

Datasets

Contents

- Datasets..... 3**
 - View Defined Datasets..... 3
 - Define a Dataset..... 3
 - View Dataset Information..... 6
 - Modify a Dataset..... 7
 - Execute a Dataset..... 8
 - Delete a Dataset..... 9

Datasets

A dataset is a specific view of data in PHEMI Central that is accessible to a given set of users. The users working with the dataset must have a role of Data Analyst. The view provided by the dataset can be across multiple data sources.

Any data elements from any data source may be included in a dataset. Defining a dataset does not modify configuration for data sources or impact protections to data.

Datasets are typically created for research or exploratory data use, and the intended use for datasets may differ from the intended use of an individual data source. In particular, the governance and controls required for a dataset may differ from those specified its constituent data sources. For this reason, you can attach an access policy to a dataset that is independent of the access policies attached to the constituent data sources.

Once a dataset is built by a PHEMI Administrator, it can be executed by users with a role of Data Analyst. A single dataset can be executed by multiple users--even by users with different authorizations. A given user's actual view into the data in an executed dataset is restricted by his or her authorization. Thus, two users with different authorizations might execute the same dataset but end up with a different set of data.

View Defined Datasets

View the list of defined datasets on the **Manage Datasets** page.

To see what datasets are available:

Open the **Manage Datasets** page, by clicking the **Datasets** icon in the left navigation bar.



The **Manage Datasets** page opens, listing the datasets that have already been defined.



Define a Dataset

Define a new dataset on the **Manage Datasets** page.



Note: Only users with a role of PHEMI Administrator can define datasets.

To define a dataset:

1.

Open the **Manage Datasets** page, by clicking the **Datasets** icon in the left navigation bar.



The **Manage Datasets** page opens, listing the datasets that have already been defined.

Manage Datasets

Datasets

☐ Select/Deselect All

☐ BCCA - OMEGA Study(not authorized)

☐ Cardiology Outcomes Analysis(not authorized)

☐ COO Task Force - Hypothesis 1(not authorized)

☐ Test Cardiology Claims(not authorized)

☐ USF Cardiology Example(not authorized)

2. Click the **Define New Dataset** button.

The Build New Dataset screen opens.

Build New Dataset

Name

Dataset Name

Description

Describe this dataset

Export Target

Hive-MR

Access Policy

Everyone can see everything

Status

Available

Available To

No Users Selected

Expires

NaN-0NaN-0NaN

Dataset Contents

Billing (0)

Cardiology (0)

Test (0)

Asset

Cardiology Pro...

com.phemi.wb.a.payer	?	-
com.phemi.wb.a.claim	?	-
com.phemi.wb.a.patientsex	?	-
com.phemi.wb.a.patientage	?	-
com.phemi.wb.a.claimid	?	-
com.phemi.wb.a.patientid	?	-
com.phemi.wb.a.admissionid	?	-

Export Preview

Select Digital Assets above to include in your exported dataset - a preview will then be shown here.

Close

Save Dataset

3. Describe the dataset.

Option	Description
Name	Mandatory. A name for the dataset.
Description	Mandatory. A brief description of the dataset.
Export Target	Optional. The format required by the system to receive the dataset. Supported values are as follows:

Option	Description <ul style="list-style-type: none"> • Hive-MR. Apache Hive MapReduce format. • JDBC. Java Database Connectivity format. • CSV. Comma-separated values format. • TSV. Tab-separated values format. <p>The default is Hive-MR.</p> <p>If you choose JDBC as the format, you have the option of selecting one of the dataset destinations that has been configured for the system. Tell me about dataset destinations.</p>
Access Policy	Optional. The access policy to protect the dataset. The access policy selected for the dataset may match or may be different from any applied to the constituent data source(s). By default, the first configured access policy is applied.
Status	Optional. The availability of the dataset. The only supported value is Available.
Available To	Mandatory. The users who are allowed to access this dataset. Select one or more users from the drop-down list; only users with a role of Data Analyst appear on the list. By default, no users are allowed to access the dataset.
Expires	Mandatory. An expiry date for the dataset. When this date is reached, PHEMI Central deletes the dataset from the system.

4. Choose the fields you want in the dataset.

You select the fields for the dataset in the **Dataset Contents** pane. This pane will show a number of tabs; these tabs reflect the data categories configured for your data. Clicking a tab will list all the fields extracted from all the data sources within the category, where each data source is represented by a column. Each data source and each field has a checkbox.

The screenshot shows the 'Dataset Contents' pane with several tabs: BCCA_LabResults (0), Billing (7), Cardiology (0), Genomics (0), BCCA_OncologyVCF (0), BCCA_PatientHistory (0), Radiology (0), and Test (0). The 'Billing (7)' tab is selected. Below the tabs is a table with columns for 'Asset', 'Cardiology Billing', 'Cardiology Claims', and 'Cardiology Proc'. Each row represents a data source with a checkbox in the 'Asset' column and checkboxes in the other columns. The following table represents the data shown in the screenshot:

Asset	Cardiology Billing	Cardiology Claims	Cardiology Proc
<input type="checkbox"/> com.phemi.wb.a.claim	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/> com.phemi.wb.a.payer	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/> com.phemi.wb.a.patientsex	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/> com.phemi.wb.a.patientage	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/> com.phemi.wb.a.claimid	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/> com.phemi.wb.a.patientid	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/> com.phemi.wb.a.admissionid	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

- Select an entire data source by checking the checkbox for the column.
- Select any number of individual fields.

Continue for each data category you want to include in the dataset.

As you select fields, a preview of your selections appears in **Export Preview** pane.

Export Preview

Billing (7)		
Asset	Destination Name	Destination Type
com.phemi.wb.a.claim	com.phemi_wb_a_claim	DOUBLE
com.phemi.wb.a.payer	com.phemi_wb_a_payer	STRING
com.phemi.wb.a.patientsex	com.phemi_wb_a_patientsex	STRING
com.phemi.wb.a.patientage	com.phemi_wb_a_patientage	LONG
com.phemi.wb.a.claimid	com.phemi_wb_a_claimid	STRING
com.phemi.wb.a.patientid	com.phemi_wb_a_patientid	STRING
com.phemi.wb.a.admissionid	com.phemi_wb_a_admissionid	STRING

- When your dataset is complete, click the **Save Dataset** button. The system confirms when the dataset has been successfully saved. Click **Close** to dismiss the **Build New Dataset** screen.

View Dataset Information

View information for a dataset on the **Manage Datasets** page.



Note: Only users with a role of PHEMI Administrator can view dataset information.

To see information for a dataset:

1.

Open the **Manage Datasets** page, by clicking the **Datasets** icon in the left navigation bar.



The **Manage Datasets** page opens, listing the datasets that have already been defined.

Manage Datasets

Datasets

☐ Select/Deselect All

☐ BCCA - OMEGA Study(not authorized)

☐ Cardiology Outcomes Analysis(not authorized)

☐ COO Task Force - Hypothesis 1(not authorized)

☐ Test Cardiology Claims(not authorized)

☐ USF Cardiology Example(not authorized)

- Click the name of the dataset you want to view.

The information screen for the dataset opens.

Dataset: Cardiology Outcomes Analysis

Name: Cardiology Outcomes Analysis

Description: Dataset contains echocardiogram information for analysis.

Export Target: CSV

Access Policy: Everyone can see everything

Status: Available

Available To: banalyst (Business Analyst)

Expires: NaN-0NaN-0NaN

Dataset Contents

[BCCA_LabResults \(0\)](#)
[Billing \(0\)](#)
[Cardiology \(6\)](#)
[Genomics \(0\)](#)
[BCCA_OncologyVCF \(0\)](#)
[BCCA_PatientHistory \(0\)](#)

3. Click **Close** to dismiss the screen.

Modify a Dataset

Modify information for a dataset on the **Manage Datasets** page.



Note: Only users with a role of PHEMI Administrator can modify a dataset definition.

To see information for a dataset:

- 1.

Open the **Manage Datasets** page, by clicking the **Datasets** icon in the left navigation bar.



The **Manage Datasets** page opens, listing the datasets that have already been defined.

Manage Datasets

Datasets

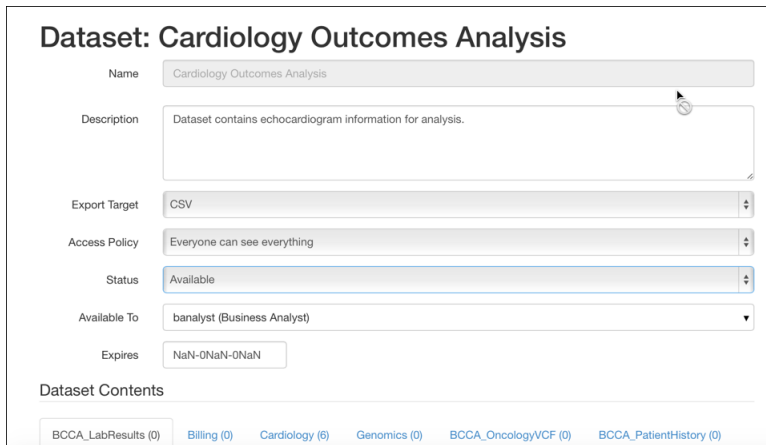
☐ Select/Deselect All

☐ BCCA - OMEGA Study(not authorized)
☐ Cardiology Outcomes Analysis(not authorized)
☐ COO Task Force - Hypothesis 1(not authorized)
☐ Test Cardiology Claims(not authorized)
☐ USF Cardiology Example(not authorized)

Execute Selected Datasets

Define New Dataset

2. Click the name of the dataset you want to view.
The information screen for the dataset opens.



Dataset: Cardiology Outcomes Analysis

Name: Cardiology Outcomes Analysis

Description: Dataset contains echocardiogram information for analysis.

Export Target: CSV

Access Policy: Everyone can see everything

Status: Available

Available To: banalyst (Business Analyst)

Expires: NaN-0NaN-0NaN

Dataset Contents

BCCA_LabResults (0) Billing (0) Cardiology (6) Genomics (0) BCCA_OncologyVCF (0) BCCA_PatientHistory (0)

3. Make your changes. *What do these fields mean? How do I choose fields for the dataset?*

Click **Save Dataset** to save your changes. The system confirms when the dataset has been successfully saved. Click **Close** to dismiss the dataset information screen.

Execute a Dataset

View the list of defined datasets on the **Manage Datasets** page.



Note: Only users with a role of Data Analyst can execute a dataset.

To see what datasets are available:

- 1.

Open the **Manage Datasets** page, by clicking the **Datasets** icon in the left navigation bar.



The **Manage Datasets** page opens, listing the defined datasets.



Manage Datasets

Datasets

☐ Select/Deselect All

☐ BCCA - OMEGA Study (not authorized)

☐ Cardiology Outcomes Analysis (not authorized)

☐ COO Task Force - Hypothesis 1 (not authorized)

☐ Test Cardiology Claims (not authorized)

☐ USF Cardiology Example (not authorized)

Execute Selected Datasets

Define New Dataset

2. Check the checkbox next to the datasets you want to execute, or check **Select/Deselect All** to select all the datasets.

You will only be able to select datasets that the PHEMI Administrator has designated as available to you in the dataset definition.

3. Click **Execute Selected Datasets**.

The dataset executes.

Delete a Dataset

Delete a dataset from the **Manage Datasets** page.



Note: Only users with a role of PHEMI Administrator can delete a dataset.

To delete a dataset:

1.

Open the **Manage Datasets** page, by clicking the **Datasets** icon in the left navigation bar.



The **Manage Datasets** page opens, listing the datasets that have already been defined.

2. Click the name of the dataset you want to delete.

The information screen for the dataset opens.

3. Click the **Delete Dataset** button. The system asks you to confirm permanent deletion. Click **Delete**.