

Frequently Asked Questions (FAQ)

Frequently Asked Questions (FAQ)	1
Link to videos	2
Access to data	3
CHoRUS data	4
Common links	4
What is the structure of the CHoRUS B2AI Data Enclave?	4
Where do I go to log in?	5
I've logged in, now what?	6
I made it to the ML Workspace, how do I get my hands dirty?	7
I've launched my own JupyterLab environment, where's the data?	9
Are there other ways to explore the data?	9
How do I train models created in Atlas in my compute instance?	Error! Bookmark not defined.
What's the deal with GitLab?	Error! Bookmark not defined.
Access to GitHub	11
I don't have a GitHub ID	11
Common Data Access Issues	11
I can't access the waveforms.	11
Common Access Issues	11
I changed institutions and need to change my institutional email to log in.	11
Trouble logging in after already accepting the invitation?	12
No accounts at your institution is able to log in?	12
Unanswered questions	13
Frequently used acronyms	14
D	14
E	14
I	14
S	14
Resources	15
For first-time data scientists	15
For first-time clinicians	15
For basic scientists	15

Link to videos

- [How to analyze PhysioNet EHR data with Jupyter/BigQuery](#)
- [Introduction to the CHoRUS Dataset](#)
- [OHDISI OMOP Standard Intro](#)

Access to data

The current data sets are:

- HITIdata60

All teams must have access to all data sets.

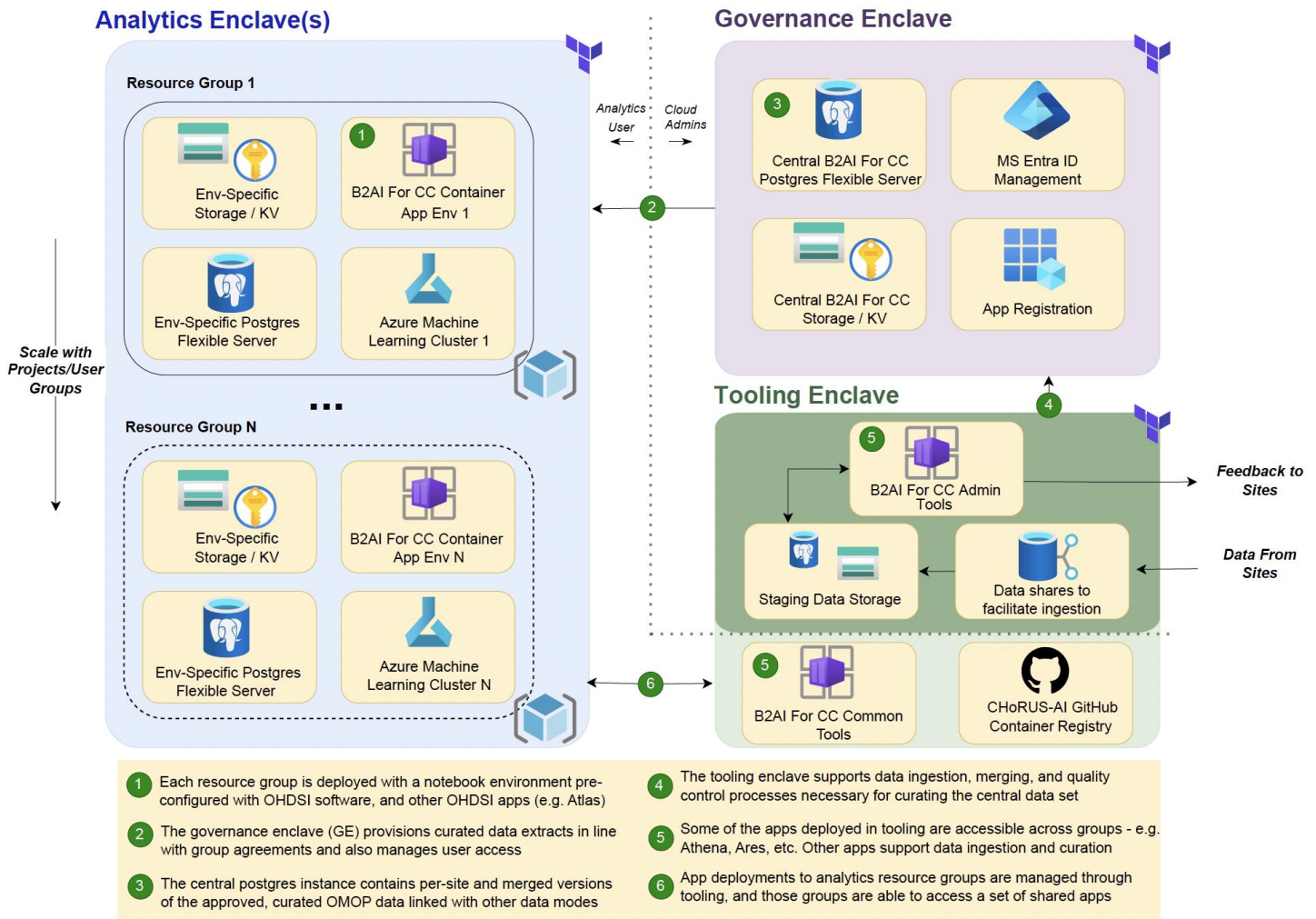
CHoRUS 2026 caveat: This dataset is in its Beta version. As the B2AI AI for Clinical Care grant is still ongoing. Data changes and updates will be ongoing through the process.

CHoRUS data

Common links

- Schema: <https://ohdsi.github.io/CommonDataModel/cdm54.html>
- portal: <https://ca-www-datathon.mangoocloud-7758f0fc.eastus2.azurecontainerapps.io/>

What is the structure of the CHoRUS B2AI Data Enclave?

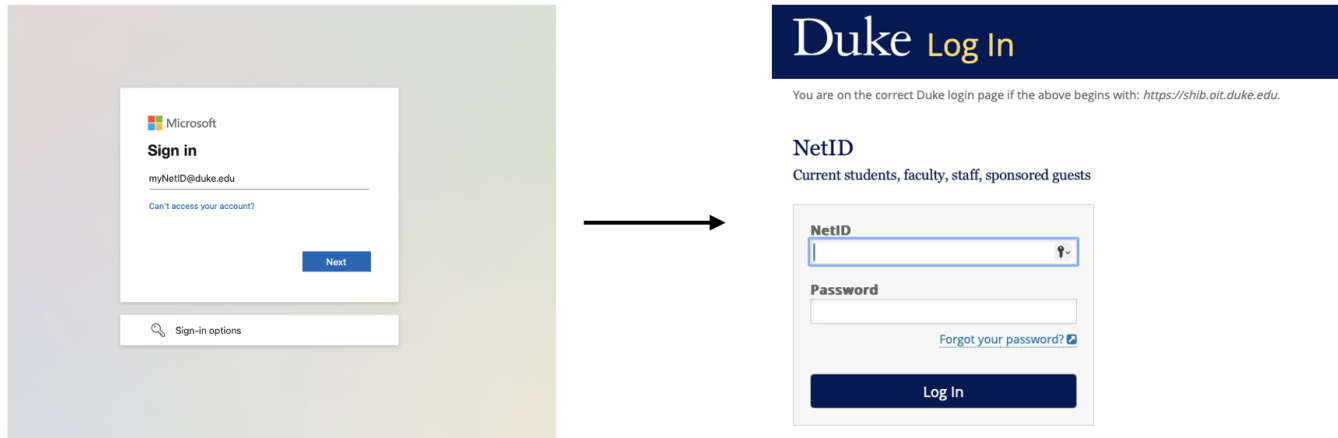


The Bridge2AI for Clinical Care (B2AI For CC, or CHoRUS) collaborative cloud consolidates rich multi-center and multimodal data in order to support complex analytic processes in machine learning (ML) and artificial intelligence (AI); such consolidation and analytic support is nontrivial and requires a flexible and scalable architecture design. We've devised a three-enclave approach (see above) that addresses the anticipated diversity of researcher profiles and analytic needs, and that ingests and combines large volumes of data from a heterogeneous network of fourteen healthcare centers.

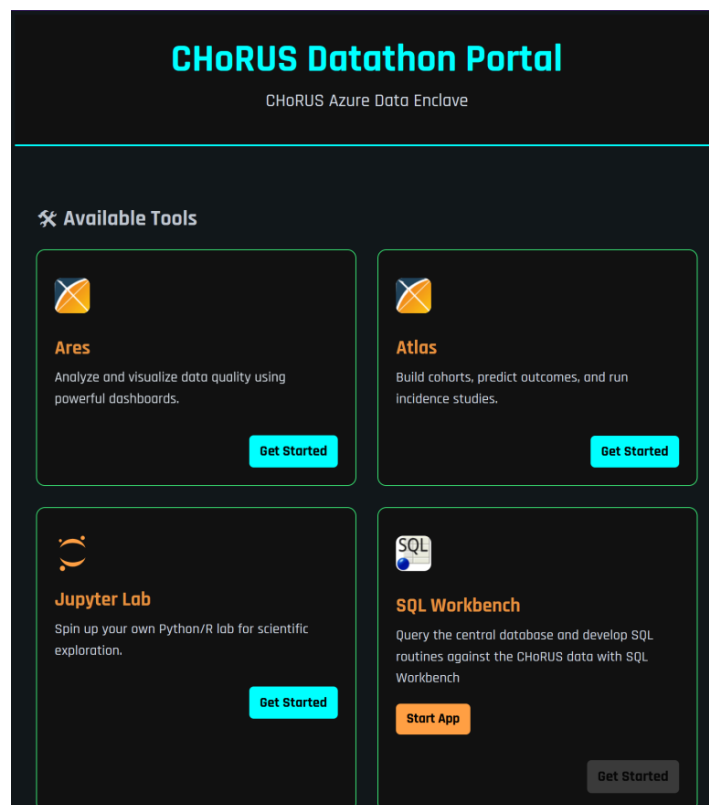
Where do I go to log in?

Please go here: <https://ca-www-datathon.mangoocloud-7758f0fc.eastus2.azurecontainerapps.io/>

You should arrive at a familiar Microsoft login screen. Here, you'll enter your **OWN INSTITUTIONAL** credentials that you gave in registration and you will be redirected to your institution's Shibboleth for a standard login.



Once successfully logged in, you'll arrive at the CHoRUS Datathon landing page:



I've logged in, now what?

You now see the 4 available applications that you can use within the environment.

- ARES
 - Analyze and visualize data quality using powerful dashboards
- ATLAS
 - Build cohorts, predict outcomes, and run incidence studies
- Jupyter Lab
 - Spin up your own Python/R lab for scientific exploration
- SQL Workbench
 - Query the central database and develop SQL routines against the CHoRUS data with SQL Workbench

The rest of this document will give a brief introduction to each of the applications.

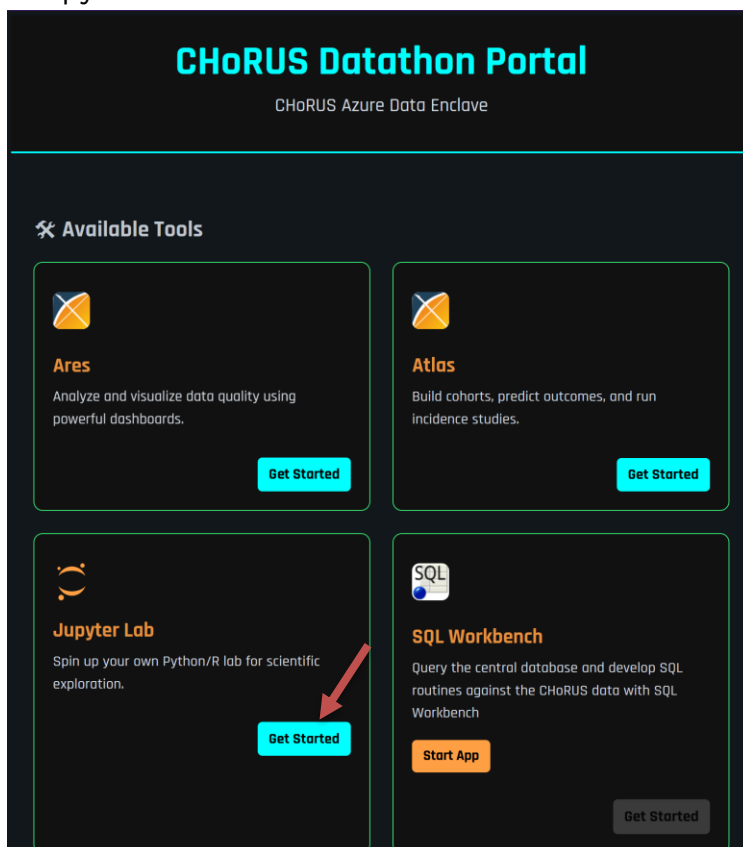
Users are assigned compute upon account creation. The CHoRUS platform uses a tiered approach for provisioning compute.

Tier	Compute Assigned	
Default	CPU: 4	Memory: 8Gi
Small	CPU: 4	Memory: 16Gi
Medium	CPU: 8	Memory: 32Gi
Large	CPU: 16	Memory: 64Gi

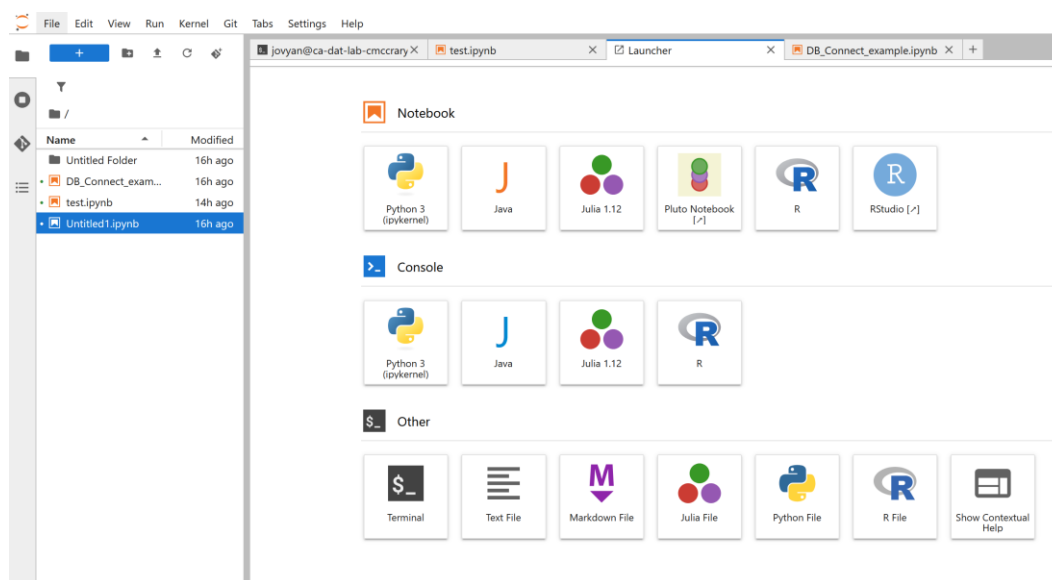
All users start at Default. If more compute power is needed, please talk to a CHoRUS representative and they can provision more power.

How do I get my hands dirty, JupyterLab?

Click 'Get Started' under Jupyter Lab.



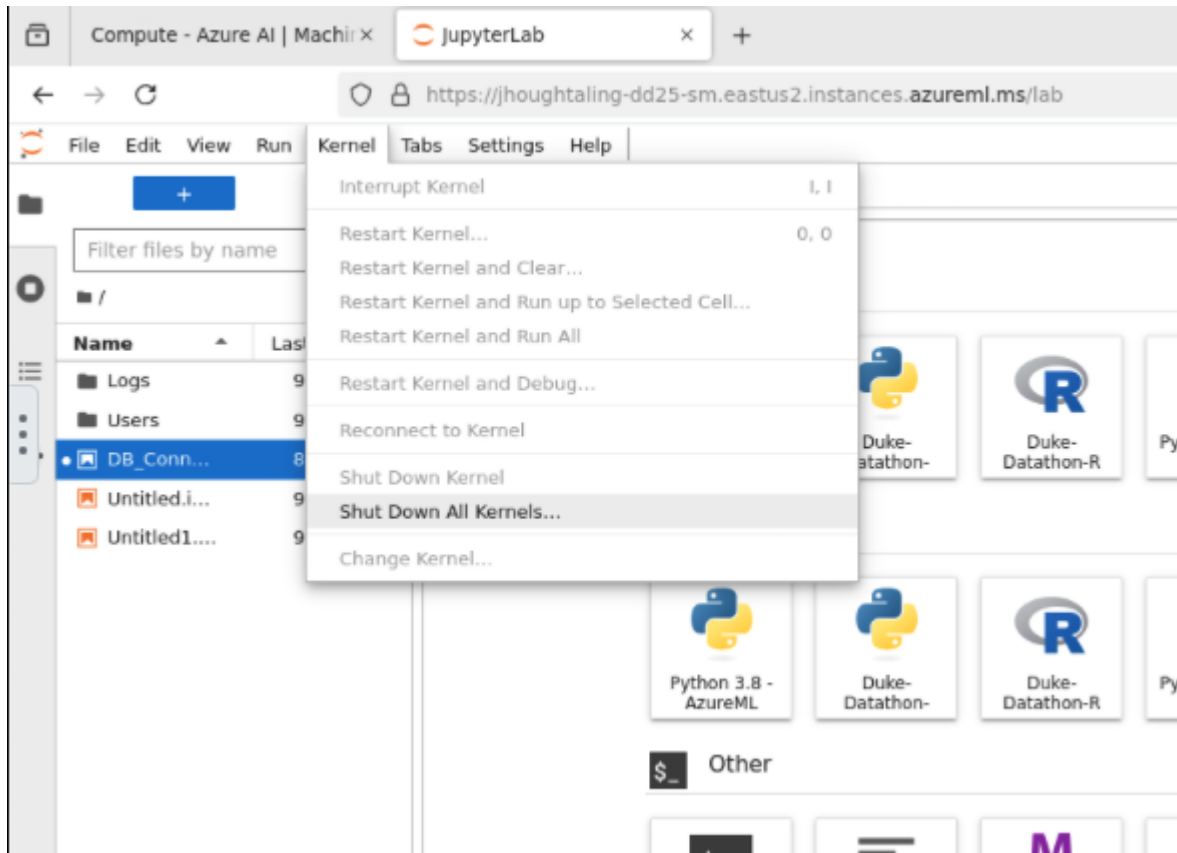
Clicking this icon will spin up your Jupyter Lab. a list of computes for the Datathon. The Jupyter Lab application takes about 30 seconds to be available. Inside Jupyter Lab you'll see the following:



To work around some interesting constraints Microsoft has placed on the compute deployment, we have created two Docker images to support custom kernels in this space. One image contains a number of Python packages relevant to observational health analytics, and the other contains the entire suite of OHDSI R packages for working with data in the OMOP CDM. Note that you can

view the definitions of these images, along with all of the other components of the infrastructure, here: <https://github.com/chorus-ai/chorus-container-apps/tree/main/mlw-jupyter-kernel>

Also note, we recommend keeping track of, and regularly shutting down, your kernels when you are not running jobs to mitigate potentially hanging processes with Docker-based kernel management:



I've launched my own JupyterLab environment, where's the data?

For OMOP-shaped tabular data, we've provisioned a common datathon Postgres instance that contains information for 20'000 individuals. You can connect to this database (datathon) using `psycopg` in Python, or `DatabaseConnector` in R, via a designated postgres user for the datathon. Note that the Datathon user has read-only access to the omopcdm schema (where the OMOP data is located) and has read/write access to the public schema in case you need scratch tables for certain processes. An example script to connect to the database can be found under the notebook called DB_Connect_exmple:

```
import psycopg2

host = "psql-hiti-datathon.postgres.database.azure.com"
port = 5432
dbname = "datathon"
user = "datathon"
password = "XXXXXXXXXX"

try:
    conn = psycopg2.connect(
        host=host,
        port=port,
        dbname=dbname,
        user=user,
        password=password,
        sslmode="require" # Azure PostgreSQL requires SSL
    )

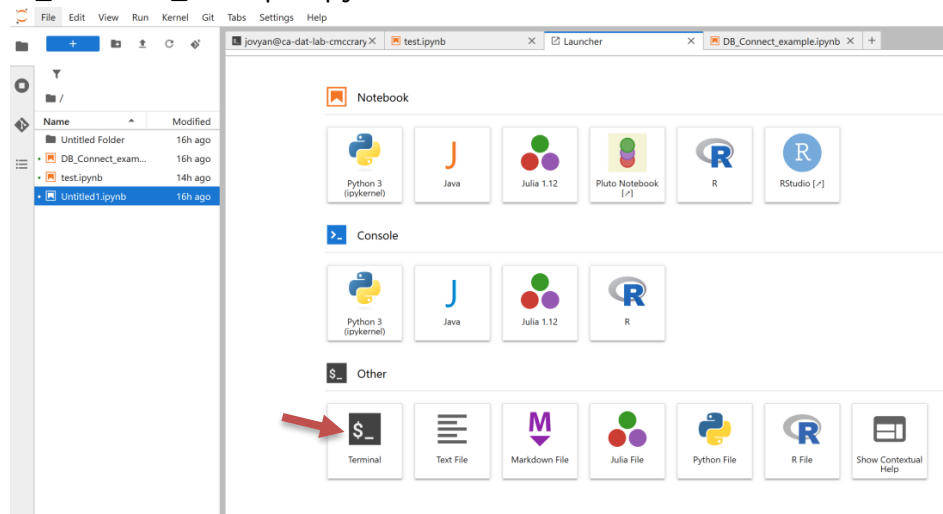
    print("✅ Connected successfully!")

    with conn.cursor() as cur:
        cur.execute("SELECT version();")
        version = cur.fetchone()
        print("PostgreSQL Version:", version[0])

    conn.close()

except Exception as e:
    print("❌ Connection failed")
    print(e)
```

If you do not see the notebook, you will need to move it from the home director with `cp /datathon/data/DB_Connect_example.ipynb ~/` under a new command terminal:



The data has a file structure of

Are there other ways to explore the data?

Yes! You can navigate through clicking 'Get Started' under ARES to check out the applicaiton to get a view on the OMOP data content and associated data quality:

[Notebooks - Azure AI | Mac](#)
[lab - JupyterLab](#)
[ARES](#)

https://ca-dukedatathon-www.blackdesert-9aaf2ae7.eastus2.azurecontainerapps.io/ares/#/cdm/DUKE-DATATHON/20250121/procedure_occurrence

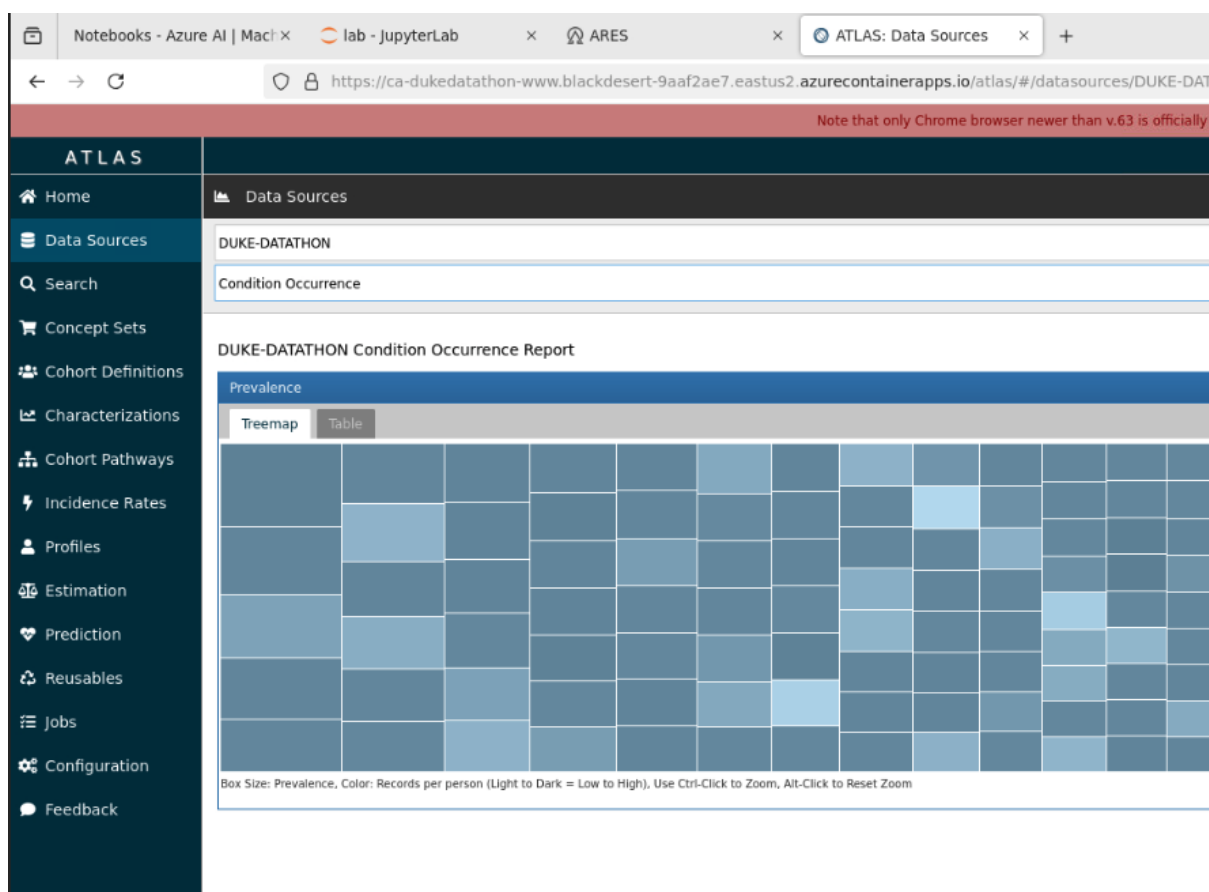
Report Category: **Data Source Release**
 Data Source: **DUKE-DATATHON**
 Data Source Release: **2025-01-21**
 Report: **Procedures**

PROCEDURE OCCURRENCE

Search in Table
 [CHOOSE COLUMNS TO DISPLAY](#)

Concept Id	Concept Name
725068	Radiologic examination, chest; single view
2514441	Critical care, evaluation and management of the critically ill or critically injured patient; first 30-74 minutes
2313816	Electrocardiogram, routine ECG with at least 12 leads; interpretation and report only
2514409	Subsequent hospital inpatient or observation care, per day, for the evaluation and management of a patient, which requires a medically appropriate history and/or examination and high level of medical decision making. When using total time on the date of t
2514408	Subsequent hospital inpatient or observation care, per day, for the evaluation and management of a patient, which requires a medically appropriate history and/or examination and moderate level of medical decision making. When using total time on the date
4145308	12 lead ECG

You can also use Atlas to create concept sets, cohort definitions/phenotypes, characterizations, prediction and estimation models, etc:



Access to GitHub

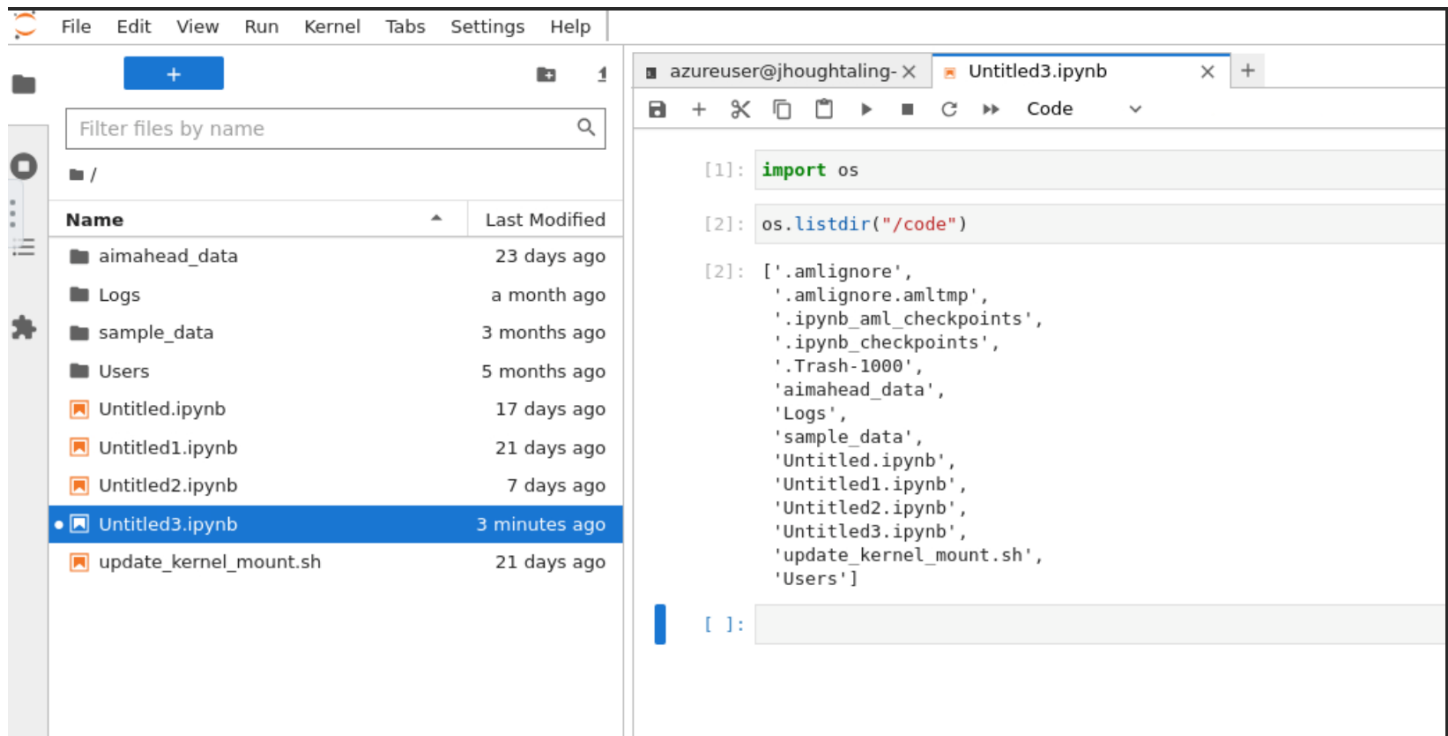
I don't have a GitHub ID

Please register for one at [GitHub.com](https://github.com). We've created a video for you if you're facing any difficulties: [How to Register for GitHub \(and get a GitHub ID\)](#)

Common Data Access Issues

I can't access the waveforms.

1. Shut down all active kernels
2. Open a terminal window in the jupyterlab application within your MLW compute
3. Make sure you're in the /code directory (you should see the same files/dirs as the screenshot above)
4. Run the script (./update_kernel_mount.sh)
5. Open a new Python kernel, and you should be able to find them located in the directory "/code":



Common Access Issues

I changed institutions and need to change my institutional email to log in.

If a user's email address changes, the CHoRUS team will need to remove it entirely, then recreate it.

Trouble logging in after already accepting the invitation?

If a users is having troubling signing in after they have already successfully accepted the invitation, then the CHoRUS team will need to reset the redemption status and ask them to login again.

No accounts at your institution is able to log in?

If a users is having troubling signing in with any institution email and other people from the institution cannot log in then the issue would be a DNS issue. This can be temporary resolved if the institution gives the CHoRUS team what IP address so we can allow access to the institution. The institution will also have to put the CHoRUS website on the whitelist.

Unanswered questions

Please use the trouble ticket system within the Connect platform <https://connect.aim-ahead.net/>

Frequently used acronyms

D

E

EHR: Electronic health record. Generally interchangeable with EMR.

EMR: Electronic medical record. Generally interchangeable with EHR.

I

ICU: Intensive care unit.

S

SCCM: Society of Critical Care Medicine. <https://sccm.org/>

Resources

For first-time data scientists

(please add below)



For first-time clinicians

(please add below)



For basic scientists

(please add below)

