PMAP Data Catalog User's Guide

1. Introduction

The PMAP Data Catalog contains information about the data that are available in the Precision Medicine Platform. The Data Catalog does not show the actual patient data. Rather, it provides information about the available data to guide subsequent requests for that data.

When the data catalog is first opened, it displays a list of the available databases sorted alphabetically by name. These databases are:

- camp – a deidentified set of data for patients with asthma. This dataset is typically used for training purposes to become familiar with working with data in the PMAP analytic environment.

- derived – a dataset of data primarily from the Epic medical record. This database contains the most frequently requested data elements, organizing the data in a way that simplifies data projections, reducing the need for researchers to link across multiple tables when they receive the projected data. The majority of PMAP data requests will utilize data from this database.

- dicom – a dataset about the imaging studies and series available in PMAP, describing information associated with the images such as date, time, description, and modality.

- edw – this dataset comes from the Epic Data Warehouse (EDW). At the current time this database only contains a small subset of data in the EDW.

- epic – this dataset contains information about the most frequently requested data from the Epic Clarity database. Because the data catalog exposes Epic Clarity table name and column names which are the intellectual property of the Epic Corporation, users of this data catalog must agree to protect that intellectual property by not posting information about tables names or columns in any forum that could make the table names or column names discoverable outside of Johns Hopkins.

- open_specimen – this dataset contains information about biospecimens that are tracked in the OpenSpecimen database. These are specimens that have been linked to a patient in the Epic medical record.

There are three levels of data elements described in the PMAP Data Catalog. These are:

Database – A collection of data from a specific data source or collected for a specific purpose. A database is composed of one or more tables. Click on a database to see the tables in that database.

Table – A subset of data in the database with some relationship in common. For example, in the derived database, the diagnosis table contains information about patient diagnoses. A Table contains one or more Columns. Click on a Table to see its associated columns.

Columns – A column provides information about a specific data element, such as a description of what that data element represents, the data type, and minimum and maximum values.

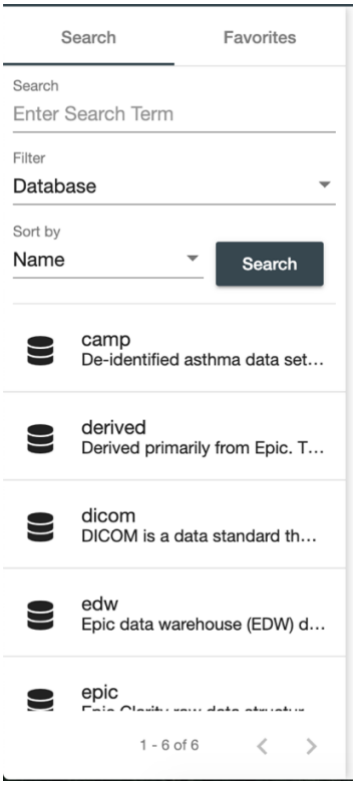
The Data Catalog has three major sections, the header bar, the navigation bar, and the catalog details.

2. The Header Bar

The horizontal header bar is at the top of the screen. Clicking on the "i" in the circle on the left side of the header bar will launch information about the PMAP Data Catalog. On the right side of the header bar we find the user id of the currently logged in user followed by a downward facing arrow. Clicking on this area will allow all users to either go back to the home screen that is seen with the Data Catalog is first launched, or to log out of the Data Catalog application. Users with special roles in the Data Catalog, such as metadata editors, will have additional administrative functions available via this menu.

3. The Search Bar

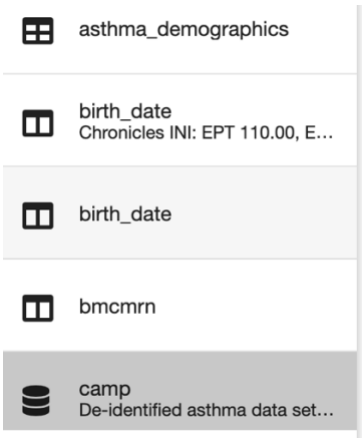
When the Data Catalog is initially launched, the user sees a list of all of the databases in the Data Catalog sorted alphabetically, as shown in the image below.



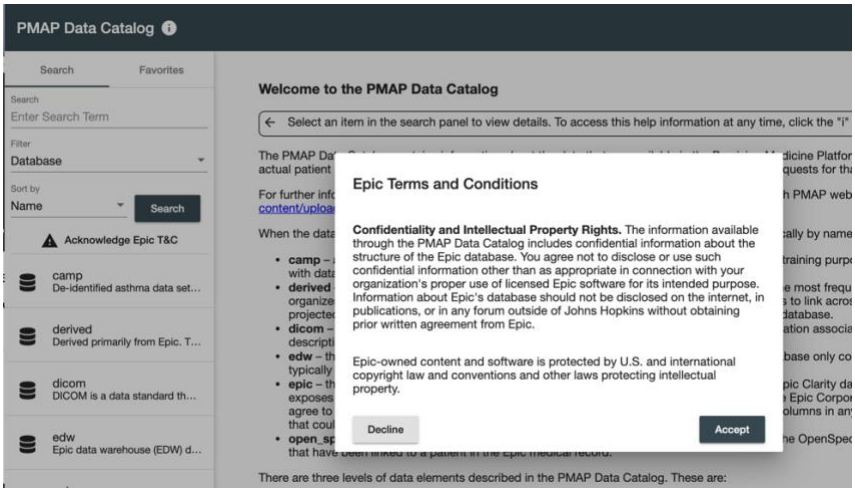
If the user clicks on one of the databases (e.g. camp or derived) they will be able to learn more about the contents of that database, as described in section 4 of this document.

Users can search the contents of the Data Catalog by typing a word or phrase into the area that says "Enter Search Term" and then pressing the Search button. After pressing the Search button, the users will see results which match the search term in the area below the Search button.

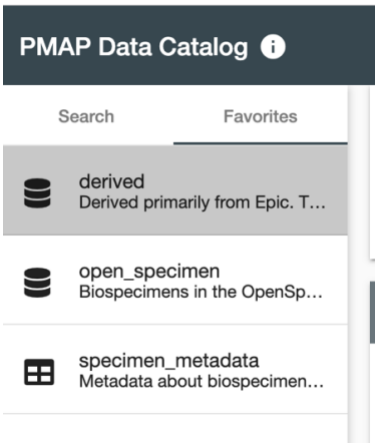
The next field is the Filter field. The Filter field controls which catalog entries are searched. The default Filter is Database. With the Database filter selected, the search will scan the database definitions but will ignore the contents of tables and columns. If the user clicks on the downward arrow at the right side of the Filter area, they will see checkboxes which allow them to select the scope of what should be searched. Available options are Database, Table, and Column. More than one option can be selected. If Database, Table, and Column are all checked, then all elements in the catalog will be searched.



Search Results: The image on the left shows results for a query that was run with the Filter set to Database, Table, and Column. The icons that precede each result indicate the type of element that matched the search. Tables have an icon that looks like a little spreadsheet with rows and columns. Columns have an icon with two columns. You will notice that there are two columns named "birth_date". This is because "birth_date" can be found in two different locations in the Data Catalog. Databases have an icon that looks like a cylinder. Search results can be sorted by the "Sort By field" in the Navigation Bar. The default search is an ascending alphabetic search, however, by clicking on the downward arrow the user can select other ways to sort the results of a search.



Epic Terms and Conditions: Users who already have an Epic account will see a prompt to acknowledge the Epic T&C. In order to see information about the Epic tables and columns, you must click on the "Acknowledge Epic T&C" field right below the Search button, and then read, agree to, and accept the terms and conditions. This must only be done one time. Users who do not have Epic access will not be able to see Epic tables.



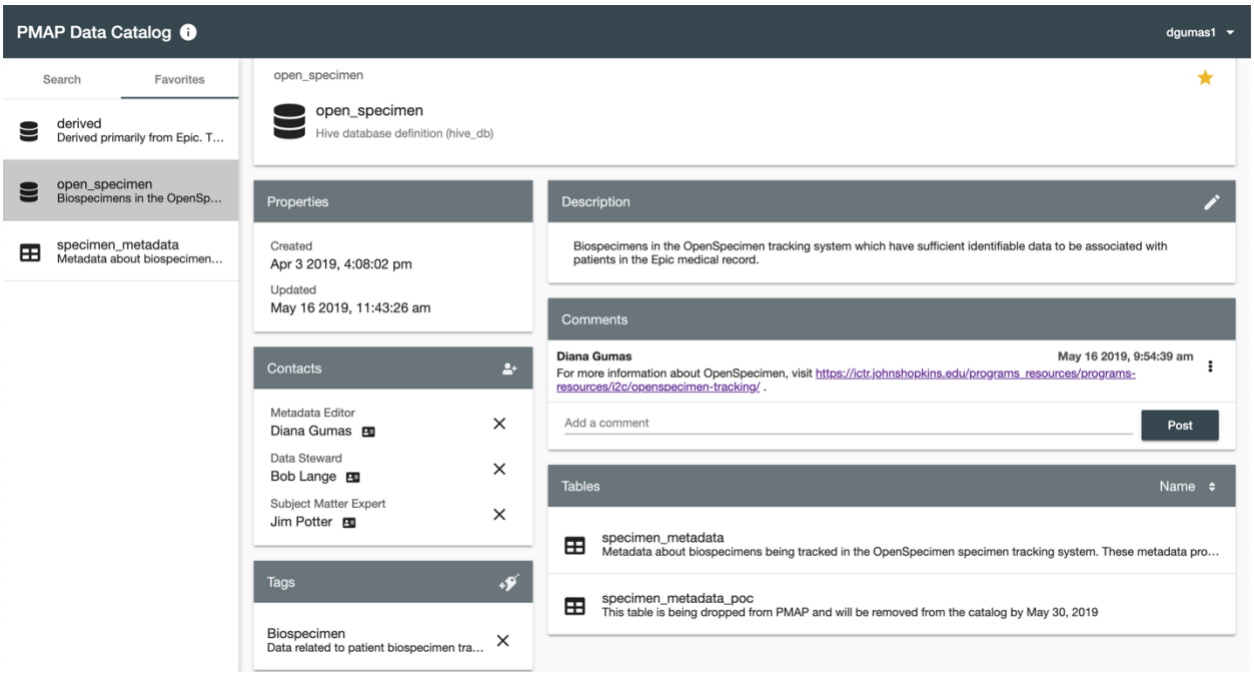
The Favorites Tab: The default tab in the Navigation Bar is the Search tab. However, there is another tab available which is the Favorites tab. By clicking on the Favorites tab the user can quickly navigate to databases, tables, or columns which they have previously selected as their favorite items in the Data Catalog. Items can be selected as a favorite item by clicking on a star on that item's main page, as will be described in the next section of this user guide.

4. Catalog Details

When an item is selected in the navigation bar, details about the selected Database, Table, or Column are displayed in the Catalog Details section. The Catalog Details section is located to the right of the navigation bar, right below the Header Bar.

4.1. Database Details

The following image shows what the user will see when they click on a database in the navigation bar.



The name of the database is listed at the top of the screen. The cylinder icon indicates that this is a database. On the right side of the screen there is a star which can be clicked in order to make this database a favorite so that it appears in the Favorites tab of the Navigation Bar. When selected as a favorite, the star will be yellow.

The Properties field displays information about when the database was created and last updated. The Contacts field displays information about people who can provide more information about the database. By clicking on the black icon of a business card to the right of each name, the user can see email and phone information for each contact. The Metadata Editor is the person who entered information into the Data Catalog. The Data Steward is the technical person who manages the source database. The Subject Matter Expert is the person who can best answer the “meaning” and use of the database.

The Tags field is used to facilitate searching. New tags can only be added by Metadata Editors.
The Description field provides a high-level description of the database and can only be modified by Metadata Editors.

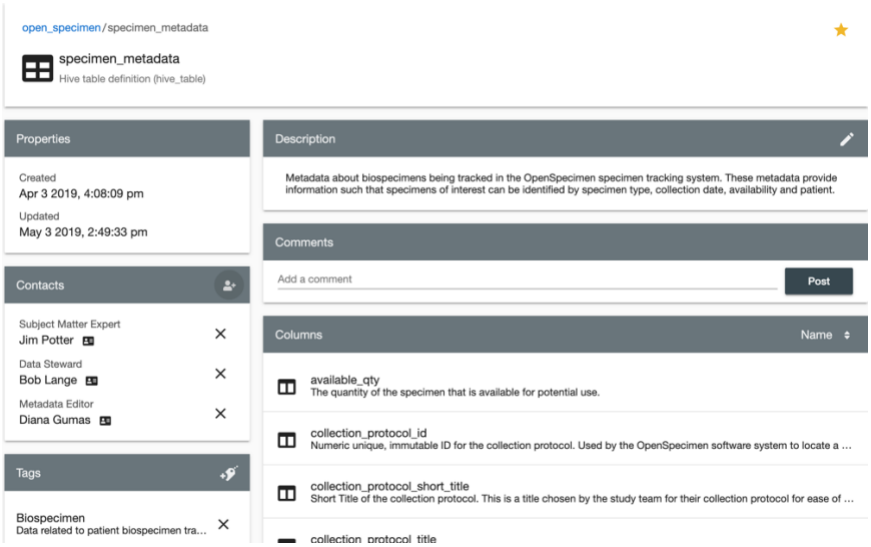
The Comments field contains additional information about the database. Any user of the Data Catalog can post a comment. We anticipate that Comments from the community will improve the quality of the information in the Data Catalog.

The Tables field lists all of the tables that comprise the database. The user can click on a table in order to see details about that table. The default sort order of the tables is ascending alphabetical but clicking on the double arrow on the right side of the Tables gray bar will provide other sort options. This is especially useful when a database has more tables than can be seen without scrolling.

An important note about the epic database: In order to see the tables associated with the epic database you must have an active Epic login and also must have accepted the Terms and Conditions

4.2. Table Details

The Table Details screen, shown below, provides information about the selected table. A table can be selected either by searching and choosing it from the navigation bar, or by clicking on the table in the Database Details screen. The sections of the Table Details screen are nearly identical to the Database Details screen described in section 4.1. The only difference is in the bottom right of the screen, which displays the columns which comprise the table. The user can click on a column in the table in order to see the details about that column.



4.3. Column Details

The Column Details Screen is nearly identical to the layout of the Database Details and the Column Details screens. What is unique about this screen is the column metadata section in the bottom right portion of the screen. This describes the data types, date ranges, and value ranges of this data element.