Administrator Quick Start Guide

Contents

Introducing PHEMI Central

To be written

Intro blurb to segue into following overview content.

PHEMI Central Overview

To be written.

This topic will provide a basic introduction to and overview of the PHEMI Central system. It will include relevant content and diagrams similar to that in the Product Description currently being developed.

Data in the PHEMI System

To be written.

This topic will follow an item of data as it traverses the PHEMI system. It will include diagrams. It will start at the data source, through collect, through curate, and through consumption by users, machine destinations, applications, and so on.

Foundational Concepts

The PHEMI system makes use of some foundational concepts. You should familiarize yourself with the concepts you need to perform initial configuration.

User Authorizations

Authorizations are configurable properties assigned to system users. Authorizations are defined by the administrator in accordance with the organization's governance policies. Authorizations are combined with data visibilities to determine what permission a user has to interact with different data. For example, a user with Clinician authorization might be allowed to access all forms of health data, including confidential or identifiable information, while a user with Researcher authorization might be allowed to consume only with de-identified or nonidentified information.

Authorizations are used in access policies together with data visibilities that have been applied to data sources. The access policy matches the authorization against the data visibility to determine what action, if any, a user may take with respect to the data.

Some examples of possible user authorizations are as follows:

- C_LEVEL: The user is a C-Level individual (for example, CEO, COO, CIO, or CTO) with a privileged level of access. Individuals with C_LEVEL authorization, for example, might be permitted to read data with CONFIDENTIAL visibility.
- DOCTOR: A user with DOCTOR authorization might, for example, be permitted to read any information, including personally identifiable information or personal health information.
- ANALYST: A user with ANALYST information might be restricted to accessing de-identified or nonidentified data.

A user can be assigned multiple authorizations. Authorizations are defined as part of system configuration.

Note: Once defined, a user authorization may be neither edited nor deleted.

User Roles

User roles define what you are allowed to do within the PHEMI Central Management Console. Users can be assigned one of three roles.

Table 1: User Roles

Role	Purpose	PHEMI Central Management Console Access
Administrator	Configures access to the PHEMI application and to data.	• System configuration: password policy, dataset destinations, data retention behavior, data categories, data visibilities, and user authorizations • Configure data sources, including deploying DPFs • Create access policies • Monitor system metrics • Manage users • Perform system maintenance • Monitor audit logs
Privacy Officer	Responsible for governance policies that define the organization's approach for safeguarding data and assigning privileges to users.	The privacy officer has no functional ability within the PHEMI Central Management Console. The infleuence of privacy officer occurs before system configuration.
Data Analyst	Submits data for ingestion and consumes data.	• Manually ingest data • Build and execute datasets

Raw Data

In the PHEMI system, raw data items are files, objects, records, images, and so on that are submitted for ingestion into the system. Raw data is stored exactly as received, along with the metadata generated for it on ingestion.

Metadata

Metadata is information about a piece of data. In the PHEMI system, metadata is information about how a given piece of data is to be managed. When a piece of raw data is ingested into the PHEMI system, information from the connection (for example, the timestamp) together with information configured for the data source (for example, the data policy and retention information) is used to create metadata properties that are stored with the data. The PHEMI system also automatically indexes and catalogues all stored data, whether raw or derived; the indexes and catalogues can also be considered a kind of metadata.

Derived Data

Derived data is data that has been parsed, extracted, or otherwise enriched or processed by running a Data Processing Function (DPF) on stored raw data. The set of derived data items can be searched, queried, further processed, or exported from the system.

Digital Asset

Once raw data has had metadata applied, it is considered a digital asset. Derived data is also a digital asset.

Data Categories

High-level data categories help you classify the data that will be stored in the PHEMI system.

Each data category can include multiple data sources, data systems (such as different databases), or data collections. For example, an organization might have a category BILLING, which could include data from several different billing systems in the organization.

Data Visibilities

All raw data and derived data stored in the PHEMI system can be tagged with attributes that provide information about the data's sensitivity and the visibility it should have to different system users. These attributes are called data visibilities.

The visibilities you define for your data should reflect the sensitivity of the data as identified by your organization. The PHEMI system predefines three visibilities pertaining to privacy:

Table 2: System-Defined Data Visibilities

Visibility	Description
IDENTIFIED	Sensitive data that contains identifying information. Such data should be secured and available only to authorized personnel.
DE_IDENTIFIED	Sensitive data that has been transformed and masked to remove identifying information
NONIDENTIFIED	Data that does not contain sensitive or identifying data.

You can define additional data visibilities to suit your organization's needs. Possible examples are CONFIDENTIAL to identify data that is sensitive and requires a user to a certain level of access or certification to access, PII to identify Personally Identifiable Information, PHI to identify personal health information, or SECRET to identify especially sensitive material.

Note: Once defined, a data visibility may be neither edited nor deleted.

The data visibilities specified for a data source are referred to in access policies and matched against user authorizations to determine what actions, if any, a given user is allowed to perform on the data.

Any data visibility defined in the system can be used by a Data Processing Function (DPF) to tag data elements, or cells, derived by the DPF. For example, the Social Security Number extracted from a health record can be tagged CONFIDENTIAL if that data visibility has been defined in the system. This mechanism provides "cell-level" privacy protection.

Access Policies

An access policy is a set of logical rules that determines how users can consume data stored in the PHEMI system. Access policies can be optionally applied to data sources and datasets.

To create an access policy, you define one or more access rules. In each rule:

- The Subject specifies a set of the user authorizations.
- The Object specifies a set of data visibilities.
- The Action specifies the action to be taken if the rule is matched.

A rule is matched when the user making the request has an authorization matching at least one of those listed for Subject, and the data being requested has a visibility matching at least one of those listed for Object. When there is a match, the user may take the specified action(s) on the data.

If there is more than one access rule within a policy, the rules are related by OR logic.

The rules in a given access policy are intended to implement specific controls over a dataset. Depending on the number and kinds of datasets your organization works with, you may need just one access policy or multiple access policies.

If you have multiple access policies, it is possible for policies to conflict with one another and still represent a consistent governance policy, provided that each access policy is used to control different data sources or datasets. For example, one access policy may allow users with Researcher authority to read CONFIDENTIAL data while another access policy does not. This can be perfectly consistent, given the policies control different data.

Data Sources

In the PHEMI system, a data source defines a set of rules and policies for managing and governing data in the system, thereby controlling consumption and access to the data. A data source should be created for any collection of data to be stored in the system and managed by the same retention, legal, and governance rules.

In the PHEMI system, data source includes information about the following:

- The data policy that applies to this data
- The Data Processing Function (DPF) to be used to create derived data from this data
- A facility for manually ingesting data

No connection information is required in data source configuration. Data is "pushed" from the site that is submitting the data; any required connection information, such as timestamp and user credentials, is extracted from that communication.

Data Processing Functions

A Data Processing Function is an executable piece of code that supplies the instructions for parsing raw data (for example, a log message or medical report) into derived data (such as a temperature reading or blood glucose measurement). The output of a DPF is structured elements, which includes a type property (for example, INT or STRING) and can include other attributes (for example, SECRET or IDENTIFIABLE). A DPF is associated with a data source as part of data source configuration.

A DPF archive is the set of code that makes up a DPF. A DPF archive is delivered as a ZIP file archive. It consists of two parts: a manifest file and a code library. To associate a DPF with a data source, the DPF archive is ``registered`` with the data source by uploading the archive during data source configuration.

Datasets

A dataset is a specific view of data in the PHEMI system that is accessible to some limited set of users. The view can be across multiple data sources.

Data sets are typically created for a research or exploratory data use falling outside the bounds of governance and controls protecting the source data. Each dataset must be associated with an access polcy to ensure that it complies with the criteria that were established in the authorization to create it.

You can define any number of datasets. Defining a dataset has no impact on configuration for data sources or protections to data. Data Source configurations and protections to data. The only requirement is that an Access Policy must be associated with the dataset to ensure that it complies with the criteria that were established in the authorization to create it.

Dataset Destinations

Datasets can be exported to consuming database systems. You define each destination database as a separate dataset destination.

Before You Configure

Before you start configuring the PHEMI system, the system must be installed and you should have your organization's governance policy clearly defined.

System Installation

The PHEMI system may be shipped as an appliance, deployed on Amazon Web Services (AWS), or deployed on your organization's VMware environment. Regardless of the deployment type, a PHEMI representative will install and set up the base system of cluster nodes with a management node and will install the PHEMI software on the cluster.

The representative will create an administrator account for you on the PHEMI system, and will provide you with the user name and password for the account, together with the URL of the Management Console. At that point, you can log on to the PHEMI management console.

Governance Policy

To protect your data and the privacy of your data, you should have your organization's governance policy in place before configuring the PHEMI system. A governance policy is a coordinated approach to protecting data and assigning privileges to users.

There are three main aspects to a governance policy:

- Data visibility. Data can have different levels of visibility or sensitivity, and your organization may need to protect different data in different ways. You may want some data visible to every user but other data visible only to selected users. In the PHEMI system, these levels are called "data visibilities." *Tell me more about data visibilities*
- User authorizations. Different users may be required to interact with data differently. For example, clinicians might be allowed to access all data, while researchers might be allowed to read only non-identified or de-identified data. *Tell me more about user authorizations*
- Access policies. Your organization's governance policy will state how users with different authorizations can interact with data of different visibility. *Tell me more about access policies*

The data visibilities, user authorizations, and access policies defined in your governance policies will drive how you configure the PHEMI system. If you are uncertain as to the appropriate data visibilities, user authorizations, or access policies to define for your organization, consult with your Information Officer, Privacy Officer, or with someone in a comparable role.

Configuration Workflow

Since some configuration tasks must be completed before others can be performed, use this workflow to configure your PHEMI system.

In these first steps, order matters. You cannot create users until you have defined user authorizations. Likewise, you cannot create access policies without defining both user authorizations and data visibilities.

1. *Log on to the PHEMI Central Management Console*.
2. *Define user authorizations.*
3. *Create users.*
4. *Define data visibilities.*

Data categories must be added before you can define data sources. Access policies are optional for data source configuration. If you are using a Data Processing Function (DPF) to create derived data from raw data, you upload the DPF archive as part of data source configuration.

5. *Add data categories.*
6. *Create access policies.*
7. *Define data sources.*

Configuring datasets and dataset destinations is optional.

8. *Configure datasets.*
9. *Configure dataset destinations.*

Log On to the PHEMI Central Management Console

The PHEMI Central Management Console is web-based.

Note: Use either Mozilla Firefox or Google Chrome to access the PHEMI Central Management Console. Microsoft Internet Explorer is not supported.

The URL for PHEMI Central is configured during installation and setup of the PHEMI system. Your PHEMI representative will provide you with the URL.

To access the PHEMI Central Management Console:

1. In the address bar of the browser, enter the URL configured for the PHEMI Central Management Console.

   The PHEMI Central login screen appears.

2. Enter your user name and password. Click **Login**.
   The PHEMI Central Management Console launches.The Management Console initially opens on the Data Sources page. Subsequently, the system remembers the last page you viewed and opens on that page.

Define User Authorizations

Define user authorizations on the **System Configuration** page.

Provisional: needs to be updated from demo system.

Note: Once defined, a user authorization may be neither edited nor deleted.

To define user authorizations:

1. Open the **System Configuration** page, by clicking the **System Configuration** icon in the left navigation bar.

The **System Configuration** page opens on the **Password Policy** screen.

2. Click the **Authorizations** tab.

   The **Authorizations** screen opens.
3. Set the authorization information.

Option	Description

4. Save the authorization information by clicking the **Save** button.

Create a New User

Create a new user on the **Manage Users** page.

To create a new user:

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

   The **Manage Users** page lists all users defined in the system so far.
2. Click the **Create User** button.

   The **Create a New User** window opens.
3. Enter the user information.

Option	Description
Full Name	Mandatory. The user's full name.
User name	Mandatory. The user ID for this user. The **User Name** field autopopulates with the first initial and last name entered for as the user's full name. For example, if the user's full name is Jane Smith, the **User Name** field autopopulates with jsmith. The user ID can be edited after it has been autopopulated. IDs can be up to 16 characters long. Alphabetic and numeric values are permitted, as well as hyphen ("-") and underscore ("_"). Spaces are not permitted.
Phone Number	Optional. The user's phone number. You must configure this field if you want the system to send alerts to the user's phone.
Email	Mandatory. The user's email address. If you configure the system to send email alerts to the user, this is the email address that will be used.
Role	Mandatory. The user's system role. Together with the user authorization configured and the visibility set for data, the user role determines what access the user will have to data. Each user has exactly one role.*Tell me about user roles.*
Authorizations	Optional. The types of data the user is authorized to access. Use either the **Shift** key or the **Ctrl** key to make multiple selections. *Tell me about user authorizations.*
Password	Mandatory. The password policy (such as minimum and maximum length, whether a password can be reused, and so on) is set by your administrator in

Option	Description
	system configuration. Passwords must be confirmed by re-entering.

4. Save the user information by clicking the **Save User** button. The system confirms when the user has been successfully saved. Click **Close** to close the screen.

Define Data Visibilities

Define data visibilities on the **System Configuration** page.

Note: Once defined, a data visibility may be neither edited nor deleted.

To define the visibility for data stored in PHEMI Central:

1. Open the **System Configuration** page, by clicking the **System Configuration** icon in the left navigation bar.

   The **System Configuration** page opens on the **Password Policy** screen.
2. Click the **Manage Visibilities** tab.

   The **Manage Visibilities** screen opens.
3. Click the **Create Visibility** button. The **Create Visibility** screen opens.
4. Enter the visibility information.

Option	Description
Name	Enter a name for the data visibility; for example, "CONFIDENTIAL," "SECRET" or "PII".
Description	Enter a brief description to remind you or someone else of the intention of the visibility.

5. Save the visibility information by clicking the **Save** button.

Add a Data Category

Use the **Datasource Categories** screen of the **System Configuration** page to add a data category.

To add a data category:

1. Open the **System Configuration** page, by clicking the **System Configuration** icon in the left navigation bar.

   The **System Configuration** page opens on the **Password Policy** screen.
2. Click the **Datasource Categories** tab.

   The **Datasource Categories** screen opens.
3. Click the **New Category** button.

   The **Datasource Categories** screen expands to show the **Category Details** area.
4. Enter a name for the category.
5. Save the data category by clicking the **Save Category** button.

Create a New Access Policy

Create a new access policy on the **Access Policy Builder** page.

Before you can define an access policy you must configure the following:

- User authorities
- Data visibilities
- Environments (access networks)

To create a new access policy, define one or more access rules:

1. Open the **Access Policy Builder** page, by clicking the **Access Policy Builder** icon in the left navigation bar.

   The **Access Policy Builder** page opens.

2. Click the **New Policy** button.

   The form for the new access policy opens, with Rule 1 ready for you to edit.

3. Enter the access rule information.

Option	Description
Subject	Mandatory. The user authorizations allowed to perform the action on the data. User authorizations are configured for the system by the administrator.
Action	Mandatory. The action(s) an authorized user may take on the data. Supported actions are as follows: • Read. The user may view the data. • Export. The user may export the data to a destination, such as a SAP system.
Object	Mandatory. The data visibilities authorized users are allowed to access. Data visibilities are configured by the administrator.

4. Add another rule by clicking the **Add Rule** button. Or, save the access policy by clicking the **Save Access Policy** button. The system confirms when the access policy has been successfully saved.

Define a Data Source

Define a new data source on the**Data Sources** page.

Before defining a data source, you must configure the following:

- Data visibilities
- Users
- Data categories

Optionally, you can configure access policies.

To define a new data source:

1. Open the **Data Sources** page, by clicking the **Data Sources** icon. in the left navigation bar.

   The **Data Sources** page opens showing all defined data categories.

2. Click the **New Data Source** button.

The **New Data Source** screen opens.

3. Describe the source.

Option	Description
Name	Mandatory. A descriptive name for the data source. Numbers, letters, spaces, and special characters are supported. **Note:** Once you save a data source, the name cannot be edited. To change the name, you must delete the data source and reconfigure it with the correct name.
Source Category	Mandatory. The data category of the data source. Choose from the drop-down list of defined data categories.
Institutional Owner	Mandatory. The individual responsible overall for data stored in the PHEMI system. Users with any role can be an institutional owner. Choose from the drop-down list of users.
Privacy Officer	Mandatory. The individual responsible for approving access policies. Only users with a role of Privacy Officer can be selected for this field. Choose from the drop-down list of eligible users.
Source Owner	Mandatory. The individual responsible for approving dataset requests involving this data source. Users with any role can be an institutional owner. Choose from the drop-down list of users.
Document Format	Optional. The kind of document that will be submitted by this data source. Examples are Microsoft Word, Excel, or JSON.
Definition	Optional. A brief description for the documents that will be submitted by this data source.
Notes	Optional. Any additional notes for the data source.
Data Sharing Agreement	Optional. A file attachment recording the data. Only

4. Save the data source information by clicking the **Save Data Source** button. The system confirms when the data source has been successfully saved.

Define a New Dataset

Create a Dataset Destination

Define a new dataset destination on the **Dataset Destinations** screen of the **System Configuration** page.

To create a new dataset destination:

1. Open the **System Configuration** page, by clicking the **System Configuration** icon. in the left navigation bar.

   The **System Configuration** page opens on the **Password Policy** screen.

2. Click the **Dataset Destinations** tab.

   The **Dataset Destinations** screen opens.

3. To the right of the **Choose Destination to Edit** field, click the **New Destination** button.

   The **Destination Details** screen opens.

4. Enter the destination details.

Option	Description
Name	Mandatory. The name of the destination.
Type	Mandatory. The destination type. Supported values are as follows: • MySQL. The destination is a MySQL database. • HANA. The destination is a SAP HANA database.
Host	Mandatory. The IP address of the destination, in dotted decimal format.
Port	Mandatory. The port on the destination to send data to.
Database Name	Mandatory. The name of the destination database.
Schema	Mandatory. The schema being used in the destination database.
User Name	Mandatory. The user name to be used to log on to the database.
Password	Mandatory. The password to be used to log on to the database.
Confirm Password	Mandatory. Re-enter the password to ensure it is correct.

5. Click the **Save Destination** button to save the new destination.