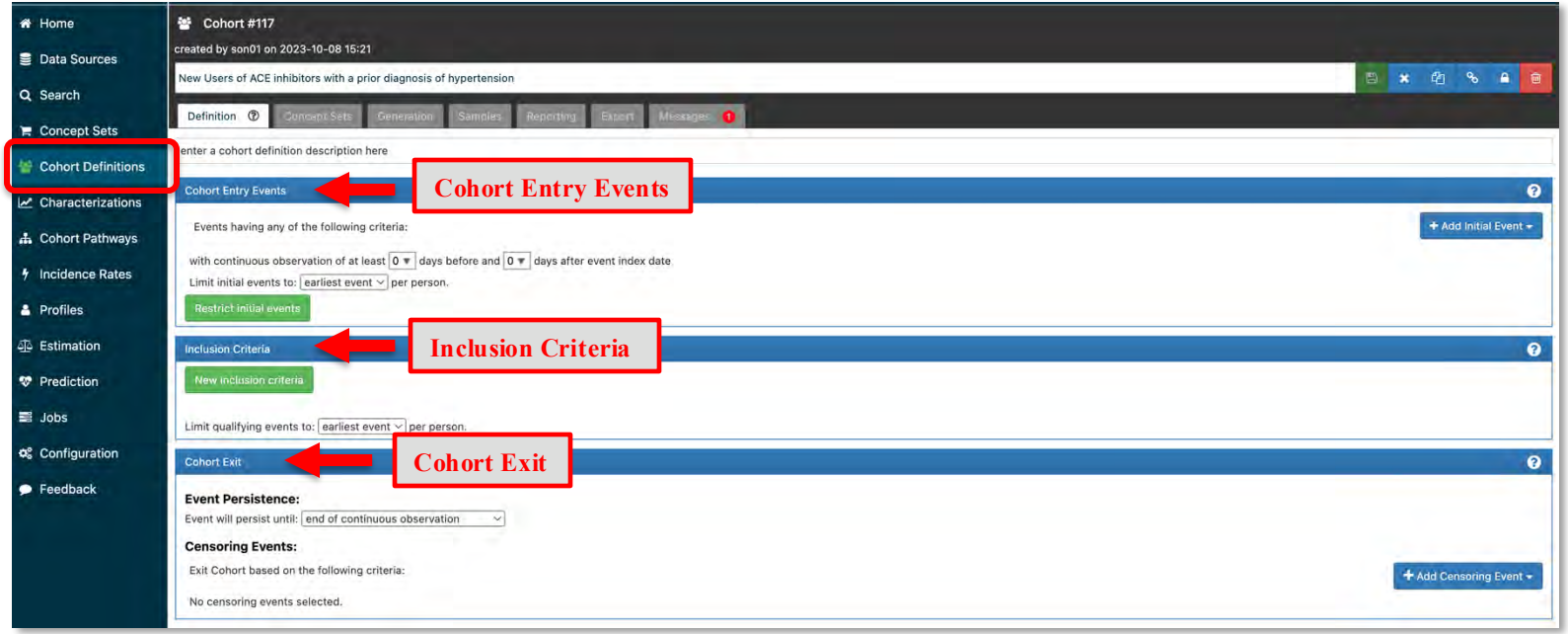
- Where you define the cohort inclusion criteria that must be satisfied for a duration of time
- Can serve as a basis of inputs for subsequent analyses
- Click **New Cohort** to create a new definition



ATLAS – Cohort Definitions

Cohort Criteria:

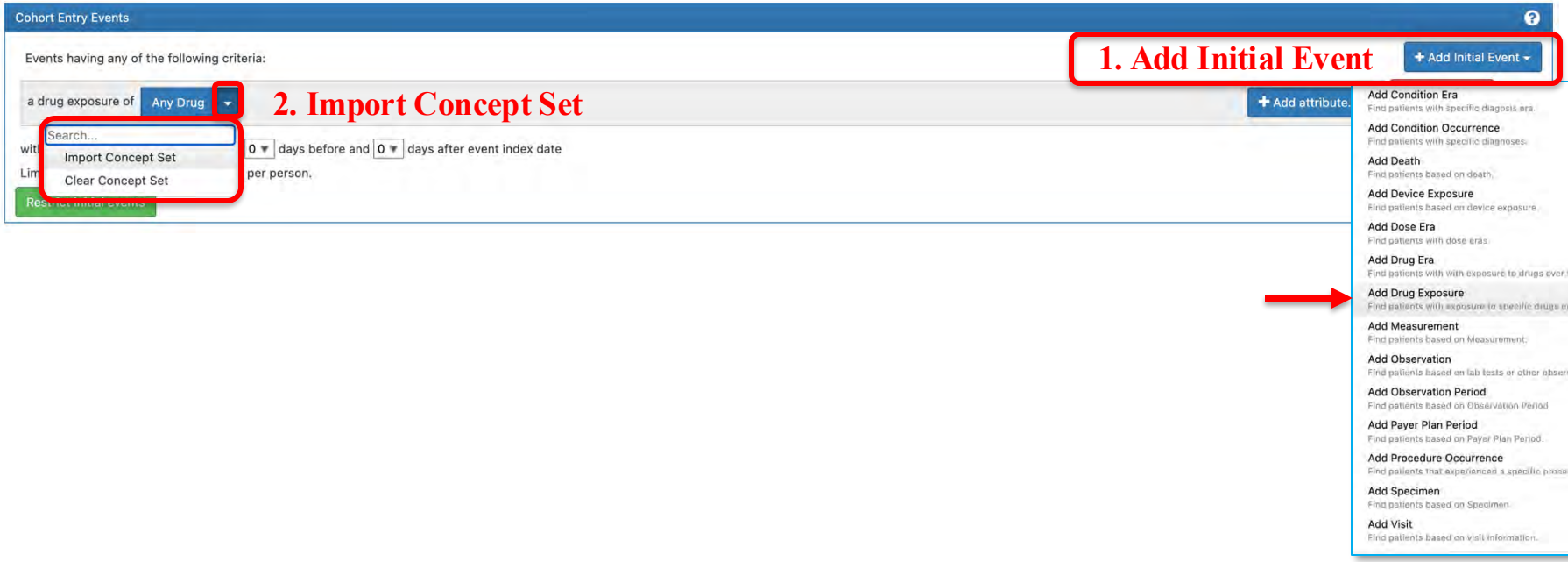
- **Cohort Entry Event:** What must be observed so that someone enters the cohort?
- **Inclusion Criteria:** Use concept sets to apply specific criteria to cohort entry event to identify subpopulation
- **Cohort Exit:** How does person leave the cohort of interest?



ATLAS – Cohort Definitions: Cohort Entry Events

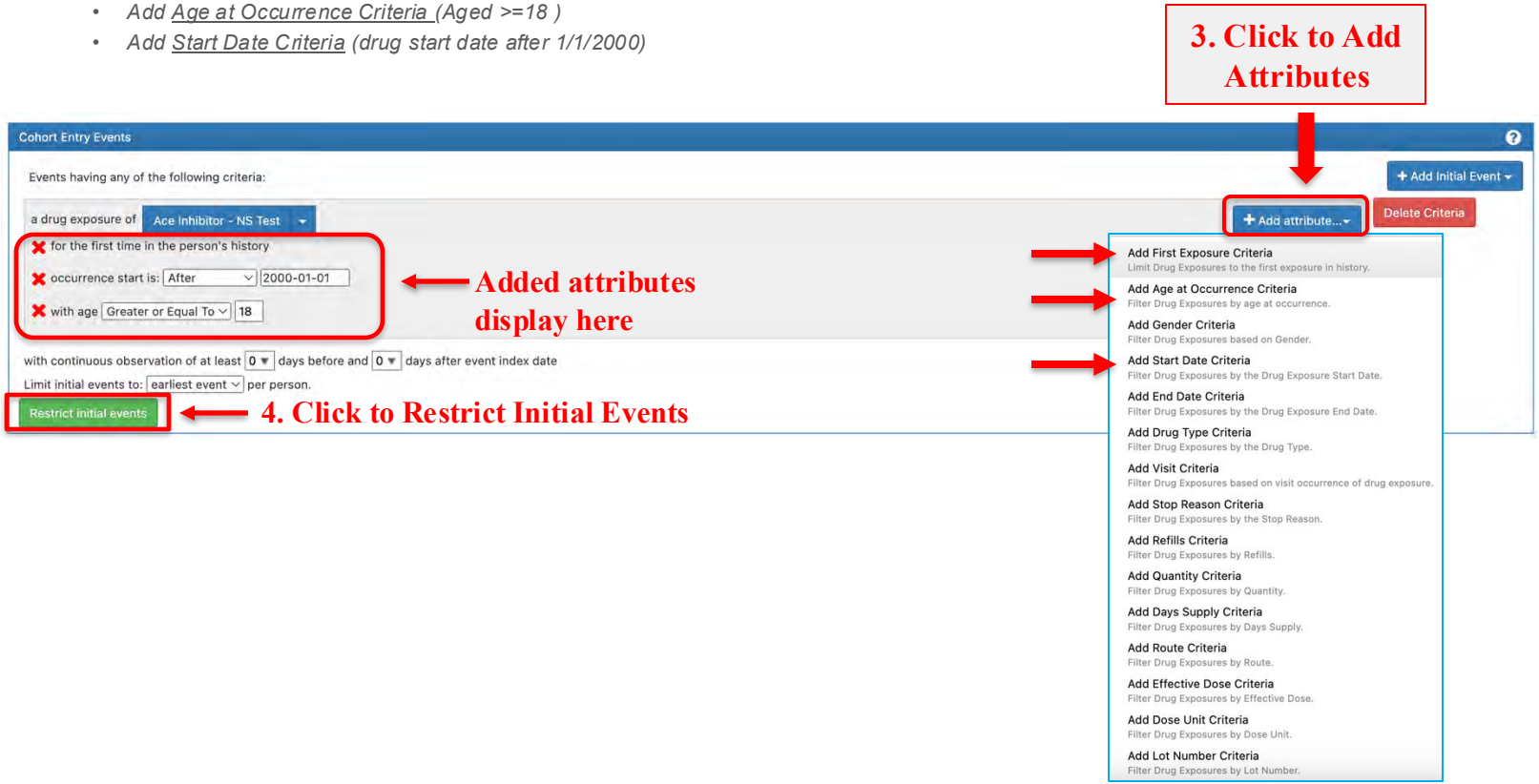
Cohort Entry Events - ***Example:*** *New Users of ACE Inhibitors*

1. Add Initial Event (ie. *add Drug Exposure*)
2. Import Concept Set (*i.e. ACE Inhibitor – NS Test)*

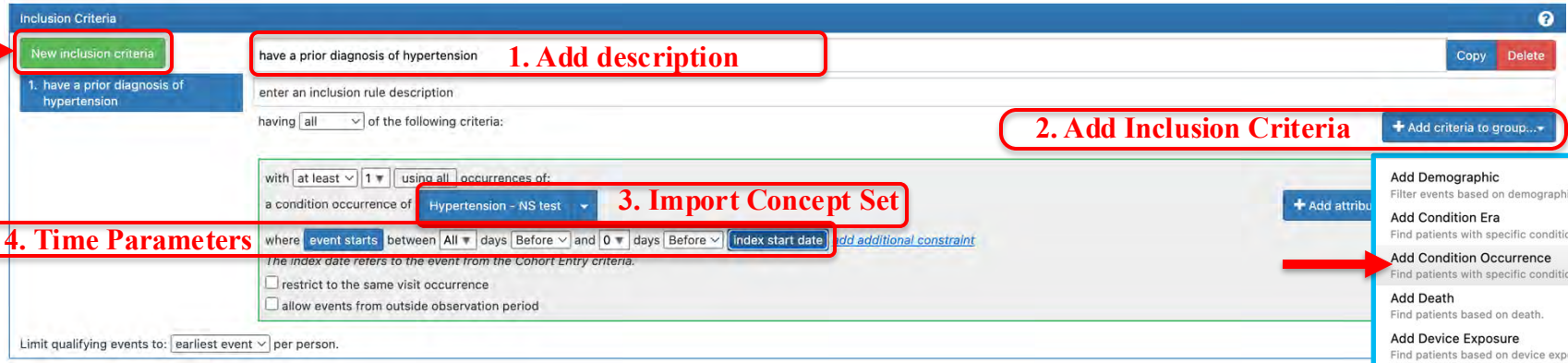


ATLAS – Cohort Definitions: Cohort Entry Event (cont.)

3. **Add Attributes** *(i.e. Add First Exposure Criteria)*
   - *Add First Exposure Criteria to define First time users*
   - *Add Age at Occurrence Criteria (Aged >=18)*
   - *Add Start Date Criteria (drug start date after 1/1/2000)*



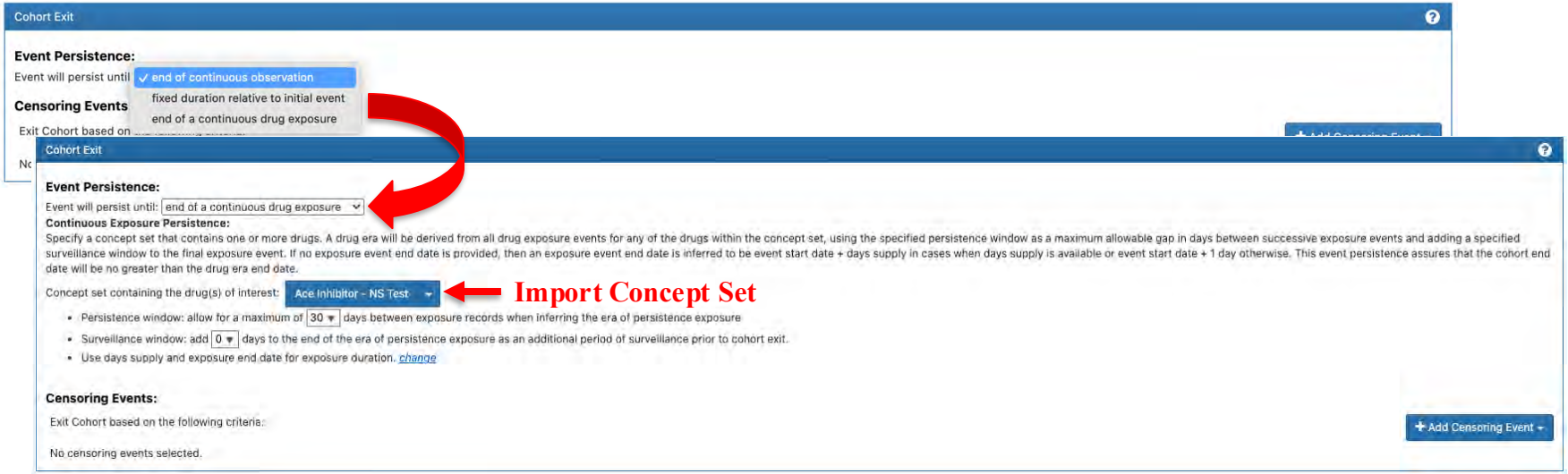
ATLAS – Cohort Definitions: New Inclusion Criteria



1. Add text description (*i.e. have a prior diagnosis of hypertension*)
2. Add criteria to group (*i.e. Add Condition Occurrence*)
3. Import Concept Set (*i.e. Hypertension – NS test*)
4. Define when the event (i.e. *Hypertension*) should occur relative to the index start date (aka Cohort entry event)

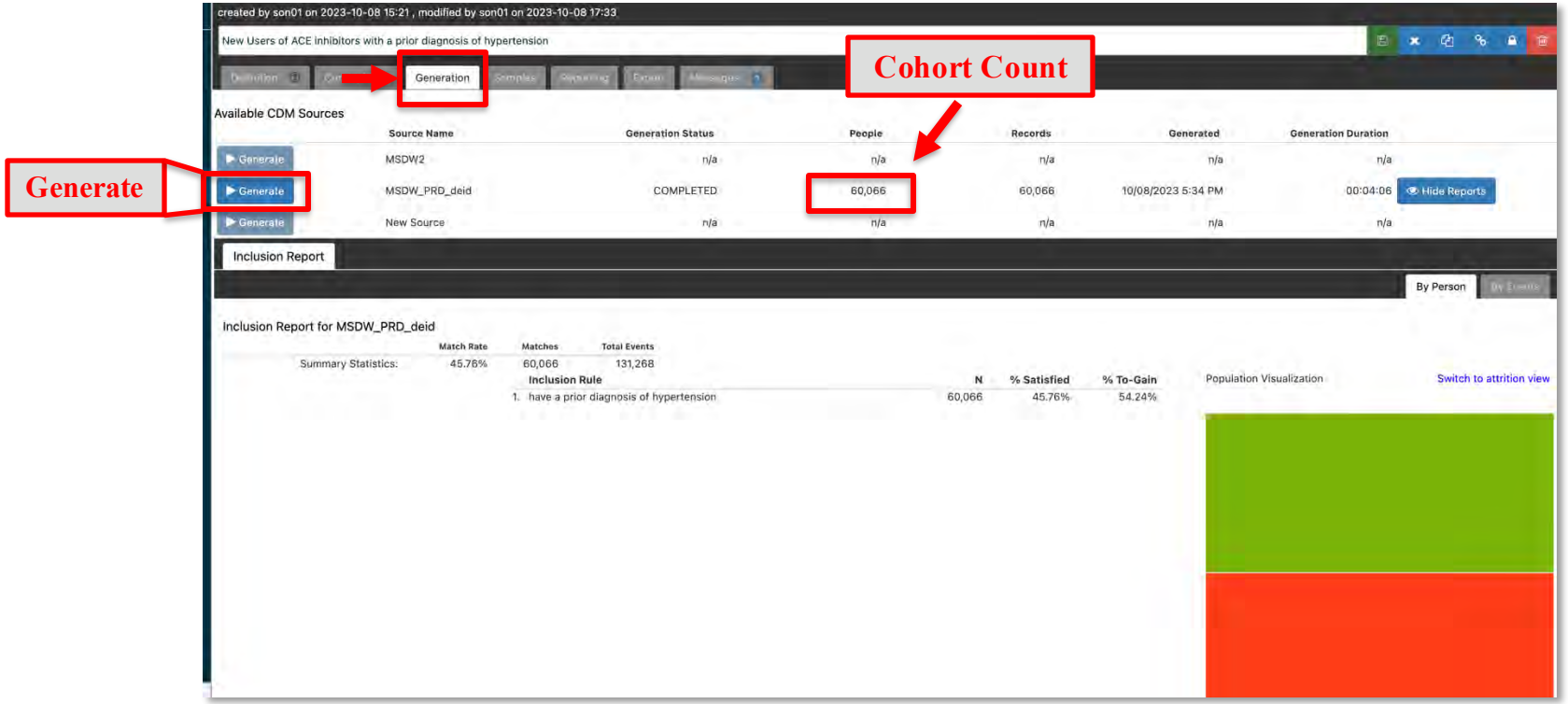
ATLAS – Cohort Definitions: Cohort Exit

- Define how a person leaves the cohort
  - select from the drop-down menu that the event will persist until a selected end
- Remember to SAVE cohort definition



ATLAS – Cohort Definitions: Generate Cohort

From **the Generation** tab, generate your cohort



ATLAS – Cohort Definitions: Additional Features

- **Concept Sets** – review concept sets within your definition
- **Export** – review a full description of what the Cohort Definition represents
- **Messages** – review warnings or memos regarding potential errors or incomplete aspects of the defined search criteria

Close Cohort Definition

Copy

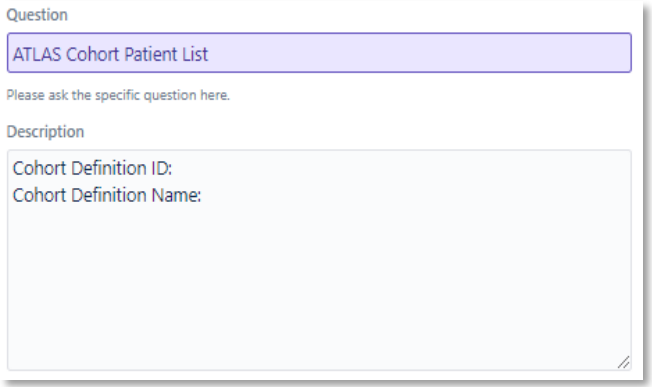
Create Link to Share Direct Access

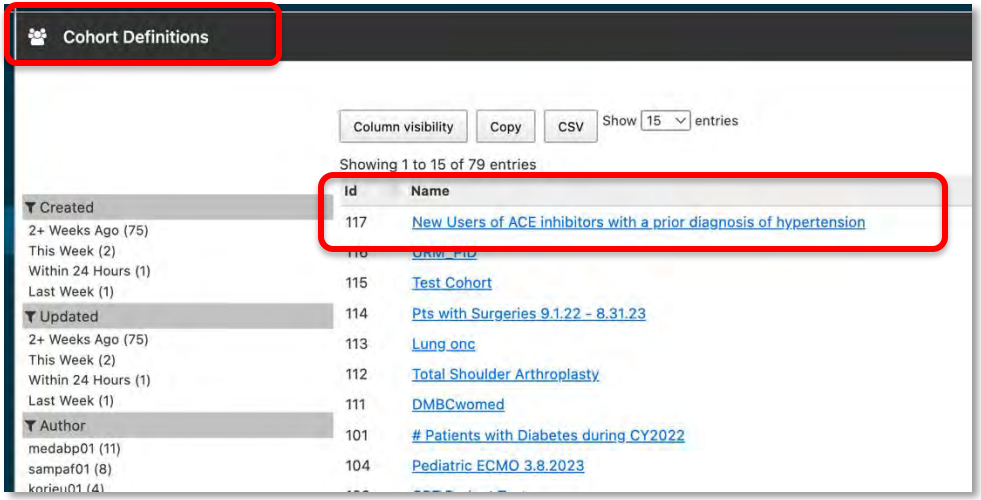
Configure Access

Delete

ATLAS – Patient List Extraction

- If you are interested in extracting the patient list, you can put in a JIRA ticket with the following details:





https://scicomp.mssm.edu/jira/servicedesk/customer/portal/4/create/100

MSDW Custom Data Request

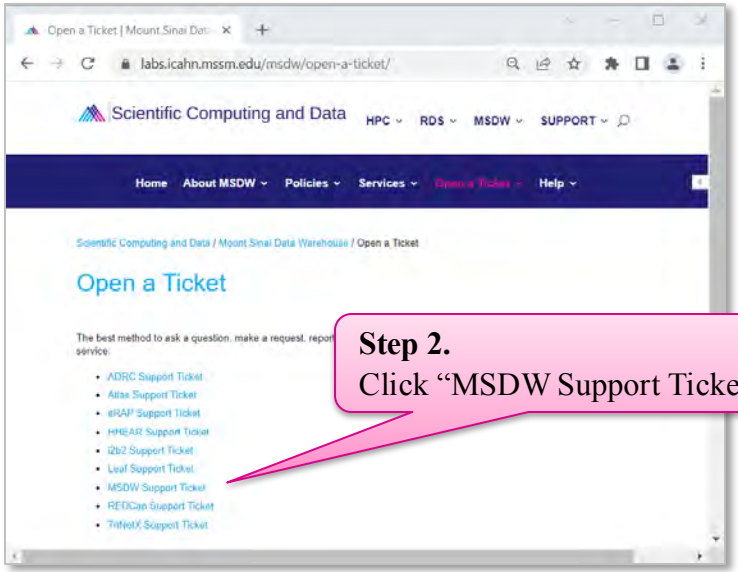
When You Need Custom Data

- Complex question that cannot be answered with one of the self-service query tools
- Need additional data that is not included in a de-identified data set
- Need PHI data for your analysis

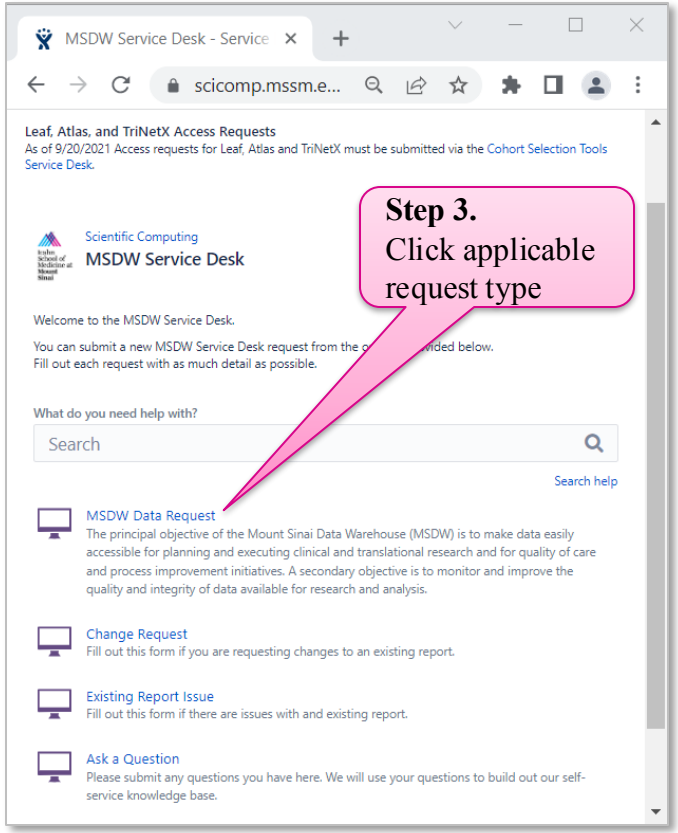
https://scicomp.mssm.edu/jira/servicedesk/customer/portal/4

How to Open an MSDW Request Ticket

Step 1.
Click “Open a Ticket”

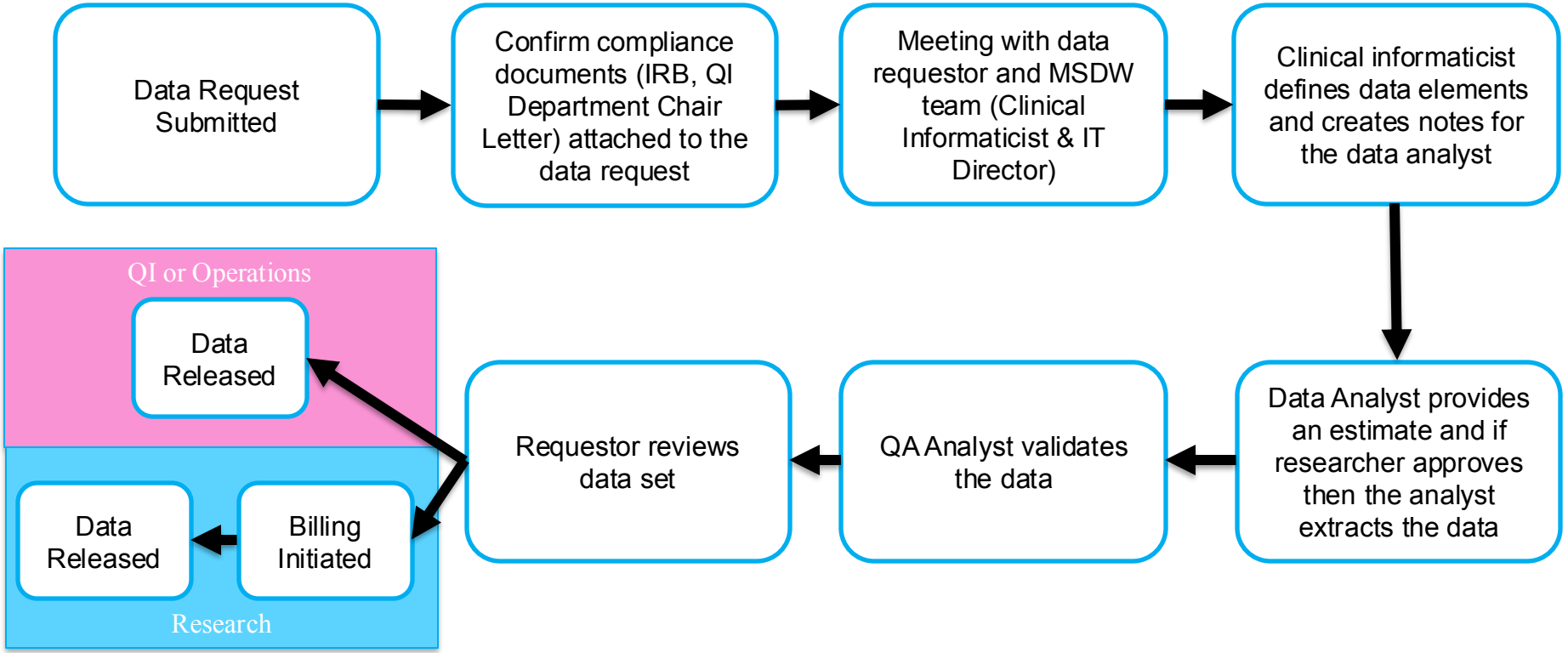


Step 2.
Click “MSDW Support Ticket”



Step 3.
Click applicable request type

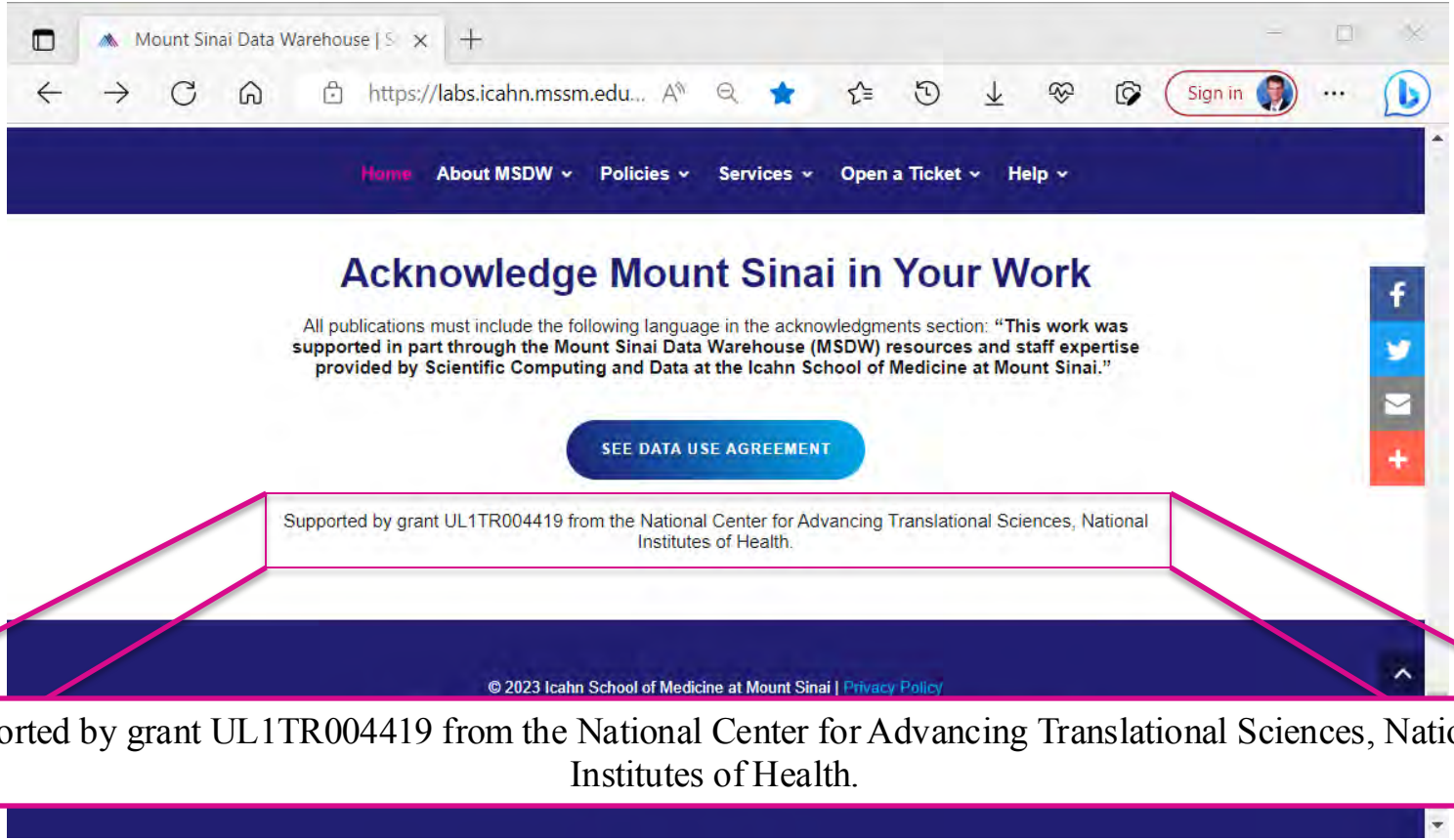
Workflow Once Data Request Submitted



JIRA ticketing system used to monitor the status of data requests

Acknowledgements

Encourage MSDW Users to Acknowledge CTSA



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

Your Publications

Report publications to Scientific Computing and Data:

All publications that resulted from Scientific Computing and Data resources and services, including Leaf and ATLAS, should be reported annually.

To report your publications, submit them here:
https://redcap.mountsinai.org/redcap/surveys/?s=HPEMDCYLNTXF3E3E

For 20 or more publications, email Maria at marajulia.castro@mssm.edu

Learn more about MSDW and Clinical Query tools from the links below:

https://labs.icahn.mssm.edu/msdw/

https://labs.icahn.mssm.edu/msdw/services/

https://labs.icahn.mssm.edu/msdw/data-sources/

"Walk-in" Digital Concierge service hosted by the MSDW

- Every Wednesday from 3:30 PM to 4:30 PM

Icahn
School of
Medicine at
Mount
Sinai

Thank you!

Thank you for your time! We hope you enjoyed this presentation.

Please take a minute to complete a short survey to provide your feedback and help improve our services:

https://redcap.link/bjpia7mf