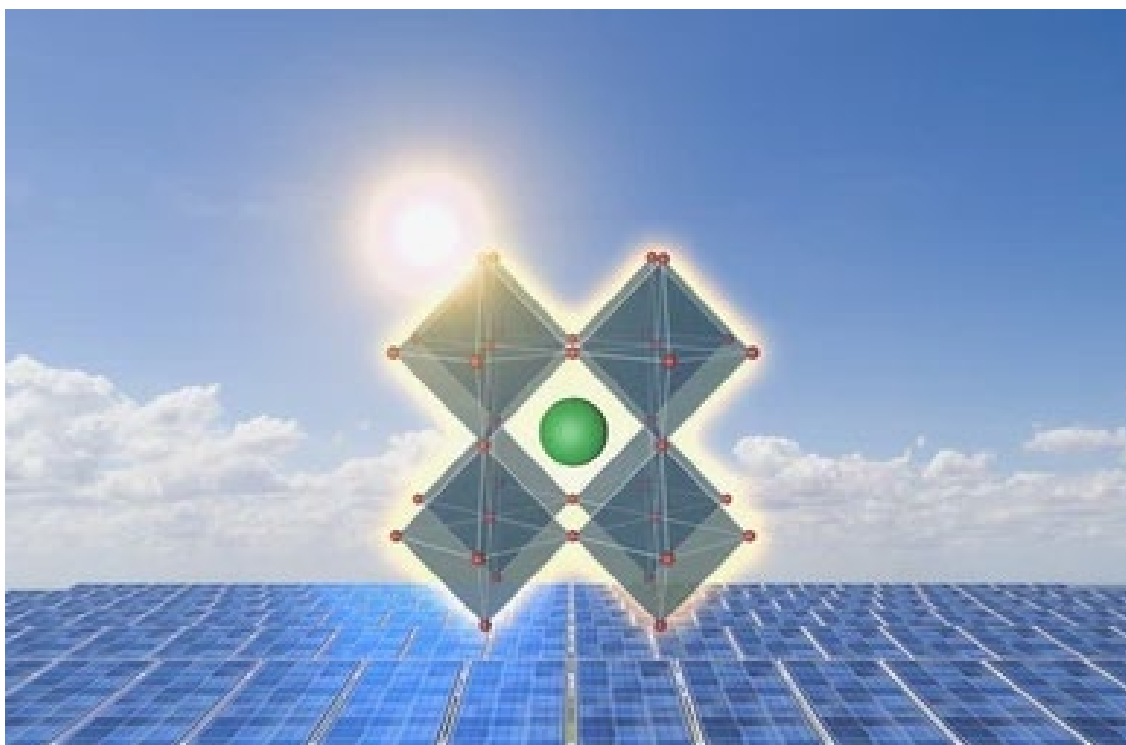


Using the Perovskites Data Hub

Supporting the US-MAP and PACT Consortia



U.S. DEPARTMENT OF
ENERGY

Energy Efficiency &
Renewable Energy



SOLAR ENERGY
TECHNOLOGIES OFFICE
U.S. Department Of Energy

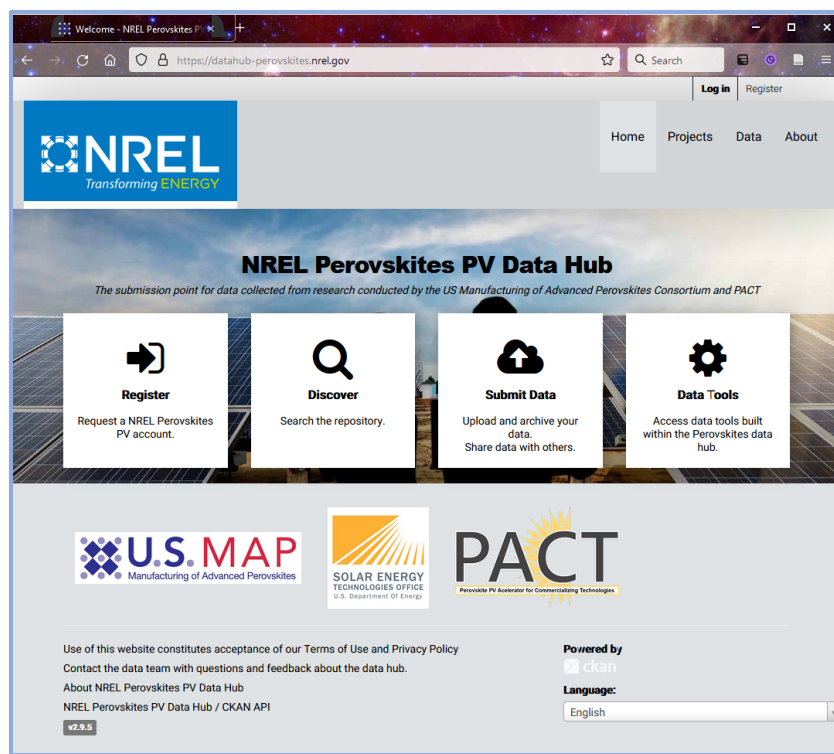
Table of Contents

Using the Perovskites Data Hub	1
Overview	3
Registration and Security	4
Projects	5
Datasets	5
Files and Resources	6
For All Researchers	7
Adding Data	7
Adding Multiple Files in One Upload (Bulk Upload)	8
Check Your Upload	10
Working inside the Dataset	10
Adding Data to an Existing Dataset	11
Adding Multiple Data Files to an Existing Dataset	12
Deleting a File or Resource	13
Viewing Data	14
Accessing the Data through the API	16
For PIs and System Administrators	18
Adding new members to a project	18
Changing permissions for a member of a project	19
Deleting a member of a project	20
Problems	21

Overview

The Perovskites Data Hub is a platform for consortium members and partners to share data and ideas. Researchers are encouraged to place their data within this hub, where it can be protected and distributed as needed. Providing data to the hub can increase communication efficiency between all parties and create a seamless environment for eventually releasing data from publicly funded research. Data can be compartmentalized and secured by project or scope and is able to store a wide variety of data types and files. The hub allows for searching the metadata and data of all resources stored within, providing a method to support discovery.

The hub is built around the CKAN software framework but has been extended through plugins and code rewrites beyond the original distribution package. The focus of the data hub is to provide the users with a system that can provide the efficiency and security for collaborative data sharing.



The architecture of the software can be divided into three main areas:

- Registration and Security
- Projects – Projects provide a means to encapsulate all the data associated with a topic or awarded research thrust. Each project can contain multiple sub projects and datasets.
- Submit Data – Immediately opens up the “Create Dataset” interface.
- Data Tools – Opens up to the list of available data tools. Links could take you to source code repos or interactive websites.

There are also many smaller interconnected software elements that support everything from API information and system documentation to visualization and search capabilities.

Registration and Security

The Perovskites data hub uses an extended security protocol, allowing for more sensitive data to be housed. The login in requires Multi-Factor authorization and draws from a special user pool defined for this application.

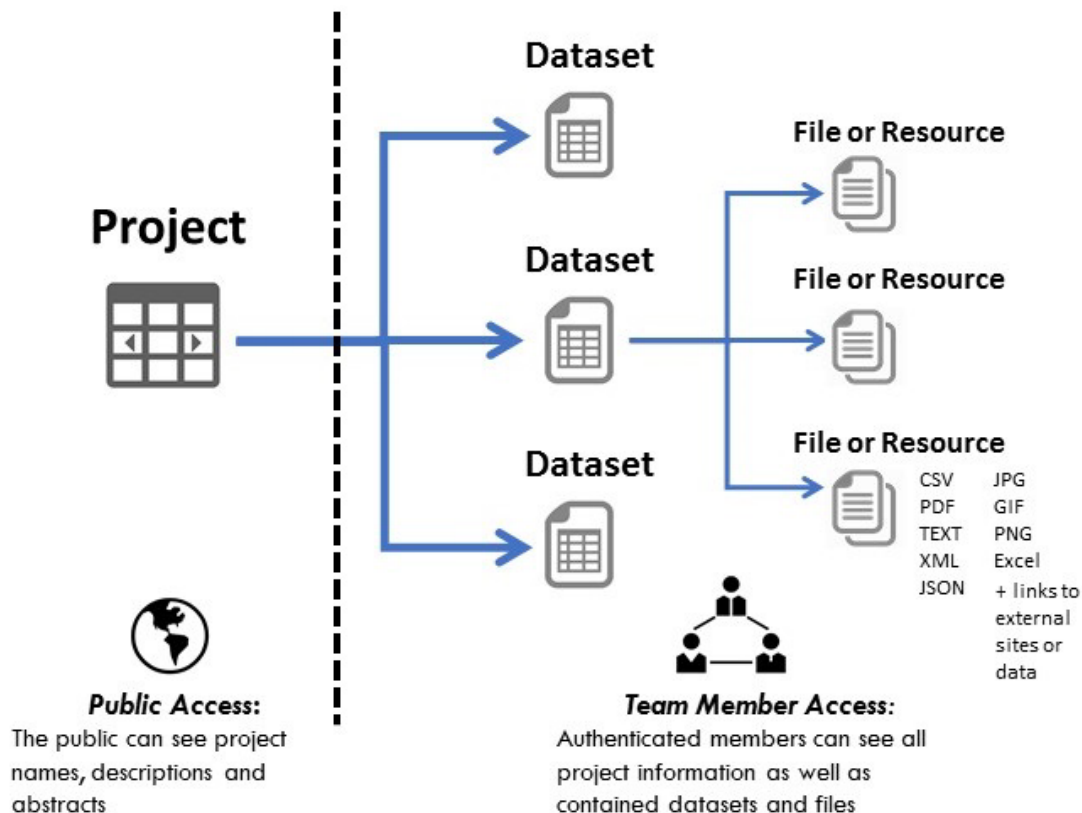
You must be a registered user of the data hub and consortium member to be able to access any files that are not deemed “Public”. The registration process is simple and begins by clicking the large “Registration” button on the main page (see image above), which will take user to the registration page.

1. Enter the required information on the registration page and click **Create Account**. Lab users should use their lab username and email address.
2. The researcher needs to email the administrator (emnadmin@nrel.gov) with the following details:
 - a. The Institution you work for.
 - b. Username
 - c. The list of projects you need access to. You can review the list of current projects without registration by clicking on the Projects tab on the menu bar at the top of the page.

Once given access to the site, you can request additional sub-projects/folders by contacting the PI for a project or the site administrator. The PI’s name is found on the Project Overview tab in the **Recipient** field. The project PI or designated representative can dynamically add or remove other researchers from the project’s access list.

By default, all datasets created, and files uploaded are designated as private, indicating only members of that project can access the datasets. To create public datasets, the user uploading the data will need to contact the site administrator to request public release. As part of the release process, the administrator will examine the datasets for completeness and that the data meets any criteria for release as set down by the consortium or DOE public data release requirements.

Unregistered or public users can see basic information about the project (abstract, principal investigator, awardee (if funded), support labs, etc.) and can view and download any designated *public data*. However, any private datasets cannot be seen or accessed by the public. Private datasets do not even show up in dataset counts within a project being viewed by a public member.



Projects

These are high level containers that are used to organize work within the data hub. Typically, these will be focused on research funded by the Perovskites consortium. However, there is often a great deal of associable current and historical data that can be added to the data hub, by request and agreement with the consortium directors.

All projects and sub-projects are created by a system administrator. In the future PIs may be able to create sub-projects as needed. Contact the [administrator](#) for any projects or sub-projects/folders you need created or adjusted.

A project can contain any number of datasets and sub-projects/folders. Sub-projects allow for additional granularity for storing data within a project. Also, sub-projects allow for restricting data within a project to only a subset of normal project members. Since security is carried by the project and sub-project membership, different authorization schemes can be established, even within a project with competing partners.

Datasets

A dataset is a low-level container or folder meant to house the files and links within the project. They are used to organize data into logical chunks that share a common source and metadata. Each dataset can contain any number of files or resource links.

By default, each dataset is established as private, giving only project members access to the data. A dataset can be made public, by notifying the site administrator and requesting public release. If you have a current dataset that needs to be made public immediately on upload, please contact the [administrator](#) and we will help you. As part of the release process, the administrator will examine the datasets for validity and completeness.

Files and Resources

All data uploaded to the hub can be either a file or a link. Collectively, these two are called resources within the data hub. The data hub is capable of housing any type of file format. We have limited the upload size to 5GB for any single file. If the researcher needs larger files stored, please contact the [administrator](#) so we may set up an authorized upload. Users should understand that moving and uploading large data files through the internet (even with high bandwidth) can take a long time and they should be prepared to shepherd the process until the data is verified loaded.

Each file uploaded or link established will have a unique set of metadata describing details about the instrument or source of the data, who created or controls the data, and any particulars with how the data was captured. Each user uploading data into the hub will need to fill out this associated metadata. In a case where the metadata or data source does not match to the pre-established ones, then the user is encouraged to upload a metadata file along with the data, to express all of the contextual information needed to understand the data.

We encourage all research consortia members using the Perovskites data hub to see the additional document on Data Standards and Best Practices (currently stored in the Data Standards project on the Data Hub). This document can help you plan before uploading the data on what might be the best formats and data structure to support the Perovskites consortium and your research. It will also advise you on what information might be needed before having to complete the metadata forms.

For All Researchers

Adding Data

The key to the data hub is uploading and archiving data that should be shared with the project members and eventually to the public. The process may require the creation of a new dataset or adding new files to an existing project's dataset. During the creation of the dataset and / or adding a resource or file, in some cases you will be prompted to provide additional information (metadata) that can facilitate understanding, searching, and organizing the data.

1. Click on the Project that will store the data

2. Click on **Add Dataset**, left panel

3. Complete needed fields on form

4. Click on **Add Data**

Verify that the set *Project* is correct. You could place the dataset in another project, but make sure you have access to that target project or the upload might fail.

For Tags use single words if possible. You need to hit **Return** after each tag for it to appear.

Selecting certain **Data Source Types** could open additional metadata fields that will need to be completed. Some data might not match the type and in that case leave this field blank.

5. Choose to upload a single file or create a link to outside data. If multiple files, see Bulk Upload below

Clicking the **Upload** button will open a dialog box for you to choose a file from your computer to upload. The **Link** button will ask you to provide a URL.

6. Complete needed Fields on the form

Depending on what instrument or source has generated the data you are uploading, you will be prompted to provide additional metadata. Some of this metadata will be required and other fields optional.

hydrophilic, ... / test_bulk_upload / Add New Resource

2 Add data

Upload Link

Name:
eg. January 2011 Gold Prices

Description:
Some useful notes about the data

You can use [markdown formatting](#) here

Format:
eg. CSV, XML or JSON

This will be guessed automatically. Leave blank if you wish

Resource Metadata

Data Quality:
-- select an option --

Data Type:
-- select an option --

Data Tool Associated With This File:
-- select an option --

Data Tools: Your data must be set up specific to the data tool in order for this metadata to enable a visualization of the data you are uploading. Please visit [User Resources](#) for a detailed description of how each data tool works and how to set up your data file to enable the visualization.

Resource Comments:
Resource Comments

Previous Save & add another Switch to bulk add Finish

What's a resource?

7. Click **Save & add another** to open a new form for another resource or click **Finish** to complete the upload.

8. Dataset creation and resource upload complete

Once everything has been completed using the Finish button, a page for the dataset will appear showing all resources currently within it.

Adding Multiple Files in One Upload (Bulk Upload)

The data Hub can add multiple files all at once to a new dataset or append multiple files to an existing data set. The researcher does need to be certain that all files being uploaded are logically relatable to each other through their dataset base metadata. To perform the bulk upload the researcher must first create the new dataset (this can be done to an existing dataset, see below) Following steps 1-4 in the Adding Data section ([above](#)).

You do not need to fill out any of the base metadata on this page if you are going to Bulk Upload. However, you will need to go back after upload and add any important metadata that should be associated with certain data. This can be through the metadata interface or by uploading a metadata file.

One method is to upload a CSV containing all the metadata values for the data files when you are doing the Bulk Upload

5. Click **Switch to bulk add**

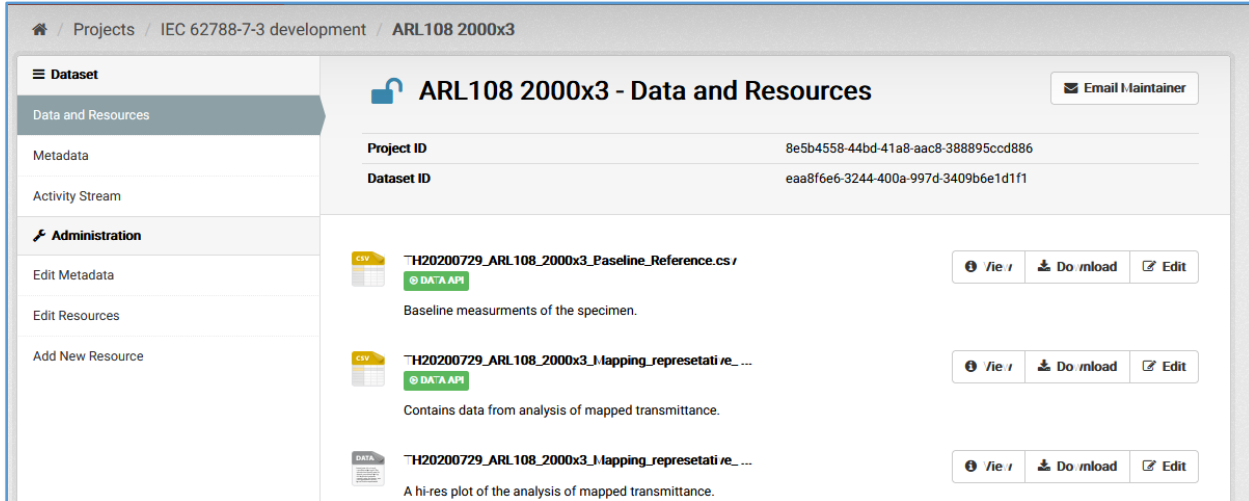
6. Click **Add Files**

7. Drag and Drop files from File Dialog to Bulk Upload page

8. Click **Finish** to complete bulk upload process

Check Your Upload

Once the file(s) have been uploaded you can now check the data and validate that the files have been processed and are available. You can also edit the file summary text (just below file name) and add or annotate the data by editing the resource metadata.



Working inside the Dataset

The navigation buttons on the left panel:

Dataset:

Metadata – View the dataset metadata.

Activity Stream – See activity concerning the data set. Note that this does not include information about downloads.

Administration: Only the editor or admin project members can perform any of these tasks.

Edit Metadata - Edit the dataset's current metadata.

Edit Resources - Opens the resource list for the dataset. From there you can add a new resource and edit the order they appear in the dataset list.

Add New Resource - Opens the resource upload page (as above).

Resource buttons:

View - If it is a “viewable” resource within the data hub, it will be displayed along with the associated metadata. Currently CSV, TXT, most picture formats, XML, and JSON are directly viewable in the hub. Link Resources will open a new tab within your browser to display the web page. Non-viewable resources will not be displayed but their metadata will be.

Download – Will download the resource file through your browser to your local computer.

Edit – Allows you to edit the metadata associated with that file or resource.

Data API (green) – Provides code samples and information for this file to be accessed via API.

Additional information and actions

Email Maintainer (Button top right) – Opens up an email to the researcher listed as the data owner.

Project ID - This is a UUID for this project. It can be used as part of the API access to the project

Dataset ID - This is a UUID for this specific dataset and can be used to access the dataset files via API.

Adding Data to an Existing Dataset

Within some working projects, depending on how the datasets are being used, you may need to continue adding data to an existing set. *Example:* A project could have a dataset for all XRD measurements. The dataset metadata could be the *Lab and Experimental Data Source Type* and set to cover all contingencies. Each data file will cover separate measurements, so these can show changing measurement conditions.

1. Click on the Project that will store the data

2. Click on **Datasets**, left panel

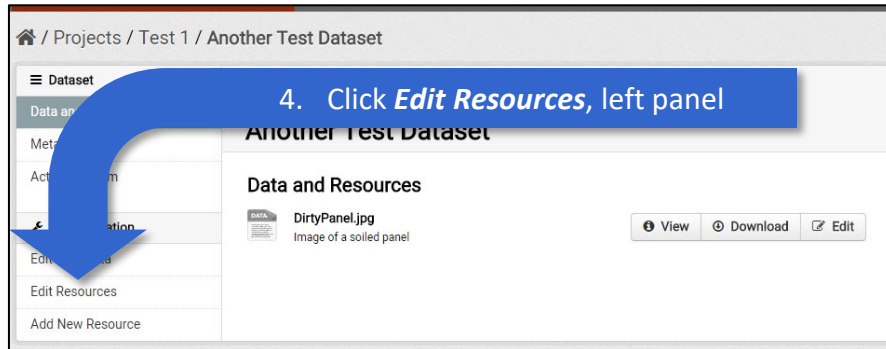
3. Click on the dataset to add data to

4. Click **Add New Resource**, left panel

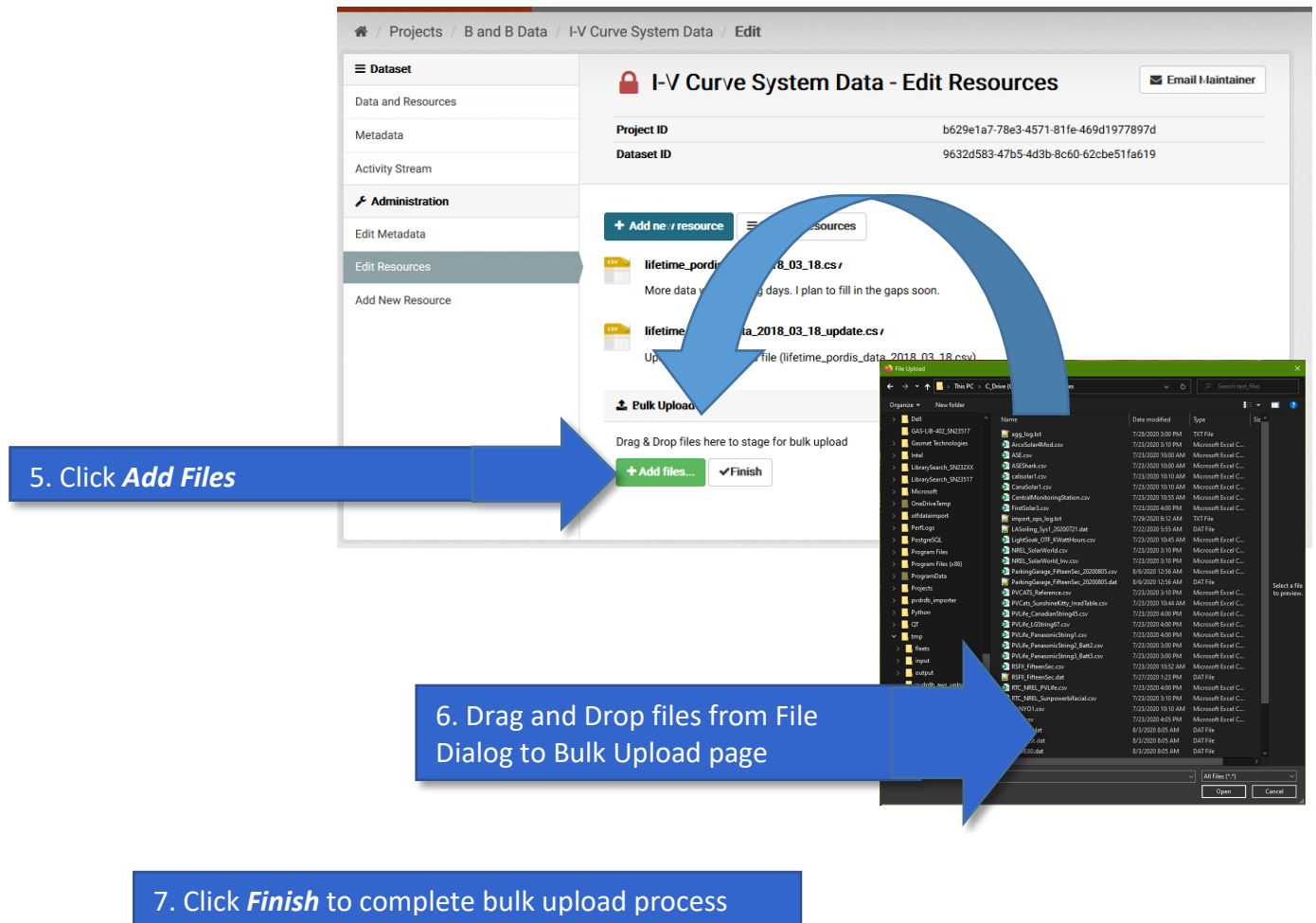
From that point follow the directions in the [previous section](#) from number 5.

Adding Multiple Data Files to an Existing Dataset

The Bulk Upload feature can also be activated to add multiple files in a single upload to an existing dataset. To access that feature for an existing dataset, follow steps 1-3 [above](#).



The Edit Resources page will open. Scroll to the bottom and you will find the Bulk Upload area.



Deleting a File or Resource

There will be times when a project editor or admin might need to delete a file or resource that has been added to a dataset. Perhaps you have uploaded the wrong file or noticed a mistake and need to reload the file. If you need assistance with this process, the system administrator can advise and assist. Contact [the administrator](#).

Starting from the Dataset's page

1. Click **Edit Resources**, left panel

2. Click on the target resource's name

3. Click **Delete**

4. A pop up dialog box will then appear and ask you to confirm deleting this file

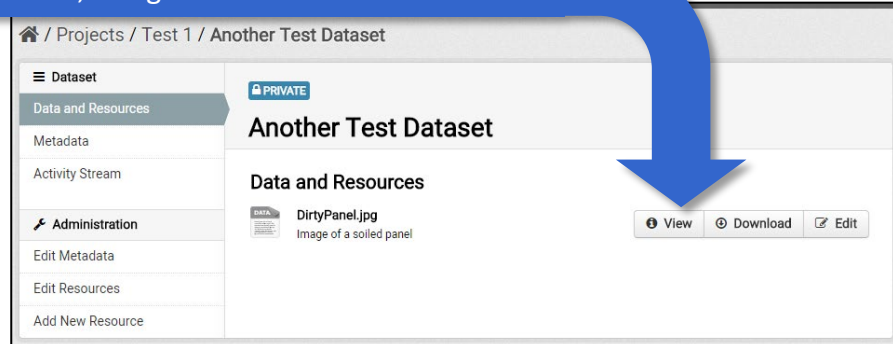
After you confirm deleting the file you will be brought back to the dataset's resource list page.

Viewing Data

In many cases the data hub has the ability to render the data file for viewing directly in the browser. This ability to view data can also be extended through custom coded plugins. To utilize the basic display capability of the system, do the following:

Starting from the Dataset's page

1. Click **View**, to right of resource name.



2. View page opens with the format's display

If the format has been predefined it will open that within the viewing engine. Additional view types can be created as plugins and tied into this same system.

On the right across from the resource name are 2-3 buttons:

Manage – This allows you to change the metadata for the resource.

Download – This will download the file through your browser to your local computer.

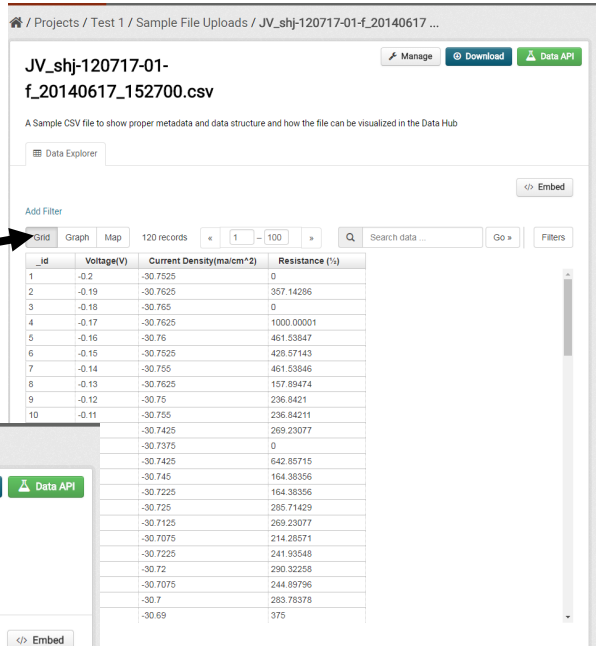
Data API (not shown) – This green button will open another dialog box listing the code snippets for accessing this data file programmatically ([see below](#)).



With CSV files the data Explorer plugin will allow for the data to be viewed in three possible ways: Table, Graph, or Map. To utilize the Graph or Map function will require interaction with the user to display the data. The map function requires there to be either Latitude and Longitude or GeoJSON as a column(s). To Graph, the data be columnar in format and the user will choose which data to plot on each of the axis. There can also be more than two traces displayed. Note that the CSVs should be in a simple format: one row as a header and data in the columns starting in row 2.

CSV Table: This is the default view of any CSV file. The headers from the file are also brought in. Each column is adjustable in size and can be filtered ascending or descending.

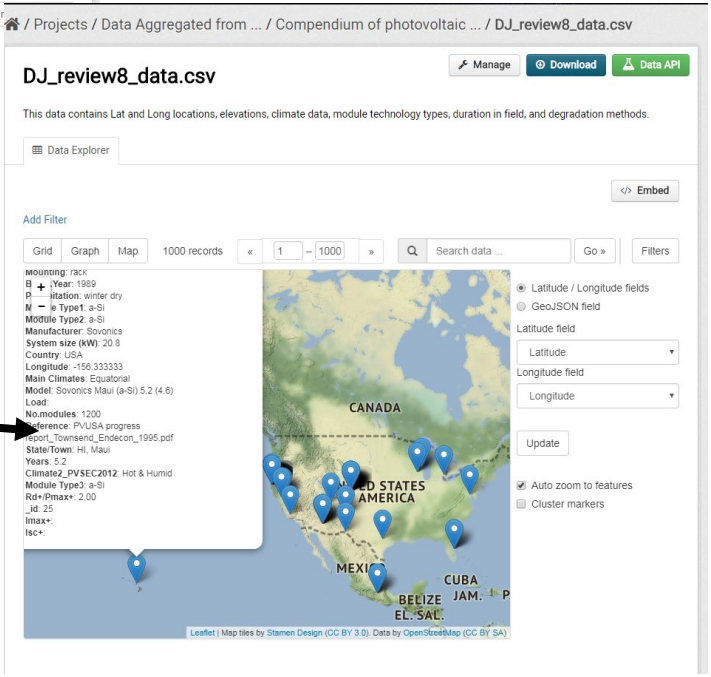
To change the display type for the CSV, Click the buttons just above and to the left of the table



CSV Plot (Graph): The user needs to connect the columns in the table to the axis. This interface is on the right side.

CSV Map: The Data Explorer will automatically detect if the file has Latitude and Longitude header names and use those by default for the geographic coordinates.

Clicking any of the map markers will display all the data for the record associated with that geographic position.



Accessing the Data through the API

Data archived within the data hub can be retrieved with the application programmatic Interface (API). Secured non-public data will still need to be accessed with credentials, but this can be passed as part of the API. Using an API can be useful if you need to upload, access, or download several datasets on a regular basis especially if they need to be processed through a software pipeline or tool set (e.g. Mathematica, Igor, Origin, etc.). The Perovskites data hub uses the native CKAN API which is built on a RESTful interface.

The first step to being able to access secure data through the API is to get your programmatic credentials. If you are accessing public data, the access key would not be needed.

The image shows a two-step process for accessing data through the API. Step 1: A screenshot of the NREL Perovskites PV Data Hub homepage. A blue arrow points to the 'Data' link in the top navigation bar. A blue callout box contains the text: "1. Click on name at the top of the page". Step 2: A screenshot of the user profile page for 'Mr. Cool Cat'. A blue arrow points to the 'API Key' field in the bottom left corner, which is circled in red. A blue callout box contains the text: "2. API Key code is on bottom left of your User page. Copy to clipboard or write it down!". The API key value is: 55053878-2f25-3878-8cf3-96012f25505.

The next step is to locate the data file you wish to access programmatically. Locate the dataset and data file you want to access. Click on the “View” for that data file (see above). You will then see the data page

1. Click on *Data API* button

2. Look through specific example codes for access to data file

3. Use the example code as a template in your favorite application

The API window will give you the code snippets to access the RESTful location for this data. Basic SQL, Python and Javascript methods are listed by clicking on the desired language type. This string can also be used as a URL and inserted into a browser. **An example string for accessing one of the public datasets follows. This will return five records from this stored CSV spreadsheet.**

https://datahub.Perovskites.org/api/action/datastore_search?resource_id=80269ad4-e83d-4793-a2fd-1b7802a21372&limit=5

For further information on using the API and your API key for secure data access please see:

[CKAN API Documentation](#)

For PIs and System Administrators

Adding new members to a project

Adding a member to the project will allow them to access datasets that have been loaded into the project container. There are three levels of access:

1. **Member** – Read only access to data in the project
2. **Editor** – Read and add datasets or data to existing datasets
3. **Admin** – As per editor, but this user can also add members to the project. Be careful with giving this level of control. Normally this should be restricted to the PIs only for the project.

1. Click on the Project that will store data

2. Click on *Members*, left panel

3. Click on *Add Member* button

4. Enter or search for username to add

5. Set permissions for that user

6. Click on *Add Member* to finalize

Changing permissions for a member of a project

Occasionally a project PI may need to adjust permissions for a member; giving or removing permissions to add or read data. To change a user's permission follow the next instructions.

1. Click on the Project in question

2. Click on *Members*, left panel

3. Find user and Click on "*Wrench*" button to right of name

4. Change permissions from list

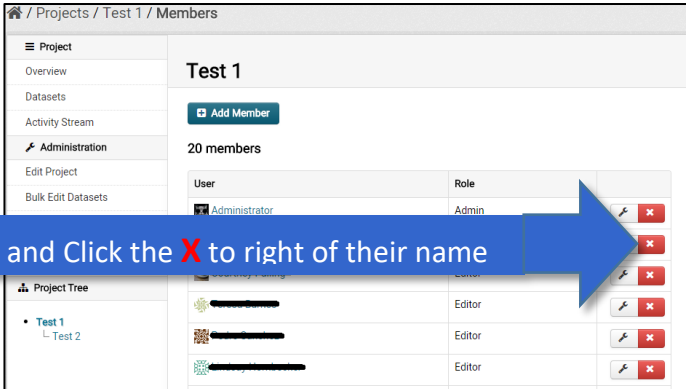
5. Click on *Update Member* to finalize

User	Role	Actions
Administrator	Admin	[Wrench] [X]
[Redacted]	Admin	[Wrench] [X]
[Redacted]	Editor	[Wrench] [X]
[Redacted]	Editor	[Wrench] [X]
[Redacted]	Editor	[Wrench] [X]









Deleting a member of a project

To remove a member from a project there are two ways to do this. Both start the same with [as above](#) up to step 3.

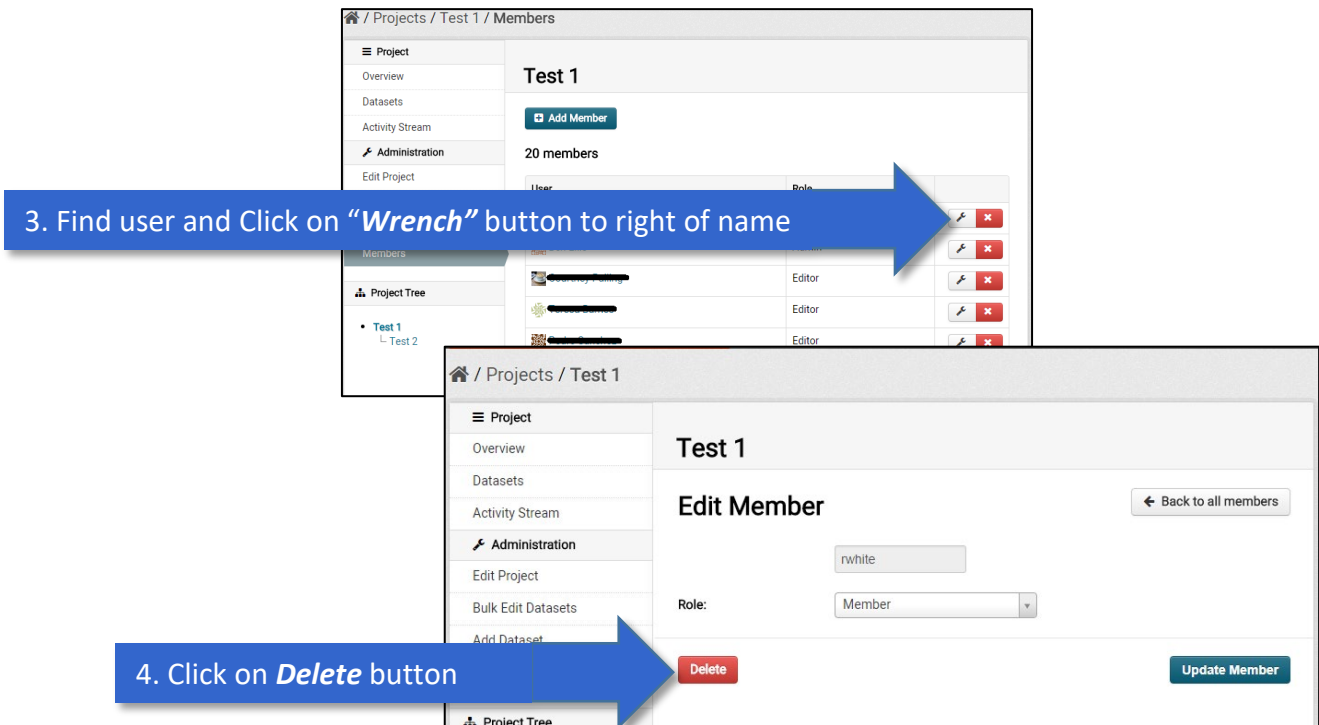
A. Deleting a user from the Members Page



3. Find the user and Click the **X** to right of their name

User	Role	
Administrator	Admin	 
[redacted]	Editor	 
[redacted]	Editor	 
[redacted]	Editor	 

B. Deleting a user from the Member's Permission page



3. Find user and Click on "**Wrench**" button to right of name

4. Click on **Delete** button

Problems



- If you cannot log on to the website
- Problems with your data
- Suggestions for improvement

Send an email to the [administrator](#) or to robert.white@nrel.gov and we will take care of the problem as soon as we can.