Overview

The NIH Rapid Acceleration of Diagnostics Data Hub (RADx Data Hub) is a centralized data repository that provides access to analytic tools and de-identified COVID-19 data from the RADx Initiative. The RADx Data Hub supports scientific efforts to better understand COVID-19 and factors associated with disparities in morbidity and mortality in underserved and vulnerable populations, by allowing researchers to discover, access, and perform analyses of COVID-19 datasets in a cloud-enabled platform.

Introduction

This tutorial provides in-depth, step-by-step instructions on how to use the features and functionality of the RADx Data Hub most effectively. If you still have questions after reading the tutorial, consult some of the other support documentation (e.g. the Frequently Asked Questions page or the Glossary) or reach out directly by using the "Contact Us" link in the main navigation bar or the footer.

Target Audience

The primary audience for this tutorial is external researchers. Internal NIH staff should consult the appropriate Standard Operating Procedures (SOPs) on the Helpful Information page.

The Features of the RADx Data Hub

The RADx Data Hub contains several different features to help you get the most out of the system and better meet your research needs. These include:

- **Common Page Navigation Tools:** The RADx Data Hub features navigation tools (e.g. the navigation bar, the footer, various widgets) that will lead you to different pages and features within the Hub.
- **Home Page:** The Home page is a one-stop shop for many of the key resources on the RADx Data Hub. From this page, you can quickly reach educational documents (e.g. the Frequently Asked Questions [FAQ], the User Tutorial), get information on news, funding opportunities, events, and study updates, search for studies, and view statistics on the information in the Hub.
- **Study Explorer:** The Study Explorer contains a number of discovery features (e.g. search, sorting, filtering) to help you quickly and easily find studies.
- **Study Overview Pages:** Each study in the RADx Data Hub has its own Study Overview page, which contains study metadata, variables, data files, and downloadable documents.

- **Variables Catalog:** This tabular resource lists all variables in each data file for each study. By viewing this resource, you can gain a deeper understanding of the key variables in a study to help you determine whether it aligns with your research goals before requesting access to the study.
- **Support Resources:** The system gives you multiple ways to submit a support request, so you can ask questions, report bugs, and request in-depth assistance from the Support team on complex questions. You can use the "Contact Us" link in the navigation bar or in the footer.
- **User Registration:** To access certain features, such as the Approved Data tab, you will need to first register with the RADx Data Hub. After you have registered, you can login using the "Login" button in the top-right of every page to access role-based features.
- **Approved Data:** After you have been approved for data, you can access the "My Approved Data" tab. From here, you can apply for a workbench instance, download data, or transfer it to your workbench instance.
- **Public Data:** The Public Data page has synthetic data files, which you can practice using our "Analytics Workbench" feature.

The Study Explorer

The publicly available Study Explorer lets you search RADx study metadata to find studies for your research. In the Study Explorer, you can:

- View available RADx Data Hub studies
- Perform free-text searches
- Navigate results
- Refine results through sorting and filtering

View Available RADx Data Hub Studies

To view available RADx Data Hub studies, click "Study Explorer" in the upper navigation bar. You will be taken to the Study Explorer, where all findable RADx Data Hub studies present, by default, in Table View.
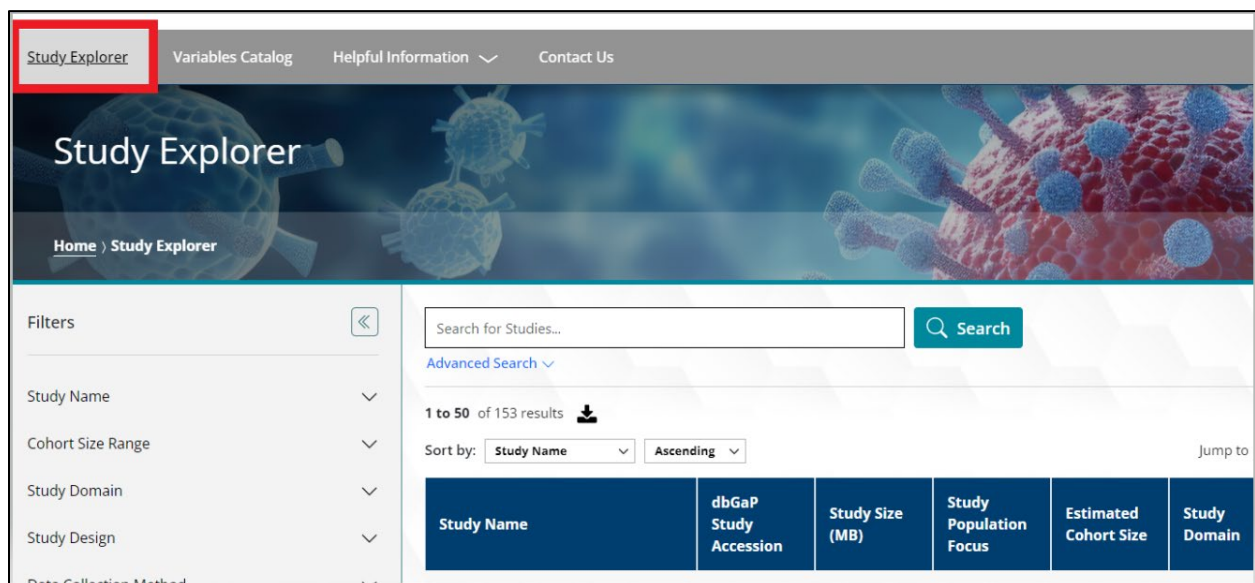
Figure 1: Study Explorer Link in the Navigation Bar

Use the toggle in the top right to switch to List View (Figure 2), which presents each search result in a vertically arranged list.
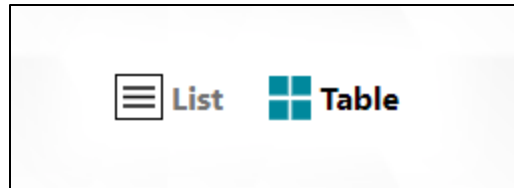
Figure 2: List View - Table View Toggle

Performing Free-Text Searches & Viewing Search Results

You can perform free-text searches by entering custom queries in the search bar. To perform a free-text search in the Study Explorer:

1. Click "Study Explorer" in the navigation bar.
2. Locate the Search bar (Figure 3).
3. Enter a free-text query.
4. Press "Enter" or click the magnifying glass icon to view results, sorted by relevance based on the entered query.

> *Tip:* You can also search directly from the Home page.
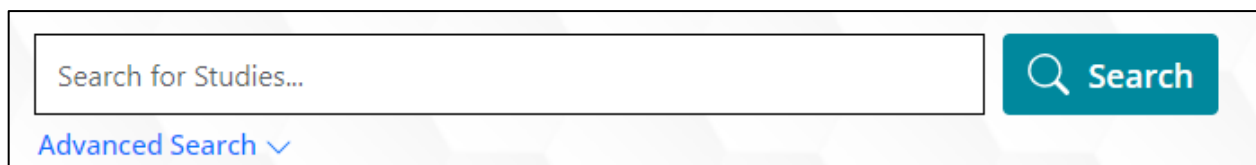
Figure 3: Study Search Bar

Navigating Through Search Results

After performing a search, use the page navigator in the top right of the Study Explorer to move through search result pages (Figure 4). To navigate through pages of the results, you can:

- *Option 1: Click the forward or backward arrows* to move one page forward or one page backward in the results.
- *Option 2*: *Click individual page numbers* (typically in the format: 1, 2, … X) to go to an individual page.
- *Option 3: To skip to any search result page*, type the desired page number into the free-text field to the right of "Jump to Page." Then, press "Go."

Figure 4: Page Navigator

In addition to changing the search results view, you can also change the number of results per page. To do this, click the "Show" dropdown at the top or bottom right (Figure 4). Then, select the number of results to show per page, and the page will automatically update.

To export search results to a csv file, click the "Download" button to the right of the sorting dropdowns.
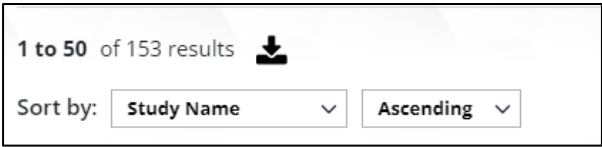

Figure 5: "Download" Button

Refining Results through Sorting and Filtering

Sorting and filtering search results can further refine a search. To sort in the Study Explorer:
1. Locate the sorting options in the top left. (Figure 6)
2. Pick either **"**Ascending" or "Descending" in the sort order dropdown.
3. Use the sort by dropdown to select a field for sorting, and results will dynamically update.
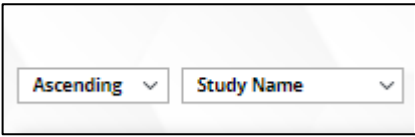
Figure 6: Sorting Dropdowns

Filtering is more complex than sorting but can help further refine a search. The filter pane is made up of two primary components: filter categories and filter values. Filter categories (e.g., "Hast Data Files" or "RADx Data Program") are high-level buckets that include multiple filter values. Filter values are the specific criteria by which you can filter search results. For example, the filter values in the "Has Data Files" category are "Yes" or "No."

The three techniques to narrow or refine a search using filters include:

Table 1: Different Search Techniques

Technique	Description	Example
Select One Filter Value	Selecting a single value in a filter category will filter the results by that value.	If you click "Yes" under "Has Data Files," you will only see studies with data files.
Select Multiple Filter Values in One Filter Category	Selecting two or more values in one filter category will function as a Boolean "OR."	If you click "RADx-UP" and "RADx-rad" within "RADx Data Program," you will see studies that align to either program.
Select Multiple Filter Values Across Filter Categories	Selecting two or more values across two or more filter categories will narrow the search and function as a Boolean "AND."	If you click "RADx-UP" from "RADx Data Program" and "Yes" under Has Data Files," you will see studies that align with RADx-UP AND have data files.

After deciding a filtering technique, select values by expanding the accordion for the desired filter category (Figure 7). Next, click the checkbox next to the filter value, and results will dynamically update. To remove a single filter value, click the checkbox a second time or click the "X" button on the filter badge above the filter pane. To remove all filters, press "Reset Search" above the filter pane.

Note: The numbers to the right of the filter values represent the number of results a selected value will return.

If you are still unable to find what you need, read the Advanced Search Tutorial to learn more search techniques.
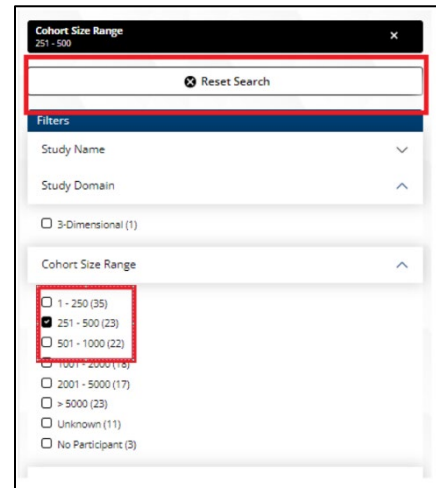

Figure 7: Filter Box

Advanced Search

Advanced Search lets you perform searches using Boolean logic. This function requires a basic understanding of the query-building principles.

Note: This tool does NOT ADD to the original keyword search. It is an independent search function.

Running a Single Query

1. Click the "Advanced Search" button to the right of the Search bar. This will disable the Search bar and open the Query Builder.
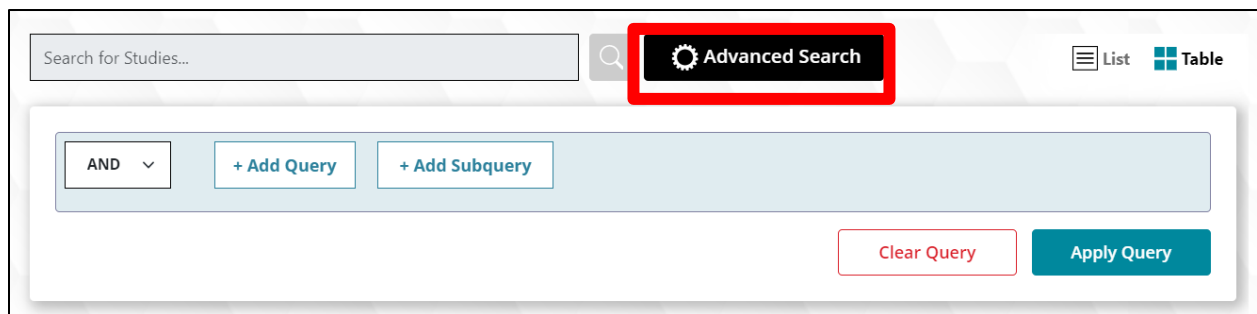

Figure 8: "Advanced Search" Button

2. Click "+ Add Query" to add a search query line to the base query group. Remove query lines by clicking the red "X."
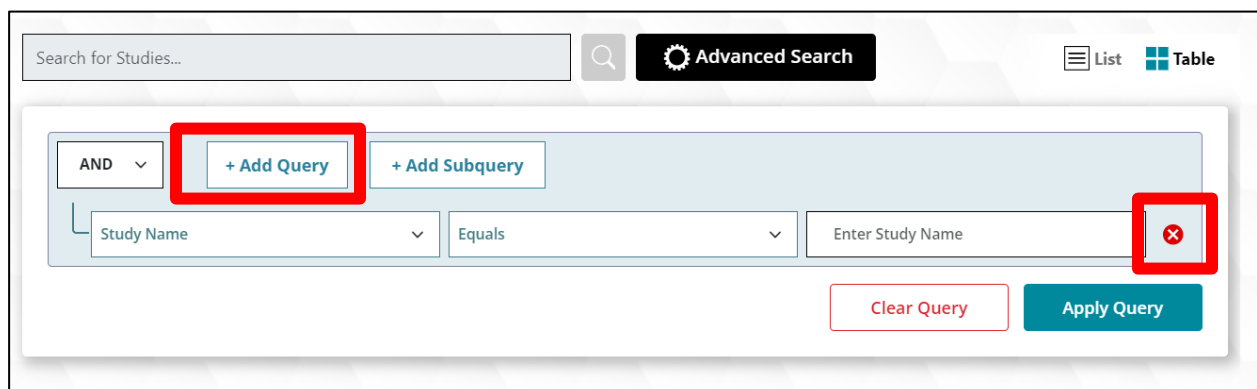
Figure 9: "Add Query" Button

Select an option from the first dropdown menu to use the provided metadata fields and direct the search.
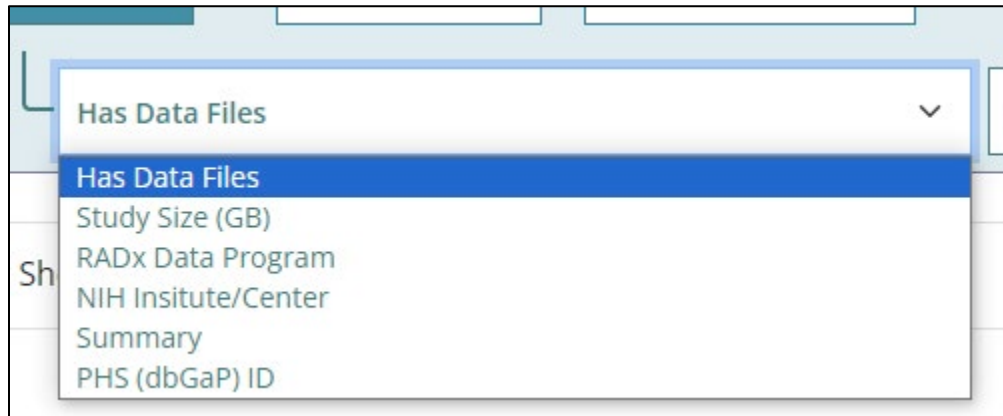
Figure 10: Metadata Attribute Dropdown

3. Choose which query type within the operator dropdown menu.
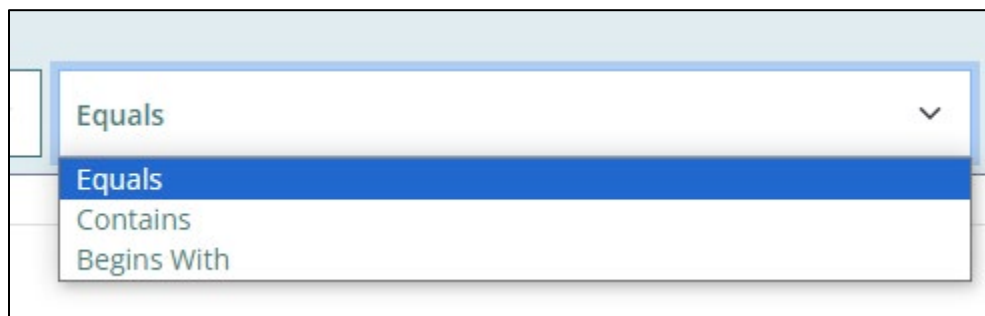
Figure 11: Query Type Dropdown

4. Enter an input value and click "Apply Query" to complete the query. The system will update the search results upon doing so.
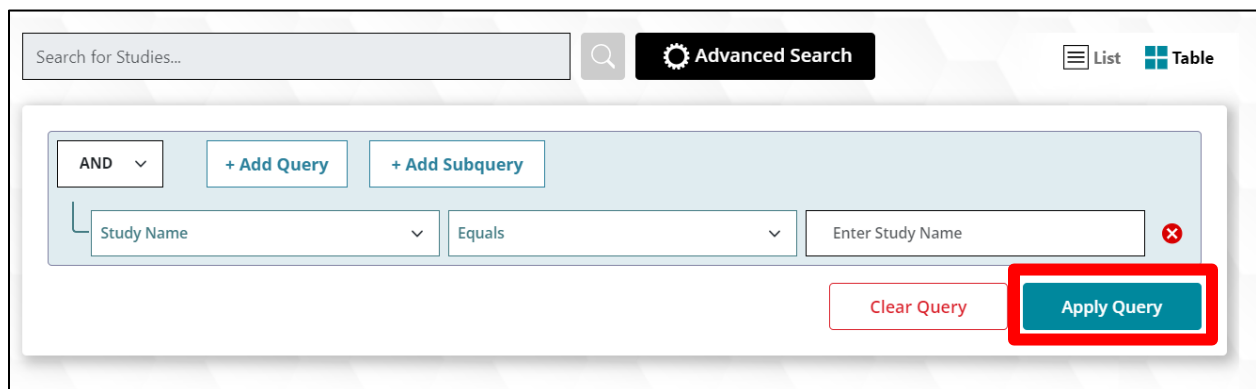
Figure 12: "Apply Query" Button

Running Additional Queries and Subqueries

After building an initial query, add another query or a subquery to further refine a search using one of two methods:

- Add a query to the initial query to refine the search by selecting an option from the Boolean Combinator ("AND" or "OR").
- Add a subquery to perform an additional logical function within the initial query.

1. Select either "AND" or "OR" from the Boolean Combinator dropdown. "AND" narrows a search based on the selected queries. "OR" broadens a search.
2. Click "+Add Query" to add another query to an initial query. Then, repeat the previous steps to build an additional query.
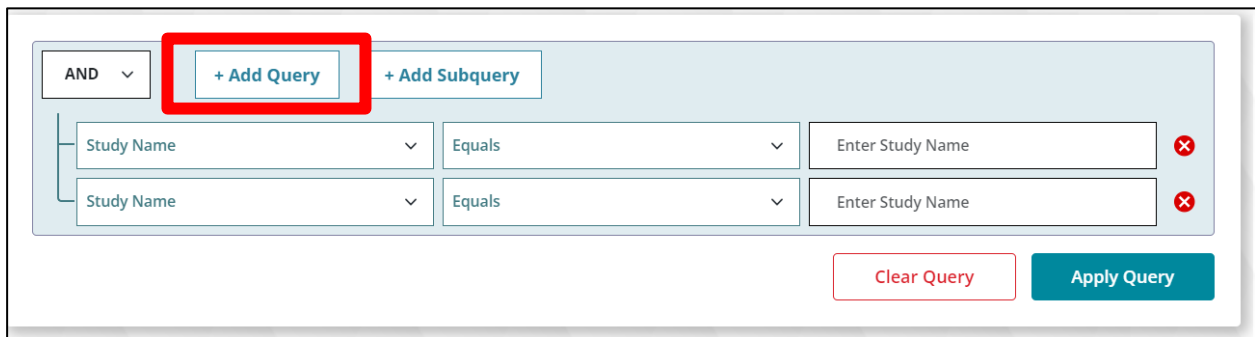
Figure 13: "Add Query" Button

3. Click "+Add Subquery" to add a subquery to an initial or subsequent queries. Since the options for building subqueries are the same as those for building queries, follow the initial query-building steps to build a subquery.
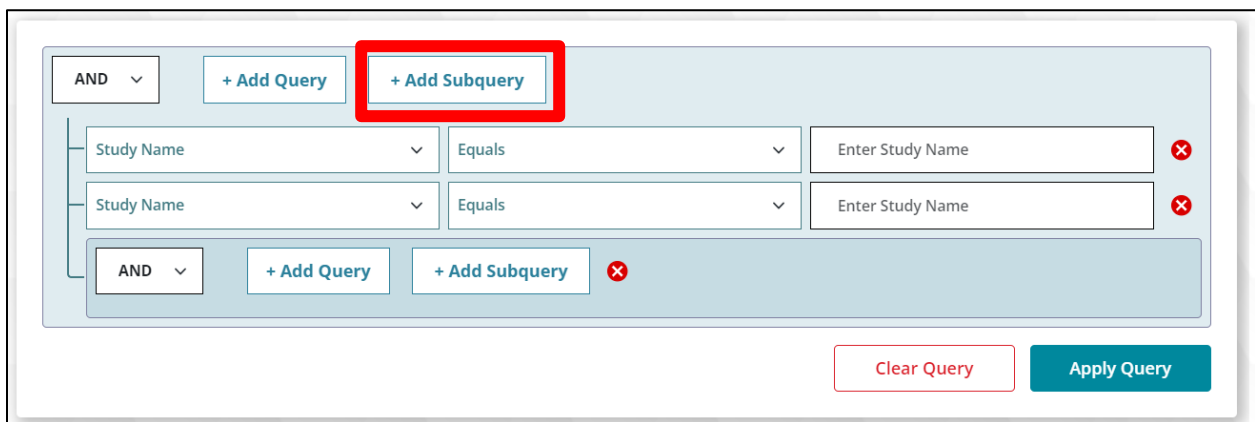
Figure 14: "Add Subquery" Button

Note: To clear a query, press "Clear Query". Then, click "Apply Query" to reset the search results.

Study Overview

Each study in the system has an overview page, which contains key documents, metadata, and variable and file information. To reach the Study Overview page, you must locate a study in the Study Explorer or Variables Catalog and click on the "Study Name."

On the Study Overview page, you can:

- View study information
- Learn how to request study access
- Download publicly available documents
- Learn about data files and download resources

View Study Information

Study Overview contains comprehensive study information, beginning with the study name, and a link to the corresponding dbGaP study overview page. Below that, a Study Information section lists several attributes to help you gain a high-level understanding of the study.

Tip: You can expand or collapse the Additional Information section by interacting with the accordion.
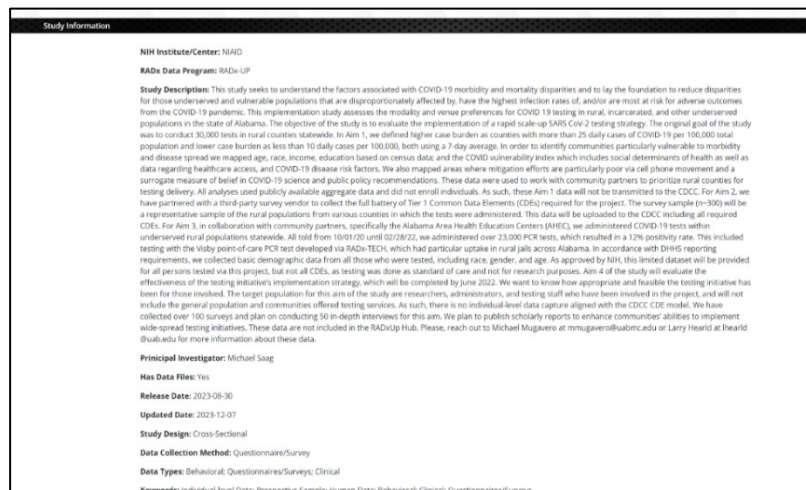
Figure 15: Study Overview Page

Learn How to Request Study Access

To gain study data access, including harmonized and non-harmonized data files, you must request access in dbGaP. To do so:

1. Login into the RADx Data Hub using the same eRA or NIH Login used for dbGaP.

2. Locate a study in RADx using the Study Explorer and click on the Study Name to go to the Study Overview page.
3. Click on the dbGaP Link in the top left of the Study Information section on the Study Overview page to go to the Study Overview page in dbGaP (Figure 16).
4. In dbGaP, submit a study data access request. For more information on requesting dbGaP study access, read the dbGaP article on requesting access.
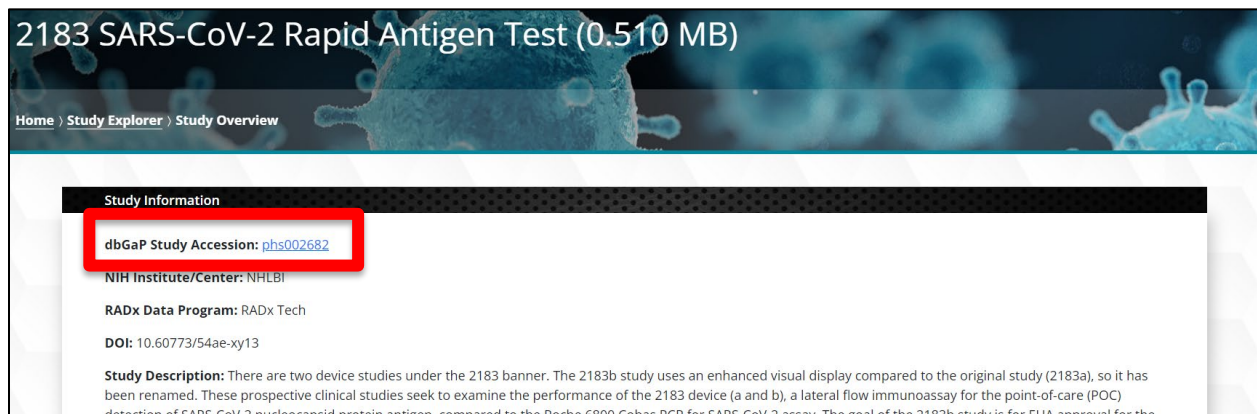

Figure 16: dbGaP Link in Study Information Section

Note: If you are accessing a DHT study, you will also see a link to the corresponding Study Page in RAPIDS in the Study Information section.

Download Documents

The Study Overview page contains downloadable, publicly available documents. There are two ways to download documents (Figure 17):

- *Option 1 - Download Individual Documents:* To download a single document, press the download icon in the download column on the documents table. The document should appear in the browser's download center for you to open or save.
- *Option 2 - Download All Documents:* To download all documents, press "Download All" in the upper right of the documents section. A zip file with all study documents will appear in the browser's download center for you to open or save.

Figure 17: Download Functionality in Documents Table

Learn About Data Files and Download Resources

The Data Files section has downloadable metadata and data dictionary files and viewable variable information to help you learn more about a study and its data, before requesting access.

In the Data Files table, data and supporting files are organized by bundles. Specifically, each bundle contains a metadata and data dictionary file, aligned to a harmonized or non-harmonized data file.

While you cannot download original or transformed files from the Study Overview page, you can download associated metadata and data dictionary files. To do this, click the download icon in either the Metadata or Dictionary columns (Figure 18). The file will show up in the browser's download center for you to open or to save.

Tip: For studies with lots of data files, use sorting to locate a file more quickly. Click a column header once to sort in ascending order, and twice for descending.

Figure 18: Data Files Table

You can view metadata files in an easy-to-read, interactive tool called the Metadata Viewer, powered by CEDAR. To access the Metadata Viewer, click on the download icon in the Data Files table. This will open a window that lists several different metadata attributes in the file. To the right of each label are help tips, which contain metadata attribute descriptions. Additionally, you can expand accordions to see more metadata attribute information.

Figure 19: Metadata Viewer

You can also view information on data file-based variables contained by clicking the carrot in the "Number of Variables" column (Figure 20). This will open a table, letting you find the data file variables.
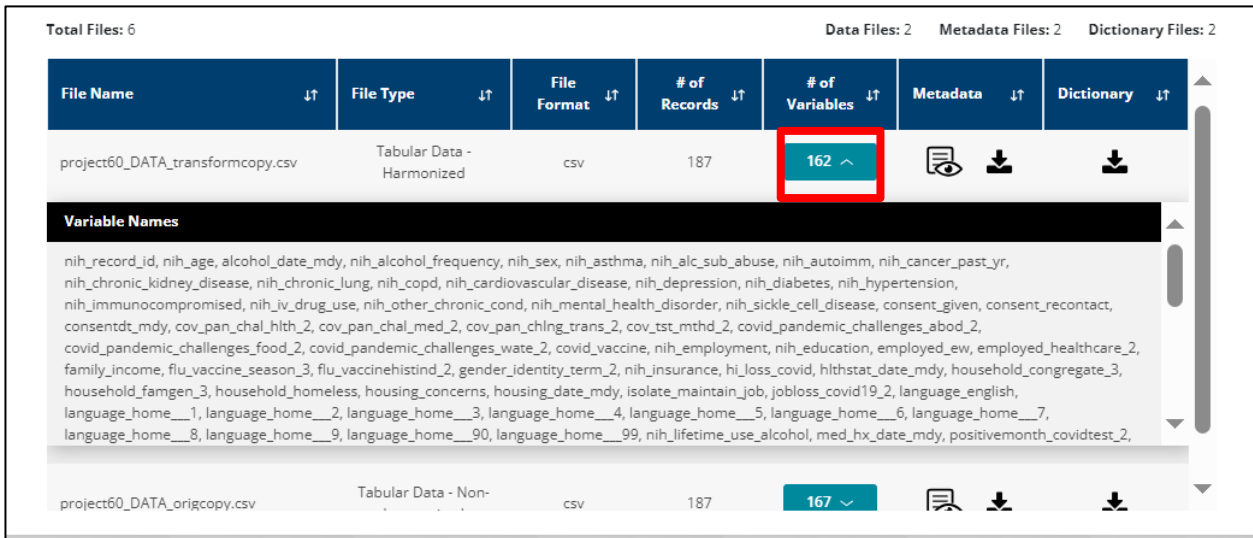
Figure 20: Variables Information in the Data Files Table

Variables Catalog

The Variables Catalog is publicly available and displays variables for each study data file as a comma-separated list. With the Variables Catalog, you can quickly understand a study's variables to make a more informed decision when requesting study data access. To reach the Variables Catalog, click the link in the top navigation bar.
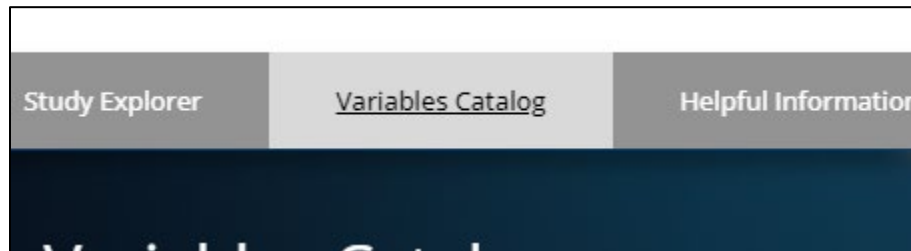
Figure 21: Variables Catalog Link in Navigation Bar

Once in the Variables Catalog, you can:
1. Find and view variable information
2. Navigate to the RADx and Database of Genotypes and Phenotypes (dbGaP) Study Overview Pages
3. Download catalog content as a CSV file

View Variable Information

View variable information in the Variable Catalog (Figure 22) in two ways:
1. *All Variables* lists out all data file variables in a comma-separated list.
2. *RADx Core Variables* lists RADx Core variables and labels in a tabular format.
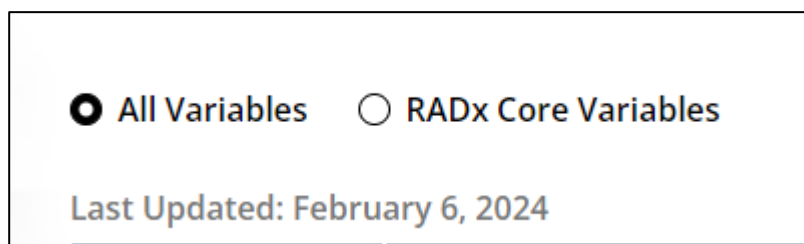
Figure 22: Variables Catalog Toggle

To switch between views in the table, select the radio button next to the desired view. After selecting an appropriate view, search for variables, studies, and more using the web browser's "Find" feature (CTRL/CMD + F).

Navigate to RADx and dbGaP Study Overview Pages

In the Variables Catalog, access the RADx Data Hub and dbGaP Study Overview pages by clicking the following links:

1. *The Study Name* links to the RADx Data Hub Study Overview page, where you can gain a better understanding of study metadata.
2. *The dbGaP ID (phs ID)* leads to the dgGaP Study Overview page, where you can request study-level access.
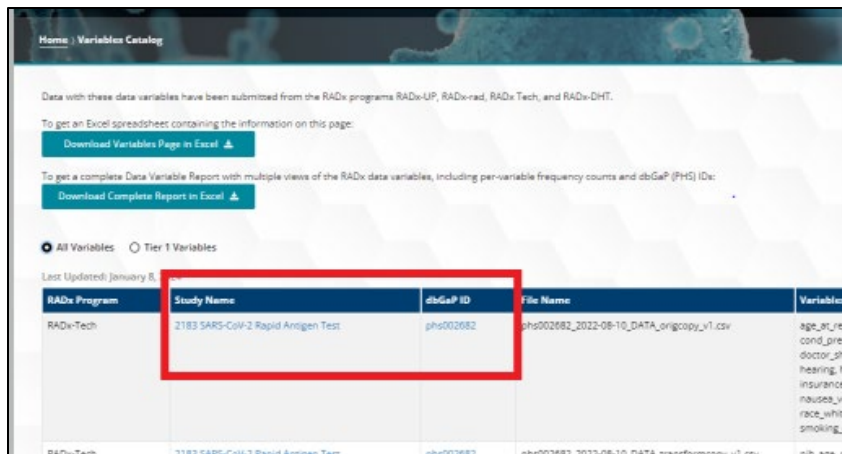

Figure 23: Links to the RADx Data Hub and dbGaP Study Overview Pages

Download Reports

In addition to viewing the Variables Catalog, you can also download the variable information as a CSV file:

1. *Default Report:* Matches what you see already on the UI with one exception; instead of having two views, this report only has one view with all the variables included. To download this report, click "Download Variables Catalog in Excel."
2. *Complete Report:* Expands on the original report with additional views and information, such as RADx program variable reports and Tier 1 variable. To download this report, click "Download Complete Report in Excel."
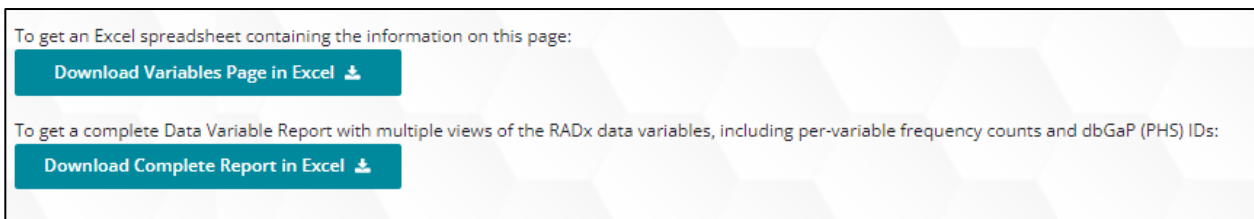
Figure 24: Downloadable Variables Reports

User Support

The Support Team can answer questions for various topics, such as technical issues and bugs, questions about analytics tools, feature requests, and more. You can reach our support team by using the support form, which is available from any page in the application. To contact our support team, follow these steps:

1. Create a support request clicking "Contact Us" on the navigation bar or in the footer. This will redirect you to the User Support Request Form.
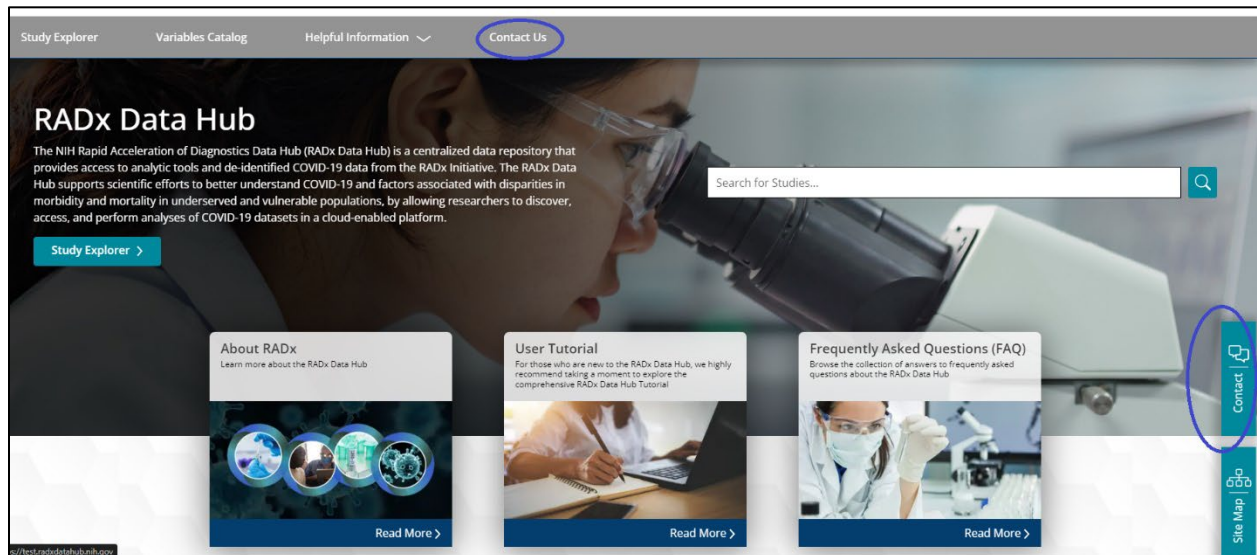

Figure 25: "Contact Us" Link in Navigation Bar

2. Complete the required fields, indicated with an asterisk.

All fields marked with asterisk (*) are required.

* Email

* Full Name

* Institution

* Request Type

* Request Title

* Request Detail

Submit

Figure 26: Support Request Form

3. Choose the appropriate option under "Request Type" to route the request.

* Request Type

General Feedback
Technical
Feature Request
Engagement
Workbench Support

Figure 27: Request Type Dropdown

- *General Feedback* – Provide Data Hub site feedback.
- *Technical* – Report bugs or other technical problems.
- *Feature Request* – Provide new feature suggestions.
- *Engagement* – Request a RADX Data Hub training, presentation, or demonstration.
- *Workbench Support* – Request help on the Analytics Workbench.

4. Under "Request Title," briefly describe the request.
5. Provide specific details about the request under "Request Details."
6. Click "Submit" to complete the support request. You will receive an automated email confirming the Support Team received the request. This will include the ticket number and any further instructions. Soon after submission, a member of the Support Team will contact you with further questions or possible resolutions.

User Registration

Create an Account

To create a RADx Data Hub account, you will first need a Researcher Auth Service (RAS) Identity Provider (IdP) account, specifically an NIH Login or eRA Commons account. Once you have an account with either IdP, you can register with the RADx Data Hub following the steps below:

1. To register, click "Login" in the top right corner on any RADx Data Hub page. The system will display a modal with a Login/Sign Up using RAS button (Figure 28).
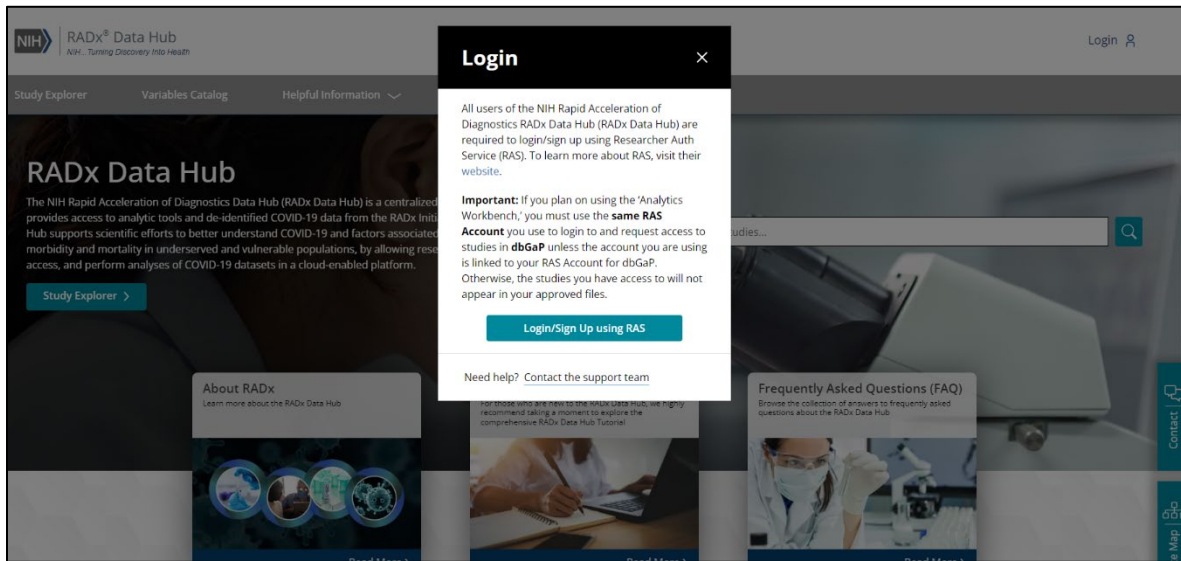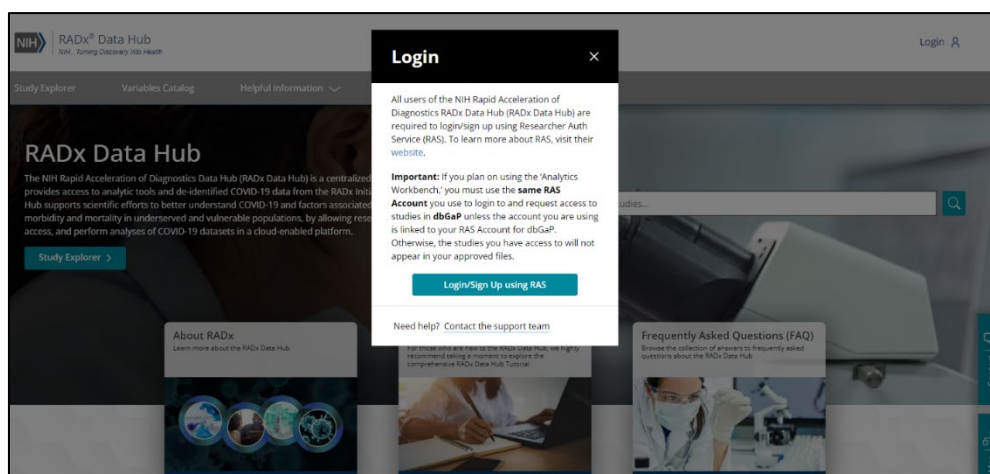

Figure 28: Login Modal

2. Press "Login/Sign-Up Using RAS," and you will be directed to the RAS Sign-In page (Figure 29).
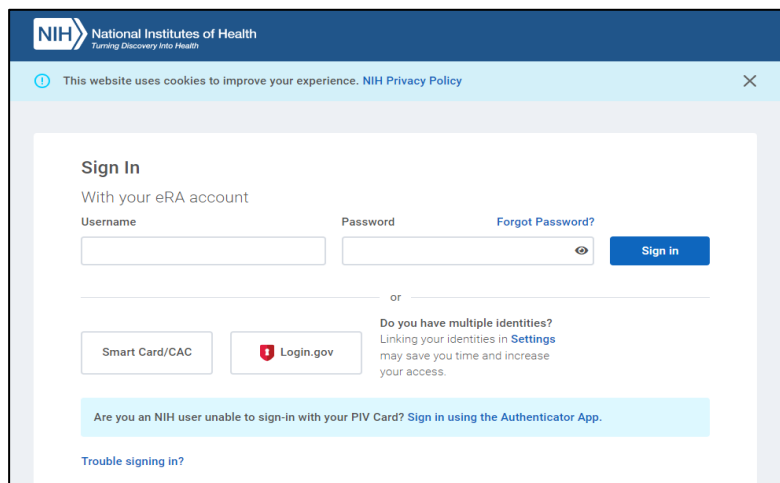
Figure 29: RAS Login Page

3. Enter your eRA or NIH Login credentials, and the system will redirect you to the RADx Data Hub User Registration page. (Figure 30).

Figure 30: User Registration Form

4. Fill in the required fields. The system automatically displays First Name**,** Last Name, Middle Initial (M.I.), and Email based on your RAS information. If these are incorrect, please contact RAS.
5. Click the Institution dropdown and add an institution. If you cannot find your institution, press "Click here to add an institution" under the dropdown. Fill out the required fields (Figure 31) and press "Add Institution." After that, the institution should appear in the Institution dropdown.

Figure 31: Add Institution Form

6. Carefully read the Terms and Conditions (Figure 32) and click the provided box to accept conditions.
7. Press "Submit" to finish registering. The system will automatically log you in, redirect you to the Home page, and send an email confirming the registration.

Figure 32: Terms & Conditions

Note: RADx Data Hub leverages dbGaP to manage study access. When registering for the RADx Data Hub, register using the same RAS credentials used for dbGaP (or a linked account). Failure to do this will make it impossible for the system to show your authorized studies available in "My Approved Data."

Login to the RADx Data Hub

1. Click "Login" in the top right corner on any RADx Data Hub page. The system will display an additional modal with a login/sign up option (Figure 33).

Figure 33: Login Modal

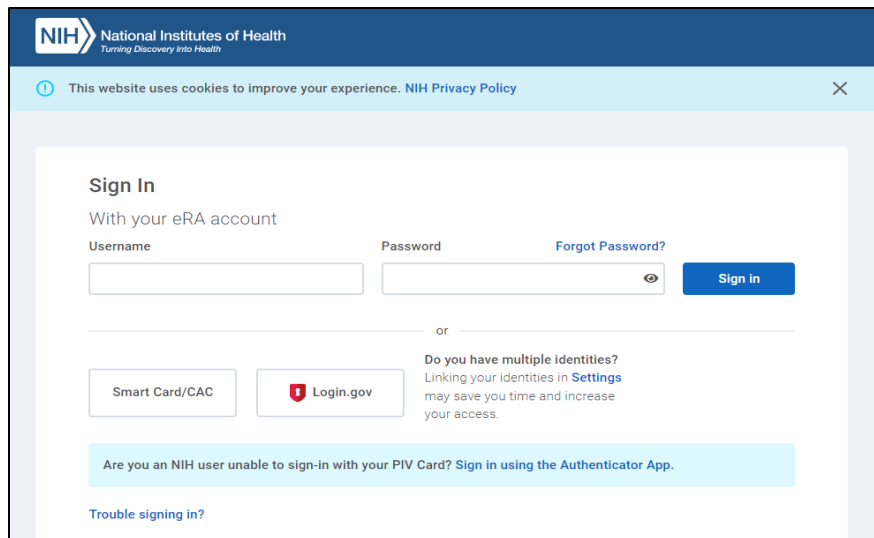2. Press "Login/Sign-Up Using RAS," and you will be directed to the RAS sign-in page (Figure 34).

Figure 34: RAS Login Page

3. Select the NIH Login or eRA option and enter login credentials. The system will redirect you to the home page after correctly entering your credentials.

Note: If you have a login.gov account linked to an NIH Login or eRA account used for dbGaP, you may use that to login.

My Approved Data

My Approved Data allows you to access data files based on your Database of Genotypes and Phenotypes (dbGaP) approvals. This page will group files into one table per study. Within each study table, the files are organized by bundles, which is a set of files including a data file, metadata file, and data dictionary (along with any SAS file equivalents if available).

After accessing My Approved Data, you can:

- [Download approved data files](#)
- [Apply for workbench add-ons](#)
- [Create and launch a workbench](#)
- [Add files to your workbench](#)

Note: My Approved Data will be empty until you are approved to access to at least one RADx Data Hub study in dbGaP.
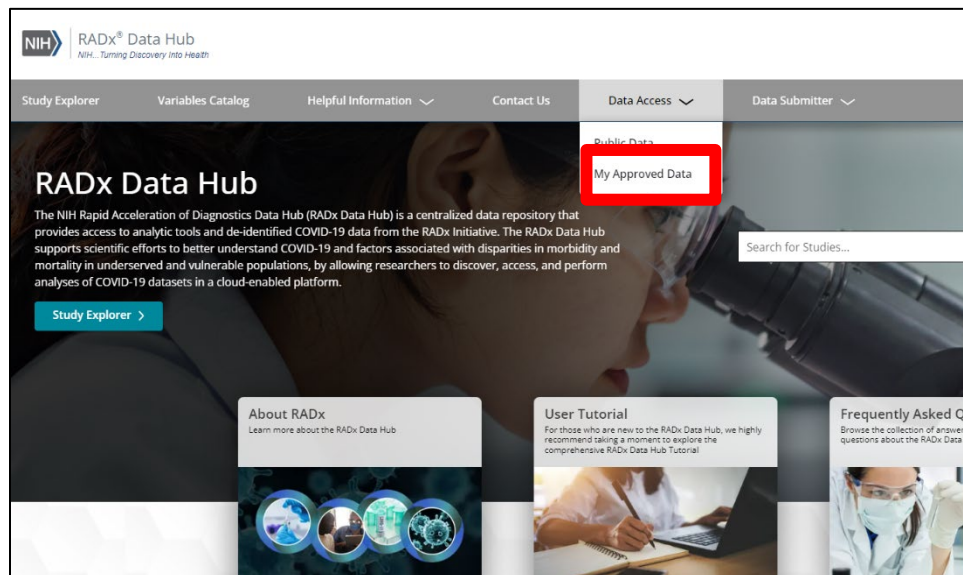

Figure 35: My Approved Data Link in Navigation Bar

Download Files

To download files, select them using one of three techniques:

- *Option 1 - Select a Bundle:* To select a bundle, check the box next to a harmonized or non-harmonized data file. This will select the high-level data file and its supporting files (i.e., metadata and data dictionary files).

- *Option 2 - Select an Individual File*: This process varies for supporting files (e.g., metadata and data dictionary files) and data files. To select a metadata or data dictionary file, check the box next to the file name. To select an individual harmonized or non-harmonized file, first check the box next to the file name. This will select the entire bundle. Then, deselect the metadata and data dictionary files.
- *Option 3 – Select All Files*: Click the "Select All" button. This will select all study table files.

After selecting the desired files, you can click "Zip & Download," and the files will appear in your browser's download center as a zip file, which you can save or open.
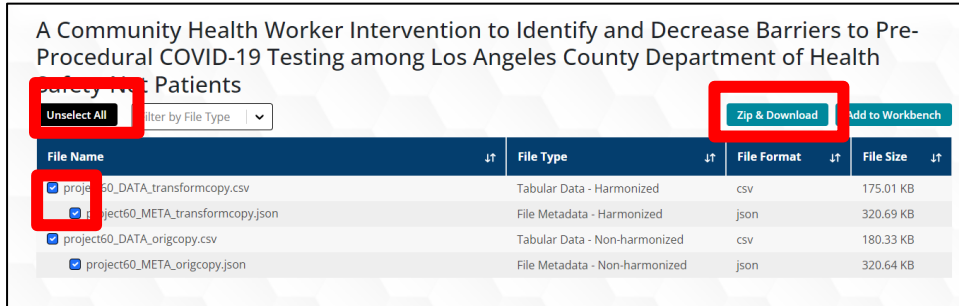

Figure 36: Selection Functionality and "Zip &Download" Button

Apply for Add-ons

After registering, you will automatically have default Workbench access, which includes JupyterLab and RStudio but does not include Data Wrangler or SAS Viya. To access those tools, you must apply for a Workbench Add-on.

To apply:
1. Press "Apply for Add-on" (Figure 37). The system will direct you to a separate page displaying the request form.
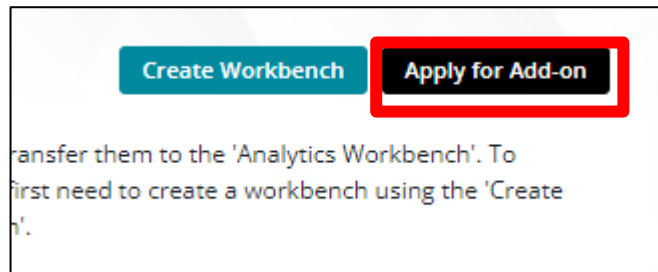

Figure 37: "Apply for Add-on" Button

2. Fill out the required fields.

Figure 38: Add-on Application Form

3. Download and read the Terms of Service. You should type your name in the "Agree to the Terms of Service" field if you agree.
4. Press "Submit." The RADx Data Hub team will review the request and email you if you are approved for a Workbench Add-on.

Note: There are a limited number of licenses available for SAS and Data Wrangler. Approval is on a first-come, first-serve basis. You may only hold one license at a time.

Create and Launch a Workbench

To create a workbench, press "Create Workbench." In the future, press "Launch Workbench" to access an existing workbench.
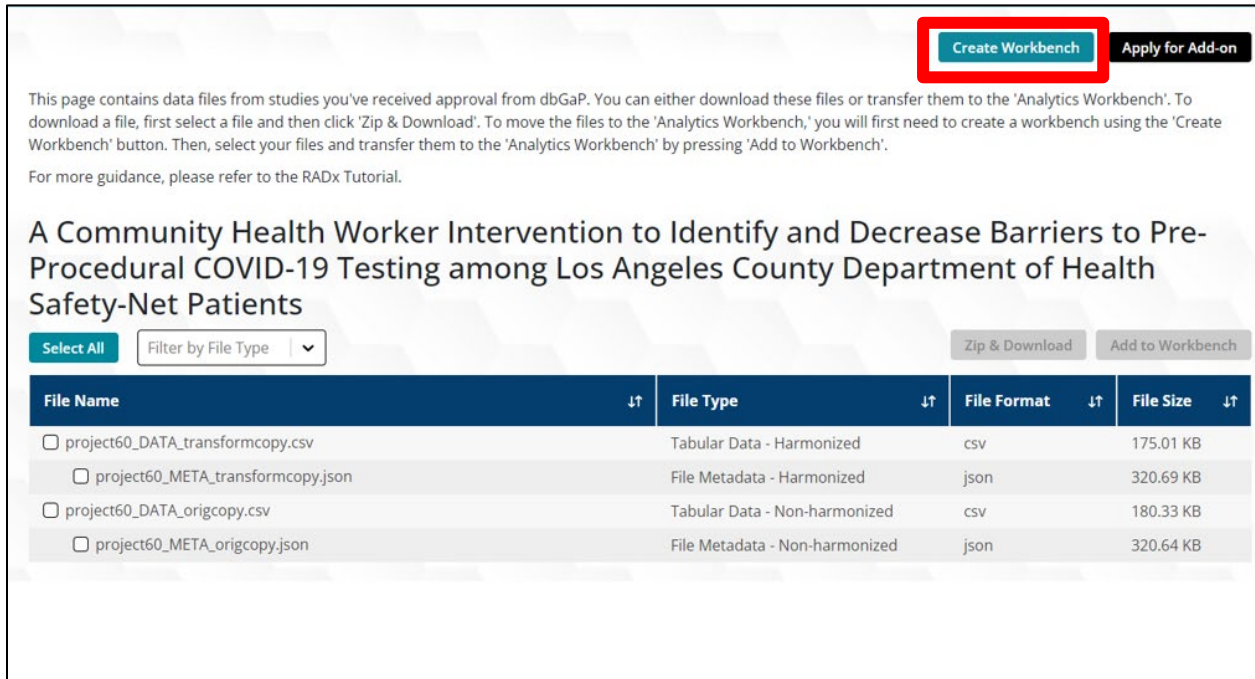
Figure 39: "Create Workbench" Button

Add Files to Workbench

To add approved files to a workbench, select the file or files you want to transfer. Then, click "Add to Workbench" in the top right of the study table (Figure 40).
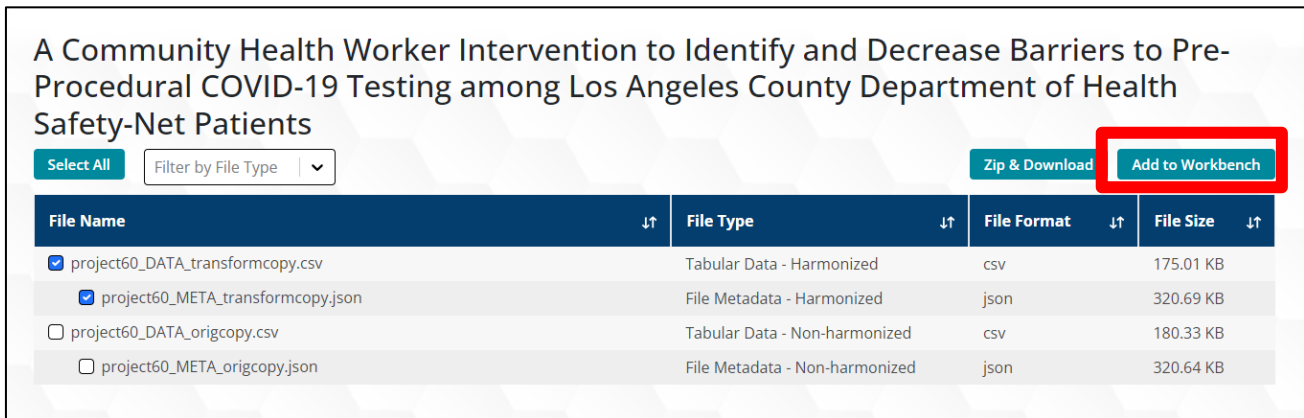

Figure 40: "Add to Workbench" Button

Public Data Tutorial

The Public Data page contains synthetic data files. Prepared using mock data, these files allow you to test analytics workbench offerings and capabilities before using real research data.

You must log into the Data Hub to access the Public Data page. After logging in, you will see a "Data Access" dropdown in the navigation bar. You can click "Public Data" to go to the Public Data page.
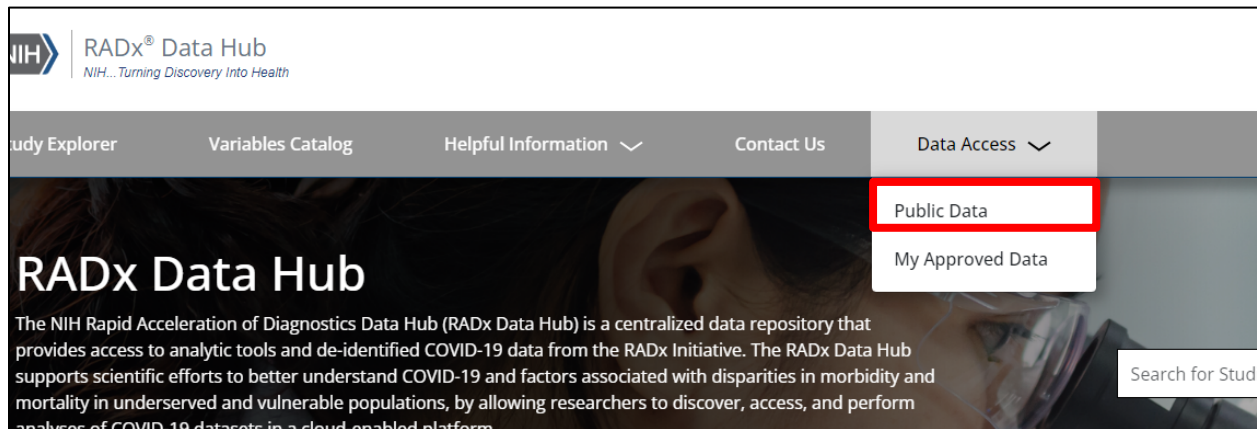

Figure 41: "Public Data" Link in the Navigation Bar

Once on the "Public Data" page, you can:

- Download public data files
- Create or launch a workbench
- Transfer files to your workbench

Download Files

To download files, select one or multiple files using one of following techniques:

- *Option 1 –Select an Individual File*: To select a single file, click the checkbox next to the file name.
- *Option 2 – Select Multiple Files:* To select multiple files, click the checkbox next to the files.
- *Option 3 – Select All Files*: Click "Select All." to select all files within the table.

After selecting files, click "Zip & Download," and the files will appear in your browser's download center as a zip file to save or open. The "Zip & Download" button will be greyed out until you select a file.
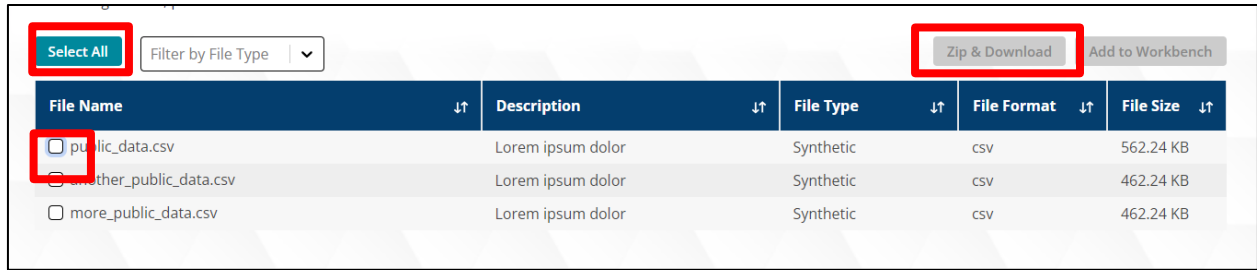
Figure 42: Selection Functionality and "Zip & Download" Link

Create and Launch Workbench

To launch the workbench, you can press the "Create Workbench" button on the Public Data page, which will automatically route you to the workbench application. After creating a workbench, you can go back to your workbench instance by pressing the "Launch Workbench" button (Figure 43).
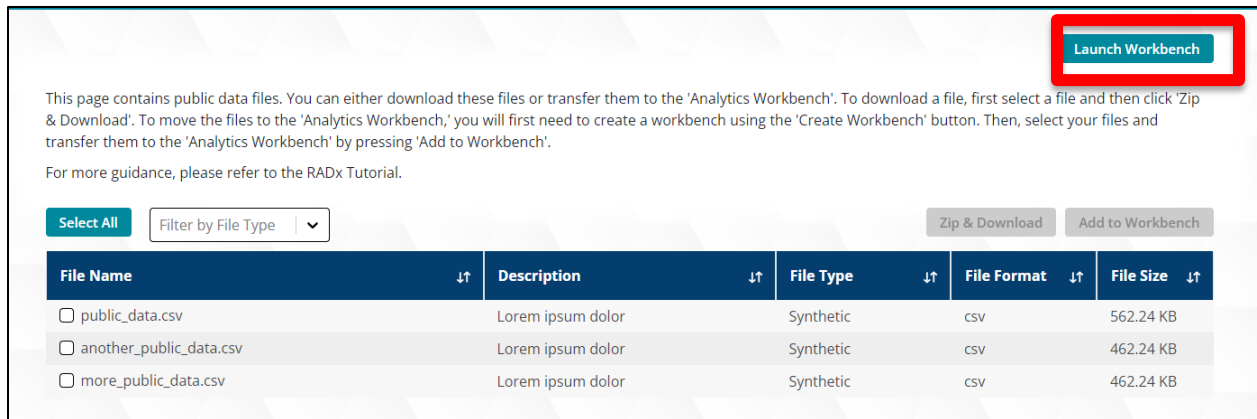

Figure 43: Launch Workbench Button on Public Data Page

Transition Files to Workbench

After creating a workbench instance, you can transfer files to the workbench. To add files to a workbench, you can first select the file or files you want to transfer. Then, press the "Add to Workbench" button to transfer them to your workbench instance (Figure 44).
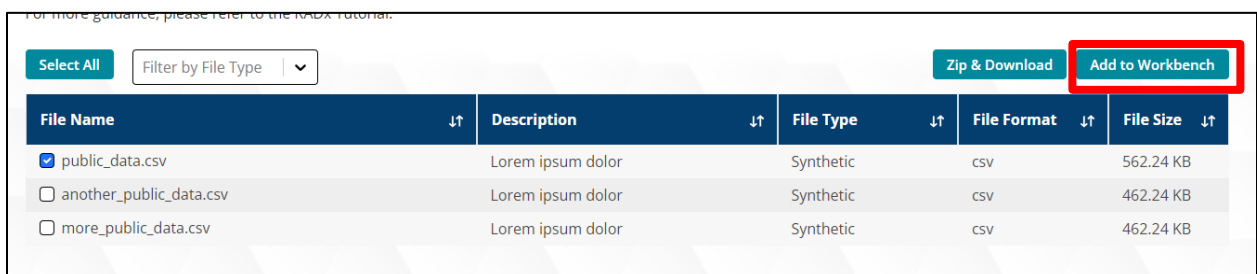

Figure 44: "Add to Workbench" Button

The Analytics Workbench

The Analytics Workbench allows users to launch compute instances with Jupyter notebooks, using Python or R, in a personal workspace environment. Workbench add-ons include 1) Data Wrangler, a no-code data transformation, analysis, and visualization option and 2) SAS Viya Analytics Pro, a cloud-hosted SAS environment that allows scalable computing, data storage, and usage tracking to enable data access, transform, analysis, visualization, and mapping capabilities.

Note: There is a limited number of licenses available for SAS and Data Wrangler. The licenses are distributed on a first come first serve basis.
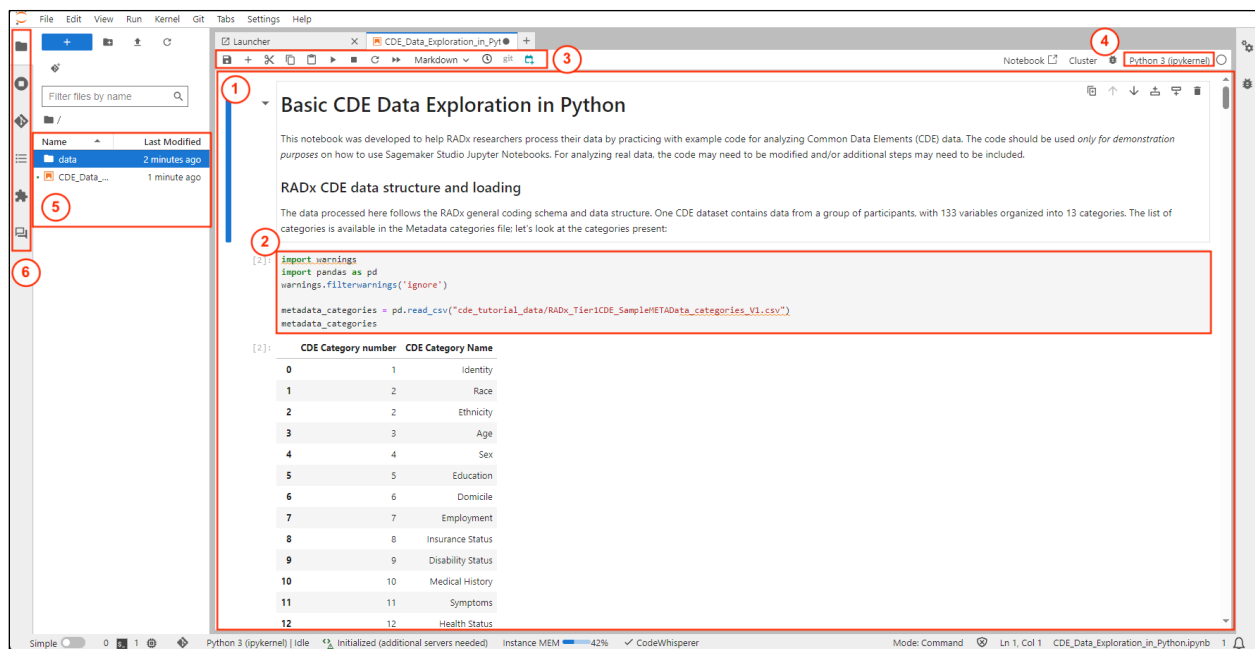
JupyterLab

Figure 45: Labeled Jupyter Notebook Interface

The main elements of the JupyterLab editor are:

1. **Notebook**: A document containing analysis code, outputs, and any additional markdown or text.
2. **Cell:** A single section of a notebook where to enter code, markdown, or text.
3. **Toolbar:** Perform the most common notebook actions, including:
   - Save
   - Insert cell below
   - Cut selected cell
   - Copy selected cell
   - Paste from clipboard
   - Run selected cell
   - Interrupt the kernel

- Restart the kernel
- Restart the kernel and run all cells
- Change cell type (i.e., Code, Markdown, Raw)
- Launch terminal

4. **Environment:** Displays the current notebook kernel type.
5. **File Browser**: Displays lists of folders, notebooks, and other files.
   - The Personal Studio environment is a private, personal Amazon EFS directory.
6. **Left sidebar**: Contains tabs to access the following functionalities.
   - **File Browser**: Displays lists of folders, notebooks, and other files.
   - **Running Terminals and Kernels**: View current kernels and terminals running in JupyterLab. Optionally shut down all or select resources (i.e., notebooks, terminals, kernels, apps, and instances).
   - **Git:** Connects to a Git repository for Git tool and operation access.
   - **Table of Contents**: Automatically generated for each notebook, Markdown file, or Python file open to navigate the document's structure with clickable entries.
   - **Extension Manager:** Enables and manages third-party JupyterLab extensions.
   - **Jupyter AI**: A JupyterLab tool to explore generative AI models and integrate them into notebooks.

Create and Launch a JupyterLab Space

The default workspace environment is a ml.t3.medium (2 vCPU, 4 GiB memory) instance type.

To create a new JupyterLab space:

1. When the Workbench is launched, select "JupyterLab" from the "Overview" section, or select "JupyterLab" from the "Applications" in the left panel (Figure 2).
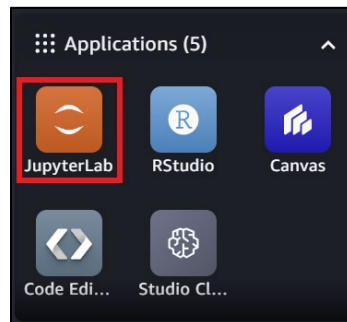

Figure 46:Workbench applications highlighting JupyterLab

2. Select "Create JupyterLab space"
   - In the "Create JupyterLab space" dialog, specify a name for the space in the "Name" field. To finish, click "Create space."

To launch a JupyterLab space:

1. From the Workbench Home page, select "JupyterLab" from the Overview section, or select "JupyterLab" from "Applications" in the left panel (Figure 2).
2. Select "Run" in the Action column of the JupyterLab space to start the workspace (Figure 3). This may take up to a minute to start.
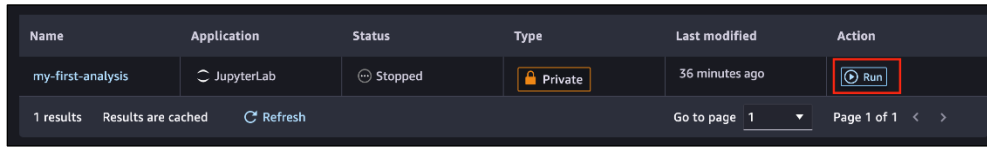
Figure 47: Start Running JupyterLab Space

3. Once the status changes to "Running", select the "Open" icon in the Action column to launch JupyterLab in a new tab (Figure 4).
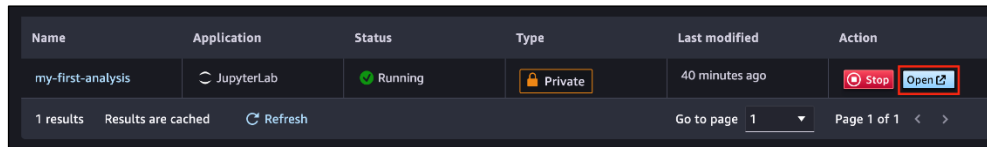

Figure 48: Open JupyterLab Space

To create a new notebook:

1. From the landing page, select "File," "New," and "Notebook" (Figure 5).
   o In the "Select Kernel" dialog, select a kernel on the dropdown menu. To finish, click "Select", which launches the notebook.
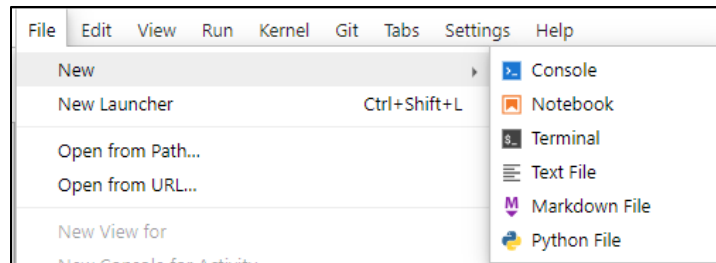

Figure 49: Launch Notebook from File Menu

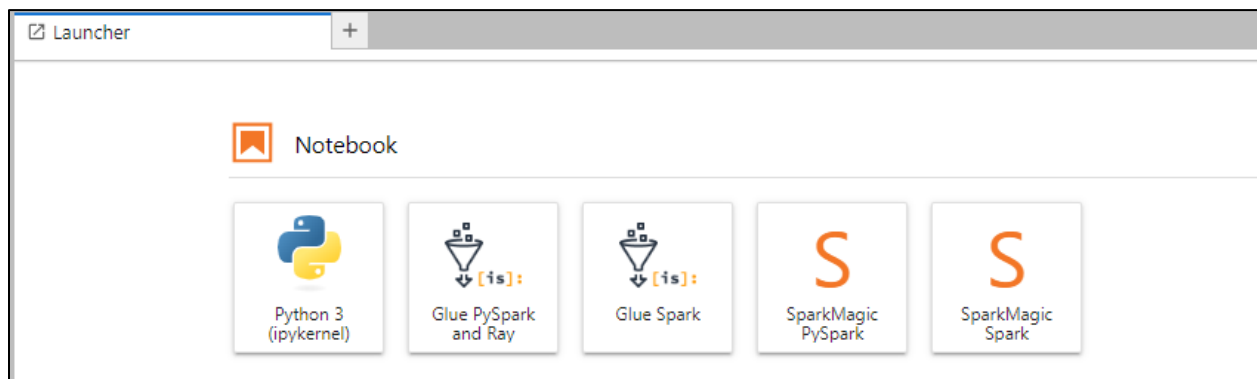2. From the Launcher page, click a preferred kernel in the Notebook section (Figure 6).

Figure 50: Launch Notebook Using Launcher

Upload and Download Files

To upload files from a local machine into a JupyterLab space:

1. In the left sidebar, choose the "File Browser" ■ icon.
2. In the File Browser, choose the "Upload Files" ▲ icon.
3. Select the files to upload and choose "Open."
4. Once the file appears in the home folder, double-click the file to open it in a new tab.

To download a file locally:

1. In the left sidebar, choose the File Browser ■ icon.
2. Right click the file and select "Download."

To download an entire folder locally:

1. From the menu, choose "File," "New," and "Terminal", which will launch a Terminal in a new JupyterLab tab.
2. Type the following command replacing folder_name and /path/to/folder:
   zip -r -X folder_name.zip /path/to/folder
3. Once the folder is zipped and it appears in the File Browser, right click the .zip file and select "Download."

Clone a Git Repository

Git repositories can be cloned into the JupyterLab home folder using the following steps:

1. Select the Git ◈ icon in the left sidebar.
2. Choose "Clone a Repository."
3. In the Clone Git Repository window, enter the Git URL (i.e., https://github.com/aws/amazon-sagemaker-examples.git)
4. Under "Project directory to clone into," enter the path to the local directory where the cloned directory should exist, otherwise Studio will clone the repository into the home directory.
5. Choose "Clone," which will automatically open a new terminal window and clone the repository. This may take up to a minute depending on the repository size.
6. If the repository requires credentials, a prompt will appear to enter a username and personal GitHub account access token.
7. When complete, the File Browser will open, displaying the cloned repository.
8. Choose the Git icon to view the Git user interface, which tracks the repository.
9. To track a different repository, open the repository in the file browser and click the Git icon.

Access Public Data

To access curated public and synthetic datasets on the RADx Data Hub's Data Access page, follow the Public Data Tutorial.

Datasets from the AWS Registry of Open Data, an AWS-hosted repository of more than 400 publicly available datasets, can be copied into a JupyterLab environment using the following steps:

1. Identify a dataset of interest and find the associated Amazon Resource Name (ARN).

      a.   For example: COVID-19 Data Lake
      b.   ARN: `arn:aws:s3:::covid19-lake`
      c.   The bucket name is `covid19-lake`

2. From the JupyterLab landing page, select "File," "New," then "Terminal."
3. Enter the following command: aws s3 sync s3://covid-lake **.**
4. Replace covid-lake with a selected dataset bucket name.

Change Environment

Notebooks launch with the minimum instance type available by default. The minimum instance type is appropriate for most tasks, however, a larger instance can be requested by submitting a Support Request. Follow the instructions in the User Support Requests Tutorial and select "Workbench Support" when choosing a Request Type. Please provide as much detail as possible in the request for the support team to determine the best suitable environment. For more detailed information about available instance types and their performance capabilities, see Available Studio Instance Types.

File Sync

If an added Workbench file does not appear in the File Browser of JupyterLab, the workspace should be resynced. Close the JupyterLab tab, and refresh the My Approved Data page. Then, follow the steps to relaunch the JupyterLab page. If the files still do not appear, the workspace may need to be manually synced with the following steps:

1. From the File menu, click "File," "New," and "Terminal."

2. Enter the following into the Terminal: ./s3sync.sh

If the files in a workspace are still missing, please submit a Support Request.

Data Wrangler

By default, Data Wrangler uses the m5.4xlarge (16 vCPU, 64 GiB memory) instance type. To request a difference compute instance, please see the Change environment section. If a Data Wrangler instance has been provisioned, a Data Wrangler flow can be created using the following steps:

1. From the Workbench Home page, select "Canvas" from "Applications" in the left panel.
2. Click "Run Canvas" to start the instance. This may take up to 8 minutes.
3. Once the instance status has changed to "Running", click the "Open Canvas" button to launch Canvas in a new tab.
4. Select the "Data Wrangler" application in the sidebar of Canvas applications and click "Create a data flow" which will open a dialog to rename the data flow for the analysis.
5. Click import data (Figure 7) and select the data file type (i.e., Tabular or Image). Select one of the data source options from the dropdown, or upload files directly. See Import to learn more about AWS data import options.
6. Data Wrangler can now be used to add transforms, analyze, and visualize your data. To learn more, see Transform Data and Analyze and Visualize.

7. To export a data flow, click **Export** from the data flow page. To learn more about exporting data transformations to other platforms, see [Export](#).
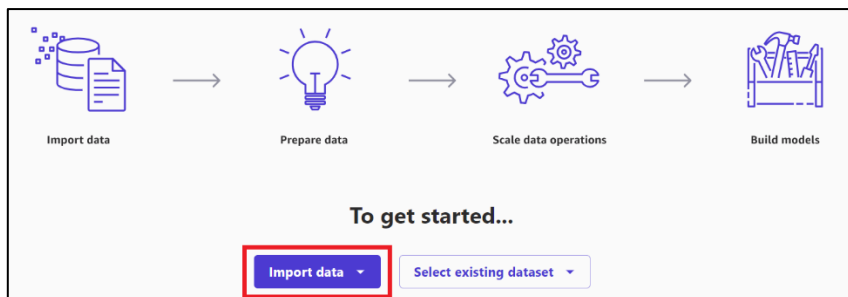

Figure 51: Import Data to Data Wrangler Flow

To stop running Canvas, click "Stop Canvas" on the Canvas homepage (Figure 8). Click the checkbox and the "Stop Canvas" button that appears to confirm shutdown. Be sure to save all work and data flows prior to shutdown.
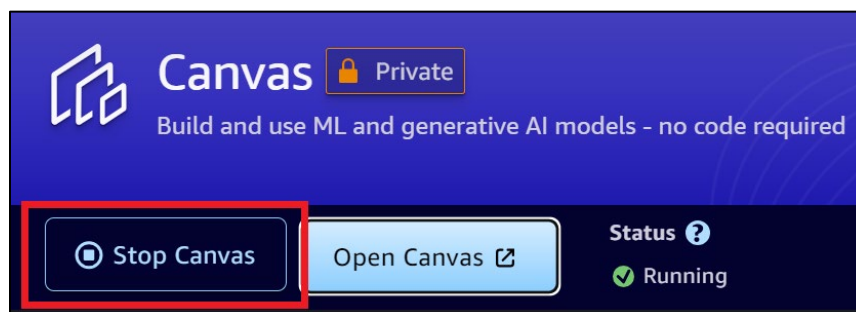

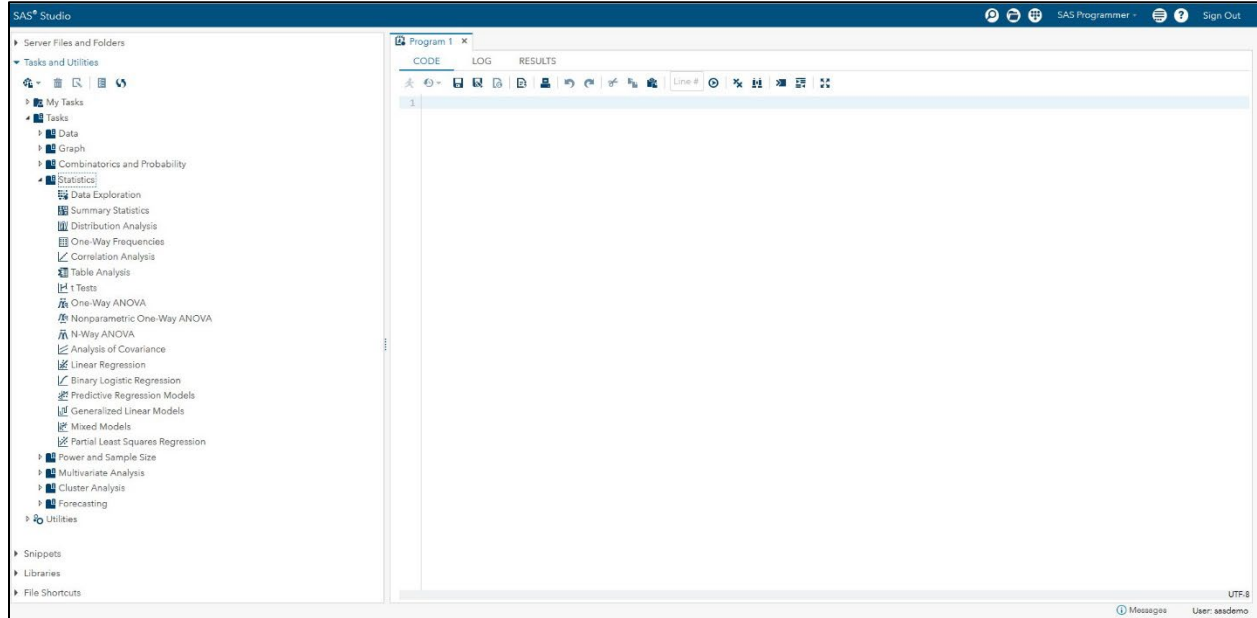**Figure 52: Stop Canvas Application**

SAS Viya

Figure 53: SAS Interface

From the My Approved Data or Public Data page, a "Launch SAS" button will appear if a license has been granted. To launch the SAS platform, click "Launch SAS."

To start a new program:

1. From the main menu, click "New", and "SAS Program". A new blank program will open in the work area.

To save a SAS program:

2. Click the "Save" 🖫 icon on the work area toolbar.
3. Select the location where files will be saved.
4. Enter the name of the program.
5. Click "Save".

To open a SAS program:

1. Click the 🗁 Open icon on the main menu bar.
2. Find the location of the program in the left panel.
3. Select the SAS program to open in the right panel.
4. Click "Open" and the program will appear in the work area.

To run just a portion of a SAS program, highlight the portion to run. To run the entire SAS program, no code needs to be highlighted:

1. Click the 🏃 Run icon on the work area toolbar.
2. Open the "Log" tab to confirm the program ran correctly.