Leaf and Atlas Clinical Query Tools

Patricia Kovatch, Dean

Sharon Nirenberg, MD

AJ Caberto

Database Analyst II

Scientific Computing and Data

Icahn School of Medicine at Mount Sinai

February 23, 2023

Icahn
School of
Medicine at
Mount
Sinai

Objectives:

1. Scientific Computing FAIR Principles of Data
2. Data Sources
3. Clinical query tools - Leaf and ATLAS
4. How to build a query using Leaf and ATLAS tools
5. When and how to request a custom dataset

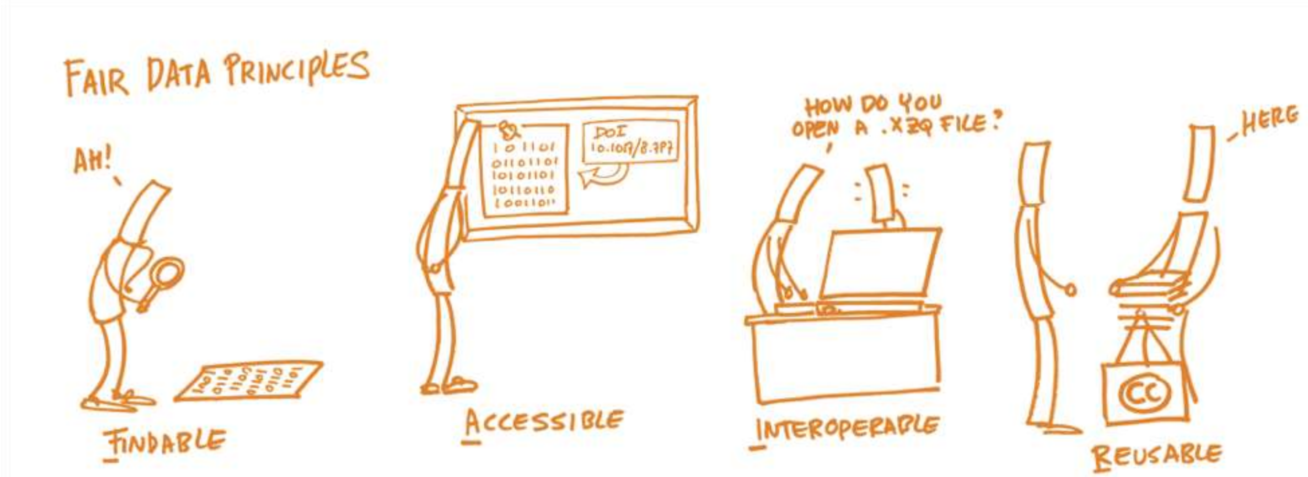
Scientific Computing FAIR Principles for Data

Findable

Accessible

Interoperable

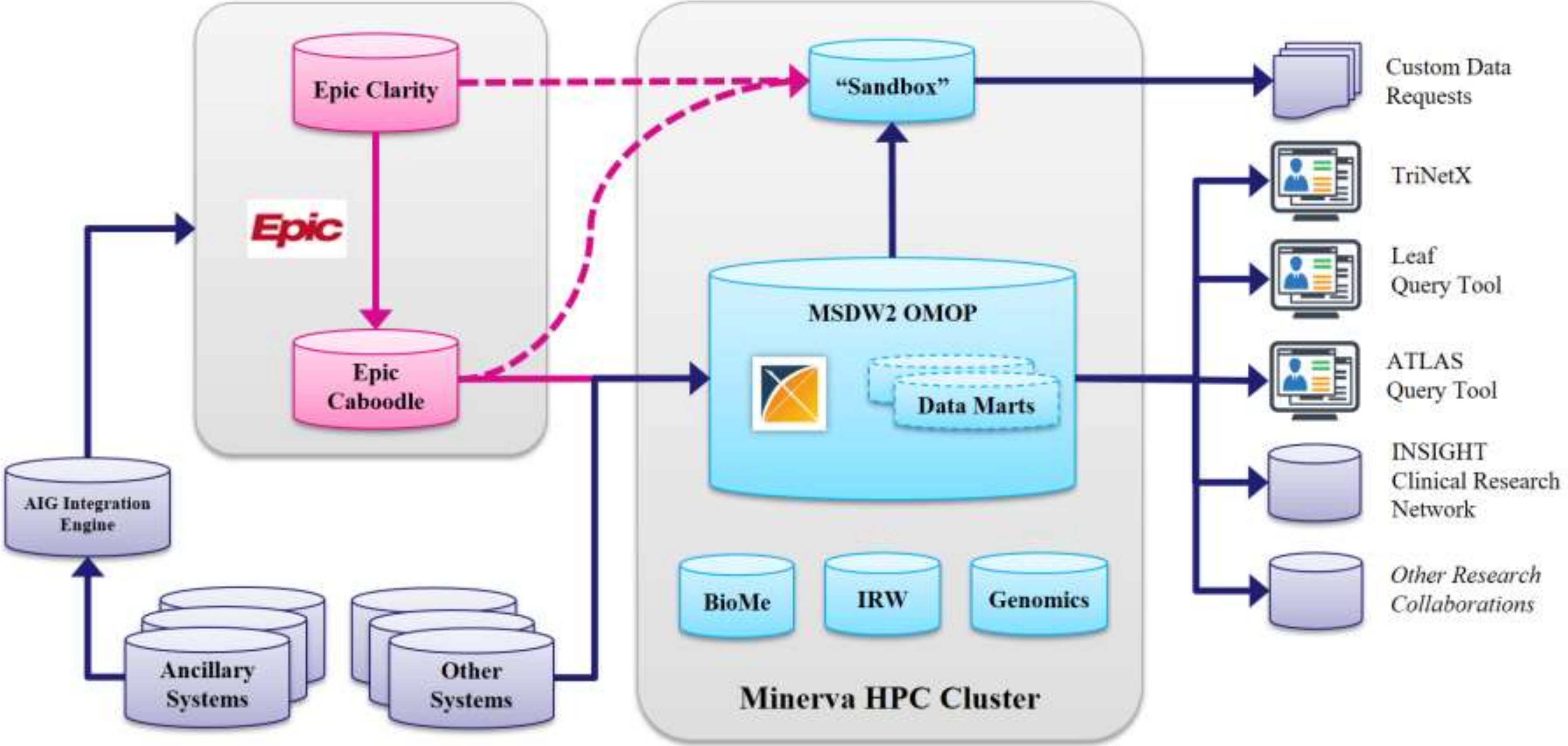
Reusable / Reproducible



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

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

MSDW2 Ecosystem

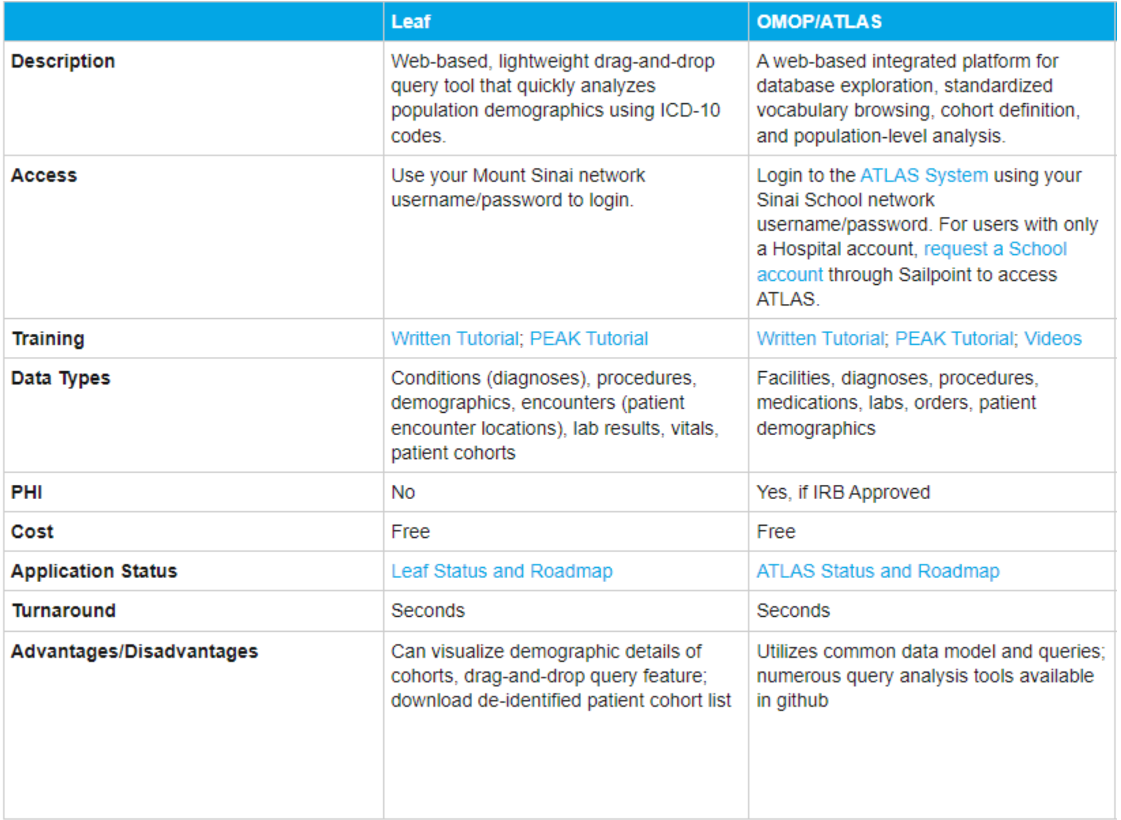


Leaf and Atlas Overview

	Leaf	ATLAS
Development	Nic Dobbins, Univ. of Washington, plus collaborators, including at ISMMS	OHDSI community: www.ohdsi.org
License	Free and Open-Source Software (FOSS)	
Trade-off	Easier, quicker, less powerful	Harder, laborious, more powerful
Data available	De-identified only	De-identified or PHI* (with IRB)
Capabilities	• Drag-and-drop Graphical User Interface • Simple Boolean logic • Temporal relationships • Predefined stats and visualizations • Can download lists of patients (privacy-preserving IDs)	• Sophisticated logic, temporal dependencies • Customized stats and 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/

Clinical Query Tools – Leaf and Atlas

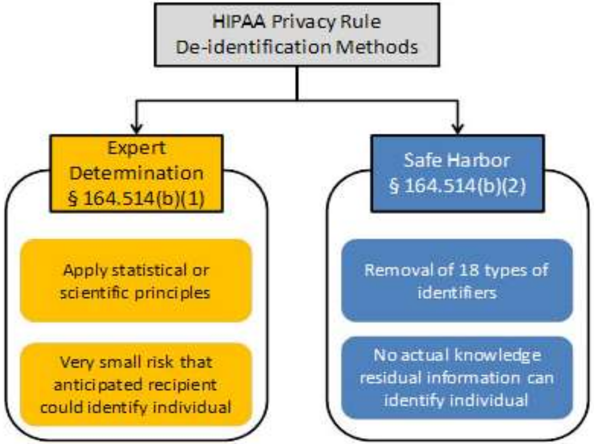


	Leaf	OMOP/ATLAS
Description	Web-based, lightweight drag-and-drop query tool that quickly analyzes population demographics using ICD-10 codes.	A web-based integrated platform for database exploration, standardized vocabulary browsing, cohort definition, and population-level analysis.
Access	Use your Mount Sinai network username/password to login.	Login to the ATLAS System using your Sinai School network username/password. For users with only a Hospital account, request a School account through Sailpoint to access ATLAS.
Training	Written Tutorial; PEAK Tutorial	Written Tutorial; PEAK Tutorial; Videos
Data Types	Conditions (diagnoses), procedures, demographics, encounters (patient encounter locations), lab results, vitals, patient cohorts	Facilities, diagnoses, procedures, medications, labs, orders, patient demographics
PHI	No	Yes, if IRB Approved
Cost	Free	Free
Application Status	Leaf Status and Roadmap	ATLAS Status and Roadmap
Turnaround	Seconds	Seconds
Advantages/Disadvantages	Can visualize demographic details of cohorts, drag-and-drop query feature; download de-identified patient cohort list	Utilizes common data model and queries; numerous query analysis tools available in github

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:



HIPAA Privacy Rule
De-identification Methods

Expert Determination
§ 164.514(b)(1)

- Apply statistical or scientific principles
- Very small risk that anticipated recipient could identify individual

Safe Harbor
§ 164.514(b)(2)

- Removal of 18 types of identifiers
- No actual knowledge residual information can identify individual

De-Identified Data Elements

- 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 numbers
- Fax number
- E-mail addresses
- Social Security Number (SSN)
- Medical Record Number (MRN)
- Health plan beneficiary number
- Account numbers
- Certificate/License number
- Vehicle identifiers, including license plate numbers
- Device identification and/or serial number
- Uniform Resource Locator (URL)
- Internet Protocol (IP) address
- Biometric identifiers, including finger and voiceprints
- Full face photographic images and other comparable images
- Any other unique identifying number, characteristic, or code

Leaf Access

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

- Indicate your reason for using Leaf: **Quality Improvement** or **Research**
- Do you have an approved IRB?
  - If Yes, consider Utilizing ATLAS
- Select **De-Identified** for the PHI question
- The next page will be the **Program for Protection of Human Subjects Access** disclosure statement. You will need to read and select ‘I agree’ to consent to the terms of use

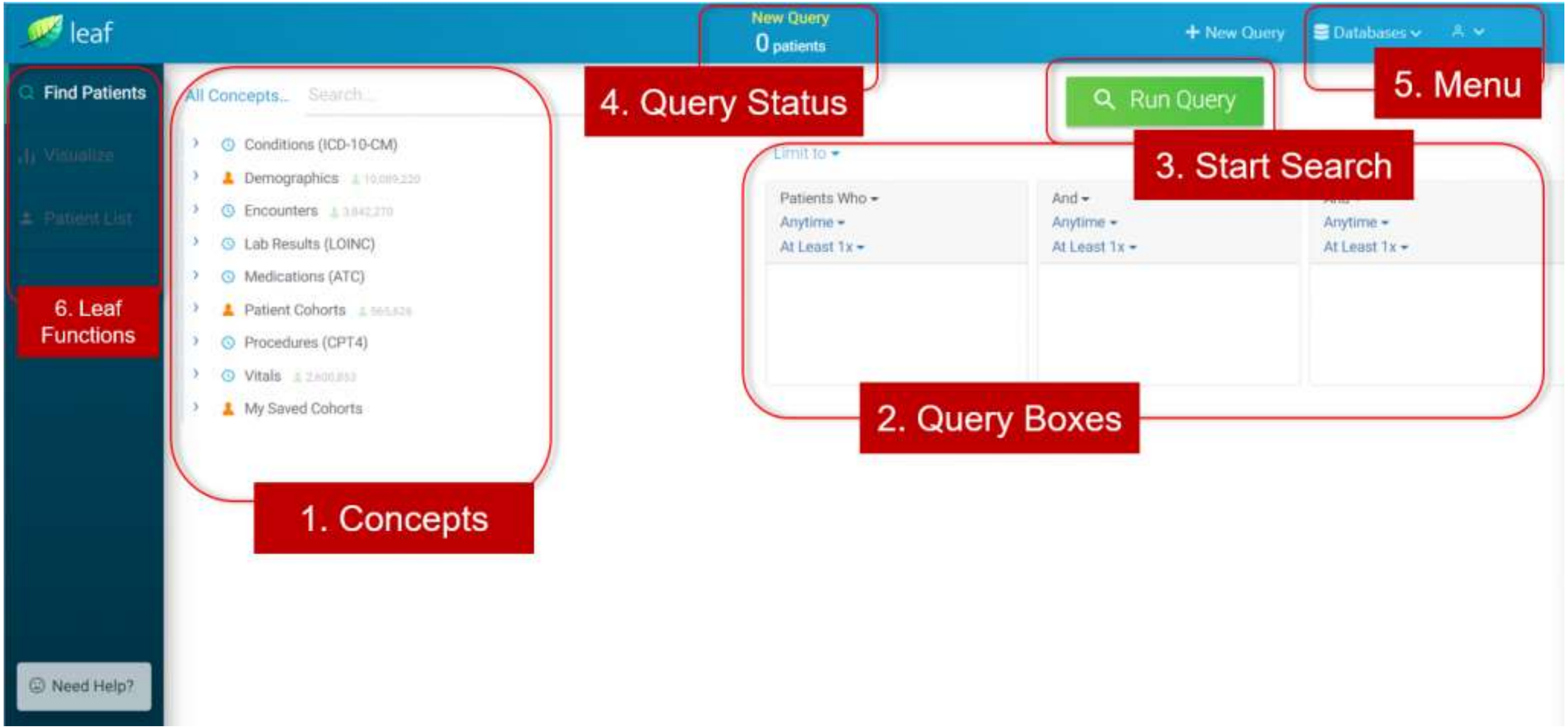
I want to find information for: Quality Improvement [✓ Research]

I have an Approved IRB: [✓ No] Yes

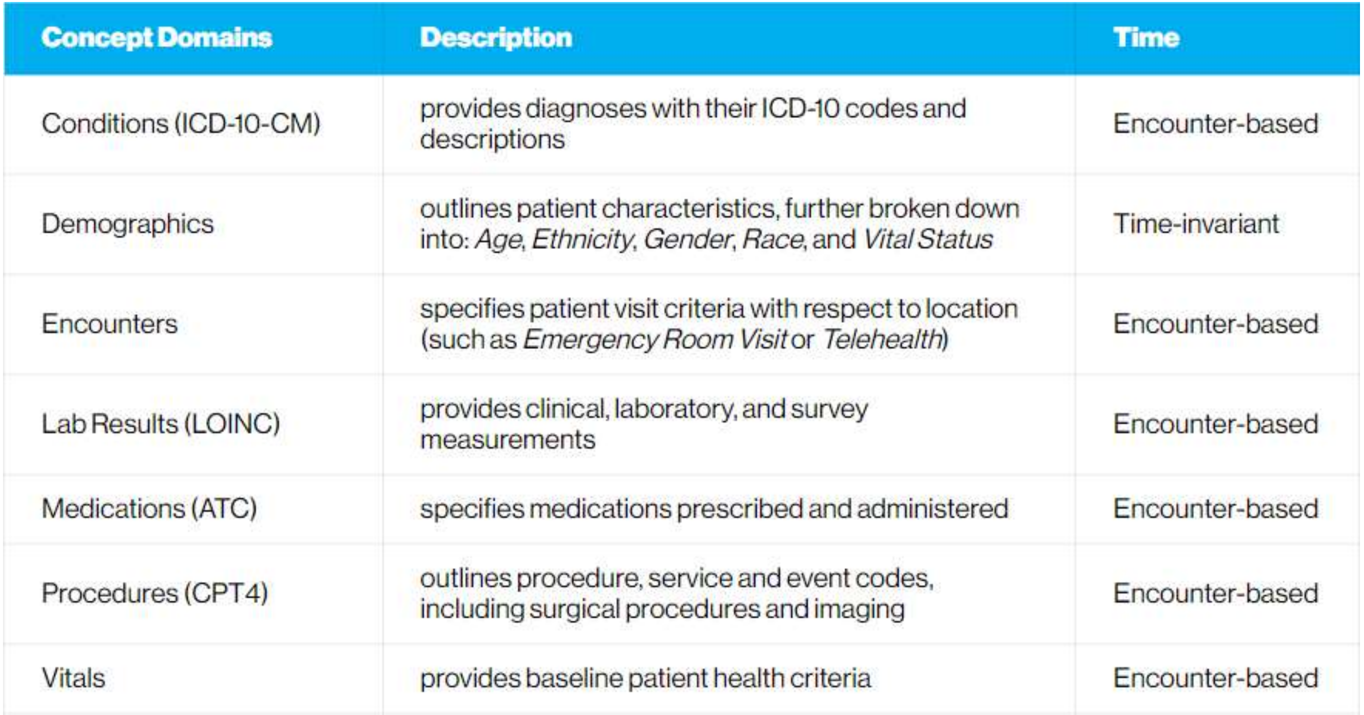
I would like Protected Health Information: [✓ De-Identified] Identified

Figure 3: Selection path to access De-Identified Patient Data for Research without an IRB

Leaf Interface



Leaf - Concept Domains



Concept Domains	Description	Time
Conditions (ICD-10-CM)	provides diagnoses with their ICD-10 codes and descriptions	Encounter-based
Demographics	outlines patient characteristics, further broken down into: *Age*, *Ethnicity*, *Gender*, *Race*, and *Vital Status*	Time-invariant
Encounters	specifies patient visit criteria with respect to location (such as *Emergency Room Visit* or *Telehealth*)	Encounter-based
Lab Results (LOINC)	provides clinical, laboratory, and survey measurements	Encounter-based
Medications (ATC)	specifies medications prescribed and administered	Encounter-based
Procedures (CPT4)	outlines procedure, service and event codes, including surgical procedures and imaging	Encounter-based
Vitals	provides baseline patient health criteria	Encounter-based

LOINC - Logical Observation Identifiers Names and Codes
ICD-10 - International Classification of Diseases
ATC - In the Anatomical Therapeutic Chemical (ATC) classification system, the active substances are divided into different groups according to the organ or system on which they act and their therapeutic, pharmacological and chemical properties.
CPT – Current Procedural Terminology

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)

Leaf Query Demo

1. Female Patients between age 50-74 and screened for mammogram
   - Add a Date Range
   - Show Visualize and Patient List
   - Save Query
2. **Quick Review:** Adult patients that received a CT Scan and diagnosed with the following codes – S02.0XXX, S02.1XXX
3. **Using Patient Cohorts:** Patients with Cancer treated with Antineoplastic Agents in the last month

Leaf Application Status and Roadmap

Last Updated: December 27, 2022

Status

Quality improvement and research users can use Leaf to conveniently query the Mount Sinai Data Warehouse. Leaf only accesses de-identified data. Documentation and user service are available.

Leaf supports these query domains:

- Conditions (diagnoses) using ICD-10-CM
- Visit (encounter) locations
- Lab results using LOINC
- Medications using ATC
- Procedures using CPT4
- Demographics, such as current age, ethnicity, gender, race and vital status
- Vitals
- Patient cohorts in Sinai repositories, starting with BioMe, BioBank, the Imaging Research Warehouse 1.0, the Cancer Institute Biorepository, and Cancer Patients

Leaf contains these known bugs:

- Patient List temporarily disabled
- Queries that last longer than 1 minute will terminate

https://labs.icahn.mssm.edu/msdw/leaf-application-status/

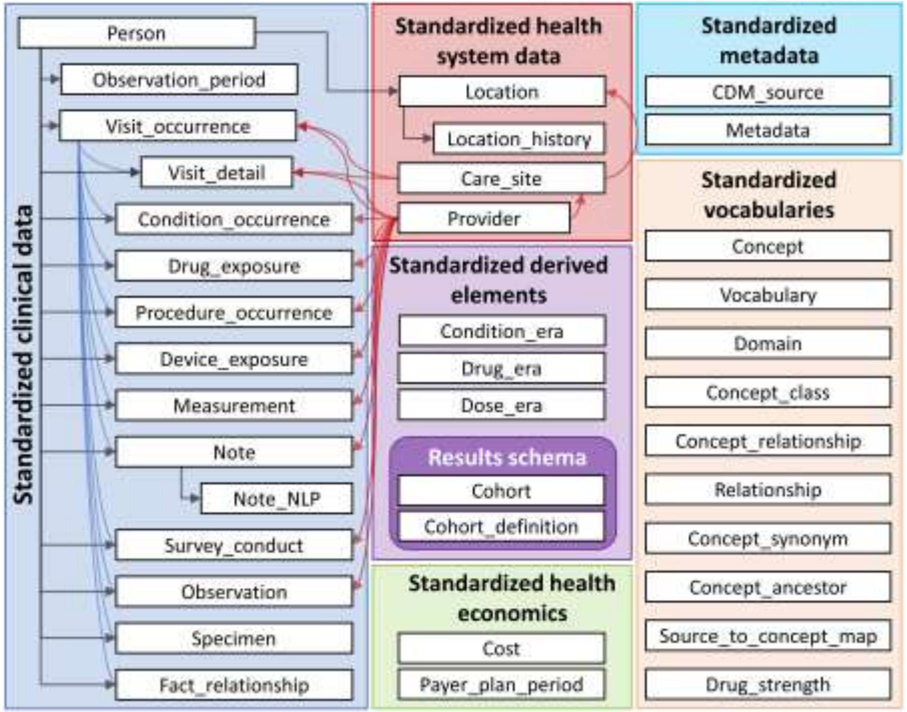
Roadmap

These improvements to Leaf are ongoing for the first half of 2022:

- Fix high-priority known bugs
- Enhance the Leaf development and configuration process to make it easier to support Leaf's users
- In addition, Leaf will access all improvements that are made to MSDW2.

OMOP Common Data Model

- **Observational Medical Outcomes Partnership (OMOP)**
  - Clinical Data
  - Health System Data
  - Standardized derived elements
  - Standardized Metadata
  - Standardized Vocabularies
  - Standardized Health Economics data
- **Common Data Model**
  - Standardizes both structure & content for interoperability
  - Produce meaningfully comparable and reproducible results



Standardized clinical data

- Person
- Observation_period
- Visit_occurrence
- Visit_detail
- Condition_occurrence
- Drug_exposure
- Procedure_occurrence
- Device_exposure
- Measurement
- Note
- Note_NLP
- Survey_conduct
- Observation
- Specimen
- Fact_relationship

Standardized health system data

- Location
- Location_history
- Care_site
- Provider

Standardized derived elements

- Condition_era
- Drug_era
- Dose_era

Results schema

- Cohort
- Cohort_definition

Standardized health economics

- Cost
- Payer_plan_period

Standardized metadata

- CDM_source
- Metadata

Standardized vocabularies

- Concept
- Vocabulary
- Domain
- Concept_class
- Concept_relationship
- Relationship
- Concept_synonym
- Concept_ancestor
- Source_to_concept_map
- Drug_strength

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

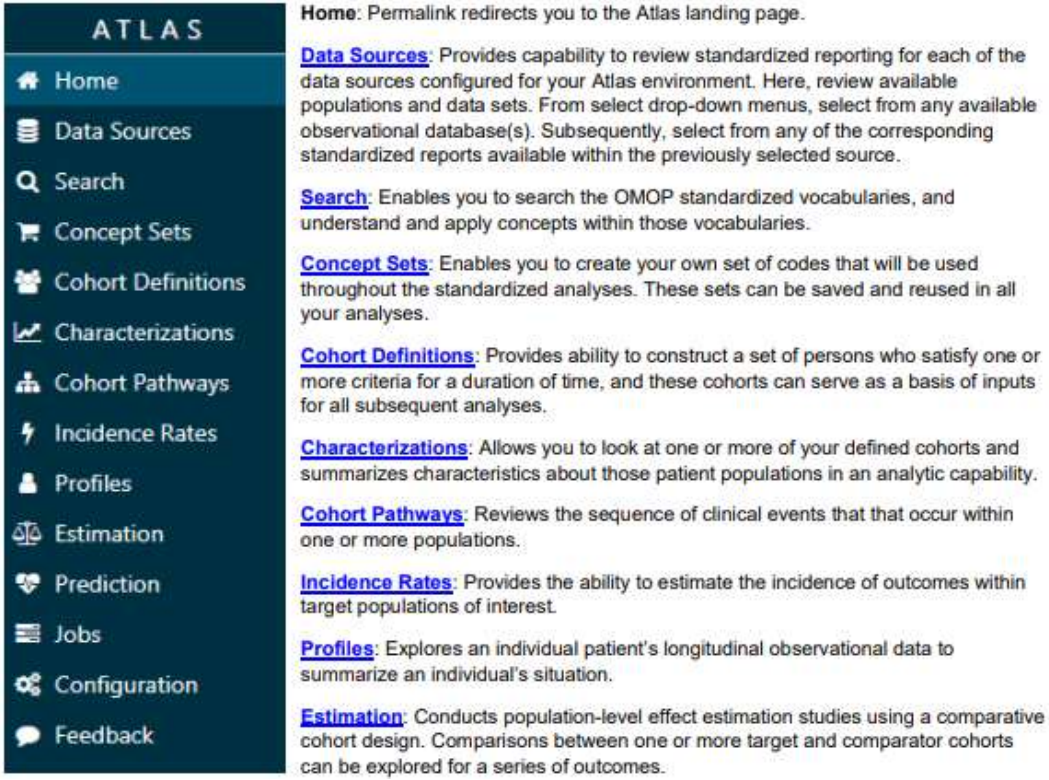
ATLAS

- 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.

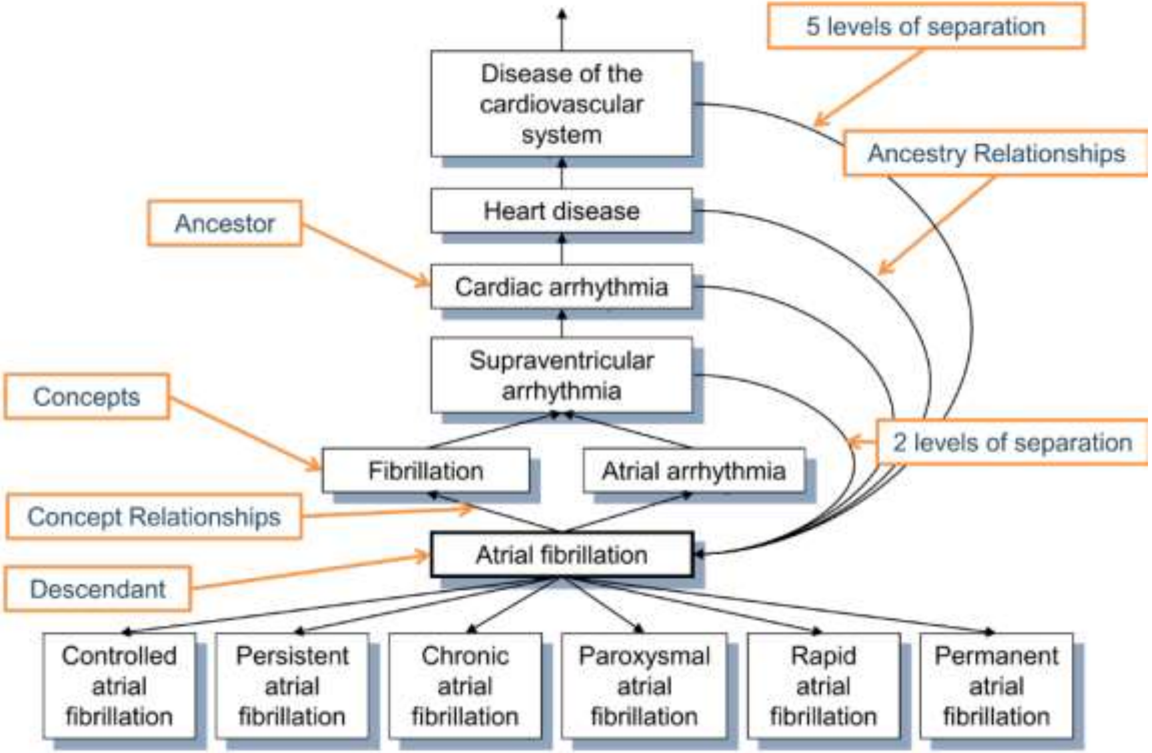
Profiles: Explores an individual patient's longitudinal observational data to summarize an individual's situation.

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 Feature

- Enables you to search the OMOP standardized vocabularies to understand and apply concepts within those vocabularies
- Enter terms, codes and vocabularies to find related concepts
- Selecting on a concepts provides additional details
  - **Details:** Descriptive values connected to the record and provides estimated record counts
  - **Related Concepts:** Provides other vocabulary for similar terms that may specify or broaden the search
  - **Hierarchies:** Indicates parent and children of the concept within the OMOP vocabulary

Disease Hierarchy (SNOMED)



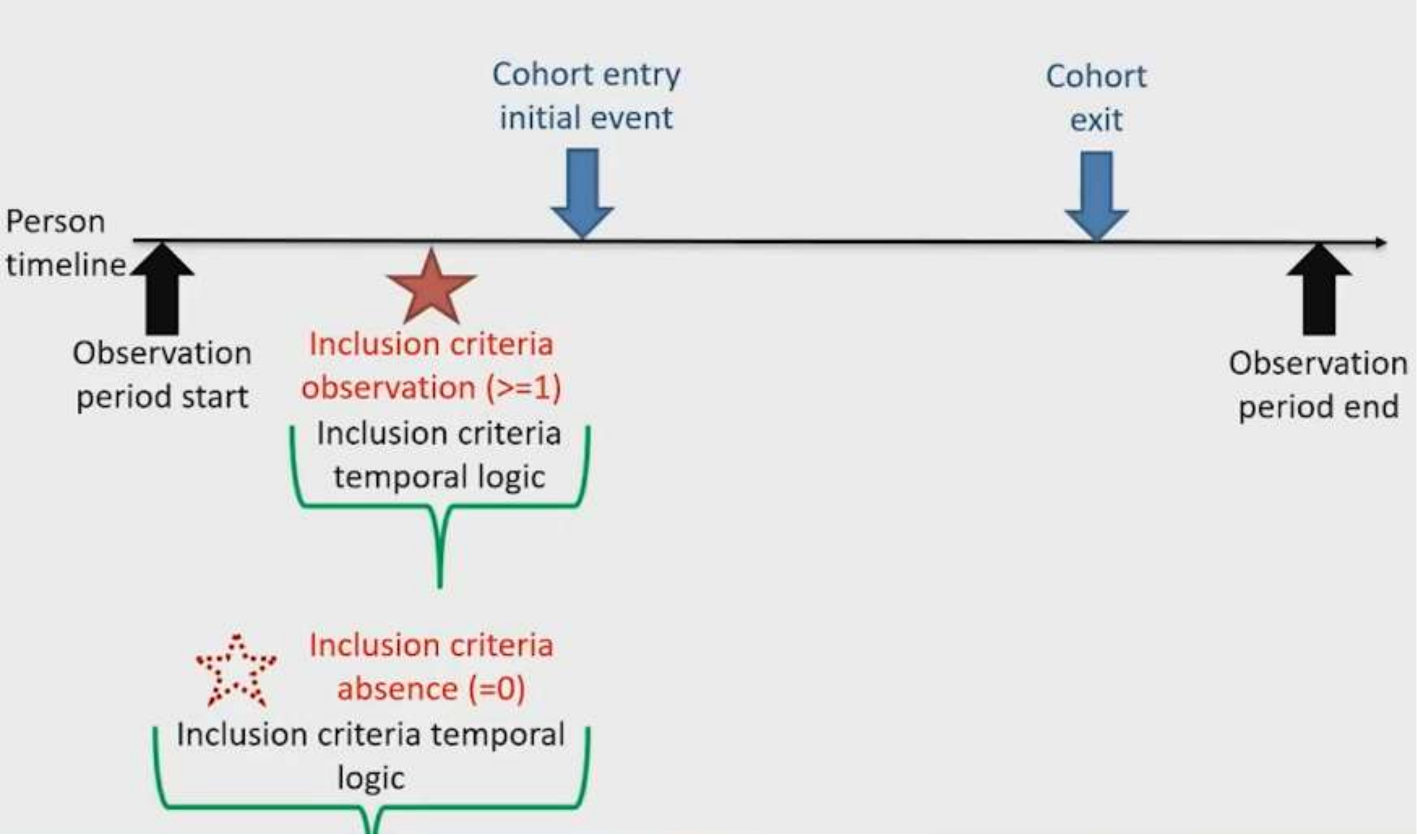
ATLAS – Concept Sets

- The building blocks of ATLAS queries
- Enables you to create your own set of concepts that will be used through the standardized analyses
- These can be saved, reused and shared for future queries and analysis
- Once concepts are added and save, the following items are essential to managing your concept set
  - **Exclude:** To exclude concepts from the cohort definition without deleting it from the set
  - **Descendants:** To add children of concepts to the set
  - **Mapped:** To add related concepts to the set
  - **Included Concepts Tab:** Includes additional concepts gathered through inclusion of descendants and mapped concepts

ATLAS – Cohort Definitions

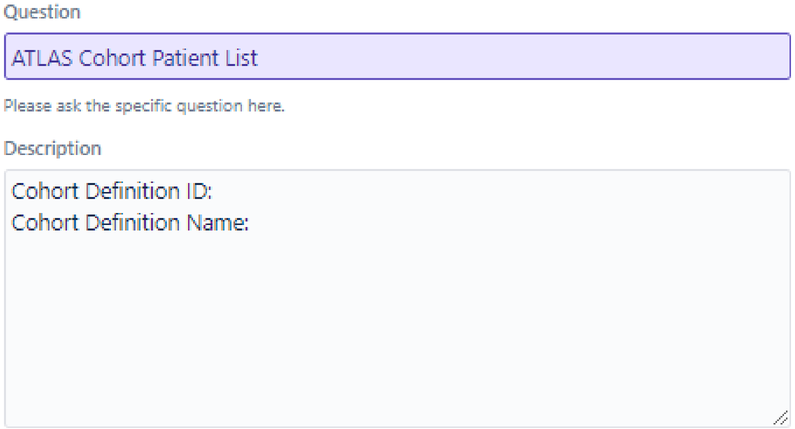
- A cohort definition defines when an individual qualifies (or does not qualify) for a cohort during this span of time
- Once saved, these cohorts can serve as a basis of inputs for all subsequent analyses
- When building cohort, it is important to visit the following sections
    - **Cohort Entry Event:** Occurs in a moment in time to qualify an individual for cohort entry
    - **Inclusion Criteria:** Use concept sets to build specific criteria to include in your cohort
    - **Cohort Exit:** Event in time at which an individual does not qualify anymore for a cohort and leaves it
- Once the definition is saved, visit the Messages tab to troubleshoot your definition
- Generation tab will allow you to see the number of patients in your population

Defining a Cohort



ATLAS – Patient List Extraction

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



ATLAS Cohort Patient List

Cohort Definition ID:
Cohort Definition Name:

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

ATLAS Application Status and Roadmap

Last Updated: December 27, 2022

Status

ATLAS application is currently in software version 1 at Mount Sinai and is available for user testing. Features and functionality are regularly updated, and scheduled maintenance is announced in advance through msdw.mountsinai.org.

As of December 2021, all ATLAS functions (left-hand tabs) are available to users.

ATLAS functionality is currently limited to include:

- De-identified data source only
- OMOP standard concept IDs are contained within only the following clinical domains: conditions, procedures, visits, vitals measurements
- All other domains can be queried using concepts in Epic vocabularies

ATLAS beta contains the following limitations:

- Data containing PHI, including datamarts, are not yet available for query
- Most ATLAS data uses OMOP standard concept IDs, but some live data uses non-standard vocabularies

https://labs.icahn.mssm.edu/msdw/atlas-application-status/

Roadmap

The following improvements to ATLAS are planned for the 4th quarter of 2021:

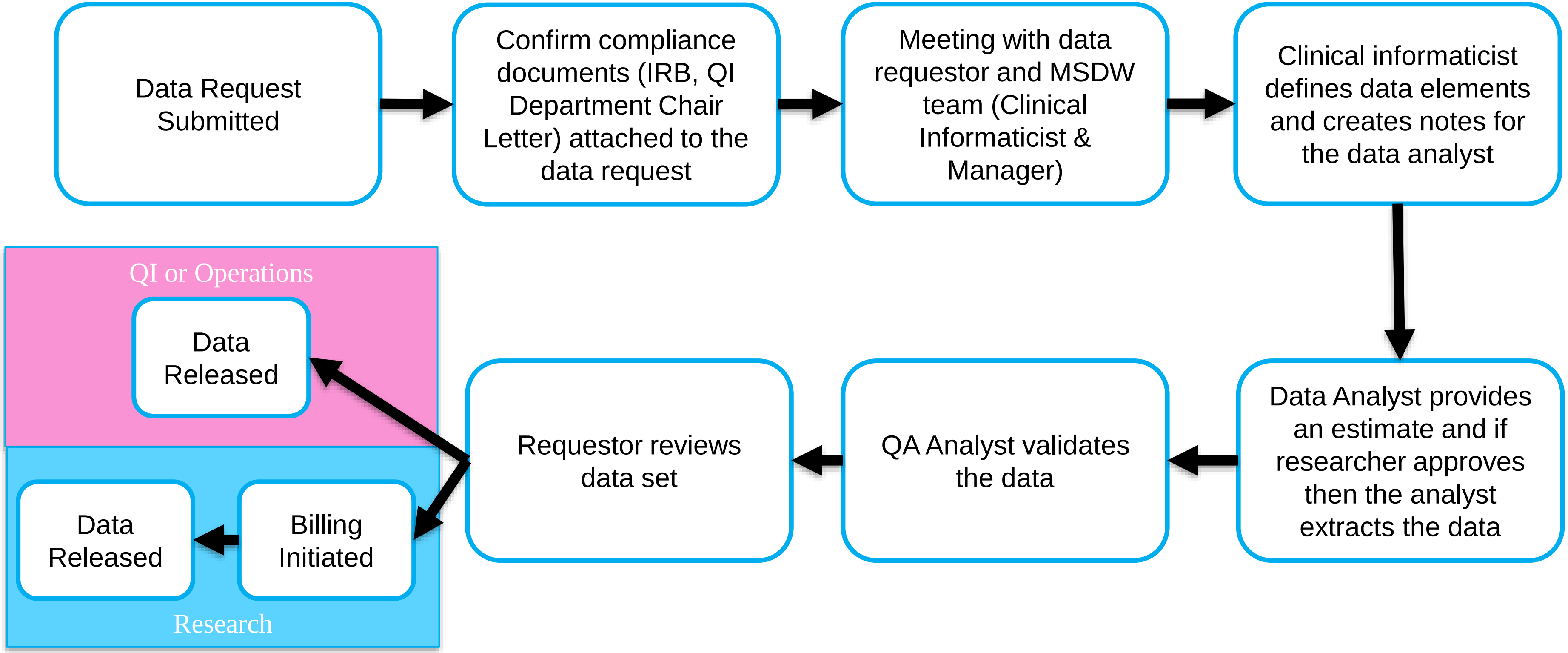
- Add four datamarts containing PHI: BioMe, Anal Cancer, Liver, and iSite
- Enhance ATLAS deployment process to support ATLAS users

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

Workflow Once Data Request Submitted



Data Request Submitted

Confirm compliance documents (IRB, QI Department Chair Letter) attached to the data request

Meeting with data requestor and MSDW team (Clinical Informaticist & Manager)

Clinical informaticist defines data elements and creates notes for the data analyst

Data Analyst provides an estimate and if researcher approves then the analyst extracts the data

QA Analyst validates the data

Requestor reviews data set

QI or Operations

Data Released

Research

Billing Initiated

Data Released

JIRA ticketing system used to monitor the status of data requests

Acknowledgements

CTSA Clinical & Translational Science Awards ®

Acknowledge CTSA:

An acknowledgement of support from the Icahn School of Medicine at Mount Sinai and the Clinical and Translational Science Awards (CTSA) grant UL1TR004419 from the National Center for Advancing Translational Sciences should appear in a publication of any material, whether copyrighted or not, based on or developed with Mount Sinai-supported computing resources:

Please use the following acknowledgement in your publications:

“This work was supported in part through the computational and data resources and staff expertise provided by Scientific Computing and Data at the Icahn School of Medicine at Mount Sinai and supported by the Clinical and Translational Science Awards (CTSA) grant ULTR004419 from the National Center for Advancing Translational Sciences.”

Acknowledge Scientific Computing and Data:

All publications must include the following language in the acknowledgments section:

“This work was supported in part through the Mount Sinai Data Warehouse (MSDW) resources and staff expertise provided by Scientific Computing and Data at the Icahn School of Medicine at Mount Sinai.”

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 mariajulia.castro@mssm.edu

Learn more about MSDW2 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/

Digital Concierge – Every Wednesday, 3:30pm – 4:30pm

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/cykp9ale