

# Using the DuraMat Data Hub



U.S. DEPARTMENT OF  
**ENERGY**

Energy Efficiency &  
Renewable Energy



**Energy Materials Network**  
U.S. Department of Energy

August 25, 2020

1

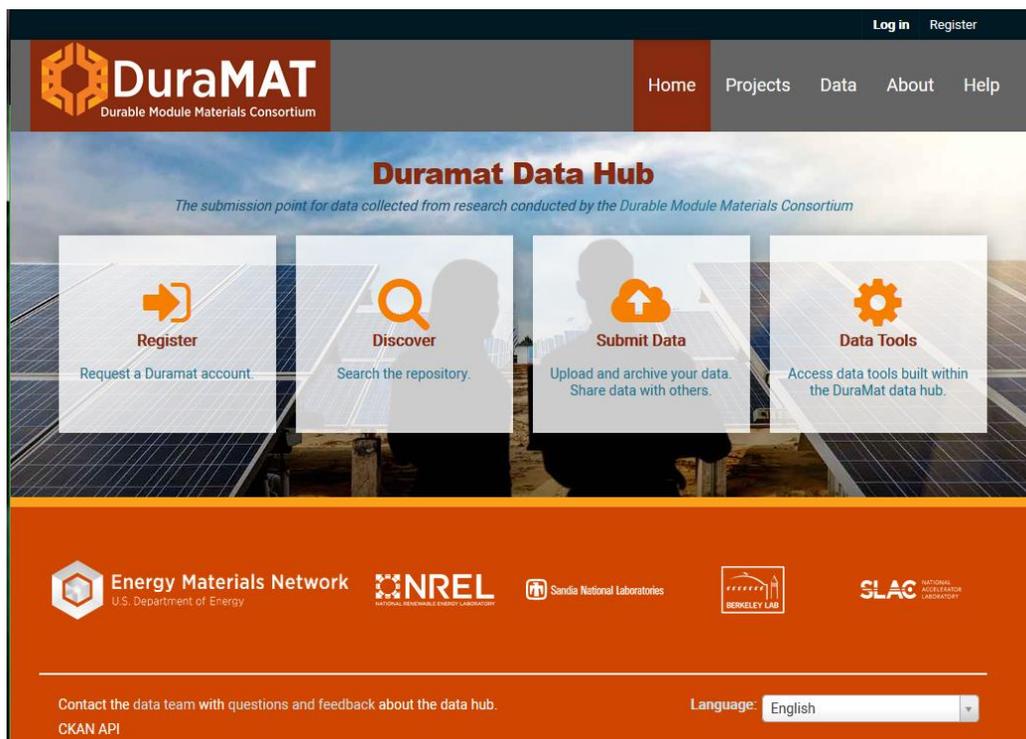
## Table of Contents

<b>Using the DuraMat Data Hub</b> .....	1
Overview .....	3
Registration and Security .....	4
Projects .....	5
Datasets .....	5
Files and Resources .....	6
<b>For All Researchers</b> .....	7
Adding Data.....	7
Adding Multiple Files in One Upload .....	8
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
<b>For PIs and System Administrators</b> .....	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 DuraMat 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 publically 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

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.

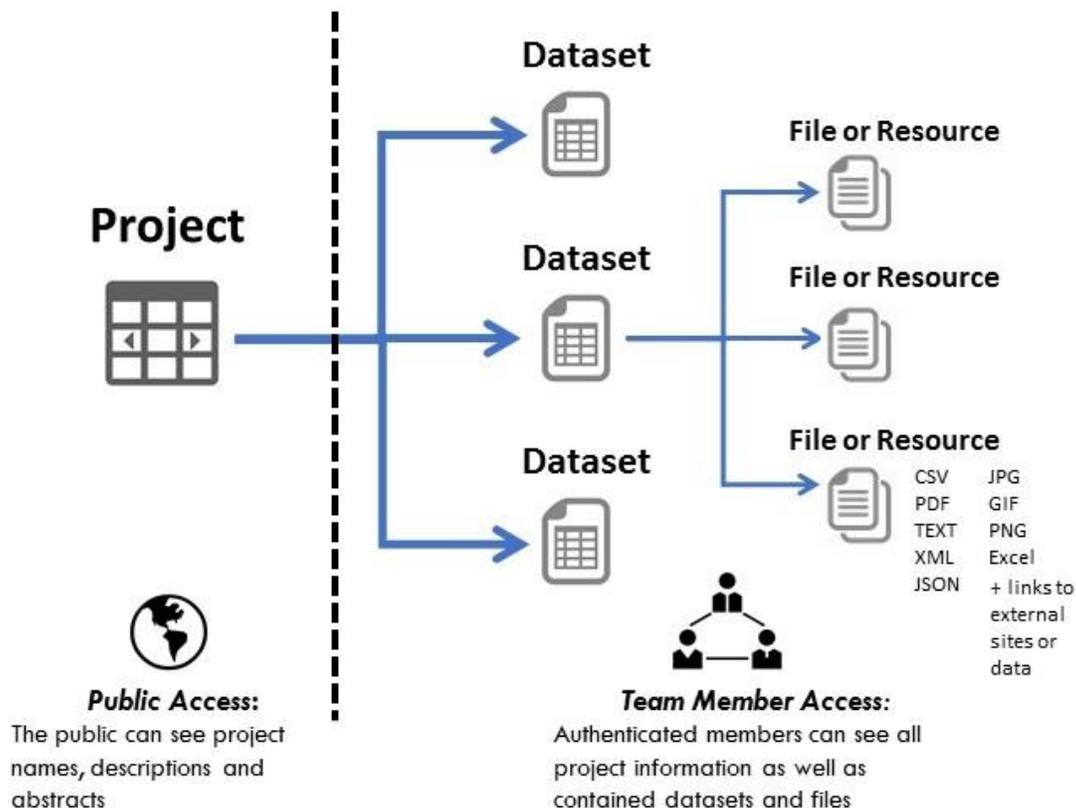
1. Enter the required information on the registration page and click **Create Account**. Lab users should use their lab username and email address.
2. The researcher needs to email the administrator ([emnadmin@nrel.gov](mailto:emnadmin@nrel.gov)) with the following details:
  - a. The Institution you 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 projects by contacting the PI for a project. The PI’s name is found on the Project Overview tab in the **Recipient** field. The project PI or designated representative has the ability to add or remove other researchers from the project’s access list.

By default all datasets created and files uploaded are designated as private, meaning only members of that project can access them. In time there will be a formal process implemented for converting the data to other access levels such as embargoed and public.

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.





## Projects

These are high level containers that are used to organize work within the data hub. Typically, these will be focused on particular research funded by the DuraMat consortium. However, there is a great deal of current and historical data that will be added to the data hub and can be housed in more generic projects, such as “Historical Data and Literature Searches”.

All projects and sub-projects are currently 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 you need created or adjusted.

A project can contain any number of datasets (see below) and sub-projects. Sub-projects allows for additional granularity for storing data within a project. Also sub-projects allow for restricting data within a project to only a sub set of normal project members.

## Datasets

A dataset is similar to a folder in most computer operating systems. 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. At a further date we will set up the procedures of public release of data. If you have a current dataset that needs to be made public immediately on upload, please contact the [administrator](#) and we will help you.

## Files and Resources

All data uploaded to the hub will be either a file or a link. Collectively, these two are called resources. The data hub is capable of housing any type of file format. We have limited the upload size to 5GB, but if the situation arises that large files may need to be stored please contact the [administrator](#).

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.

We encourage all research participants in DuraMat 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 DuraMat 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 data of all kinds that should be shared with the project members and eventually to the public if desired. The process may require the creation of a new dataset or it could be new data files added to an existing project's dataset. During the creation of the dataset and / or adding a resource or file, 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 all Fields on form

4. Click on **Add Data** to move to next page

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.

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

For the moment, just use the default license. This will be removed at a later date.

Selecting certain **Data Source Types** could open up additional metadata fields that will need to be completed.

**Create Dataset**

1 Create dataset 2 Add data

Title: eg. A descriptive title  
\* URL: datahub.duramat.org/dataset/<dataset> Edit

Project: Test 1

Description: eg. Some useful notes about the data  
You can use Markdown formatting here

Tags: eg. economy, mental health, government

License: Creative Commons Attributi...  
License definitions and additional information can be found at [opendefinition.org](http://opendefinition.org)

**Dataset Metadata**

\* Author: admin

Institution: NREL

Data Source Type: Lab and Experimental

**Lab and Experimental**

Accelerated: -- select an option --

Environmental Conditions: -- select an option --

Measurement Types:

The data license you select above only applies to the contents of any resource files that you add to this dataset. By submitting this form, you agree to release the metadata values that you enter into the form under the Open Database License

\* Required field

Next: Add Data

## 5. Choose to upload a file or create a link to outside data (If not Bulk Upload)

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.

The screenshot shows the 'Add New Resource' form. At the top, there are two buttons: 'Upload' and 'Link'. Below these are fields for 'Name' (with the example 'eg. January 2011 Gold Prices'), 'Description' (with the example 'Some useful notes about the data'), and 'Format' (with the example 'eg. CSV, XML or JSON'). There is a note: 'This will be guessed automatically. Leave blank if you wish'. Under 'Resource Metadata', there are dropdown menus for 'Data Quality', 'Data Type', and 'Data Tool Associated With This File'. Below these are 'Resource Comments' and a text input field. At the bottom right, there are four buttons: 'Previous', 'Save & add another', 'Switch to bulk add', and 'Finish'. Blue arrows point from the text boxes to the 'Upload' and 'Link' buttons, and from the 'Save & add another' and 'Finish' buttons to the instruction box for step 7.

## 6. Complete all 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.

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

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, it is advisable to go back after upload and add any important metadata that should be associated with particular files.

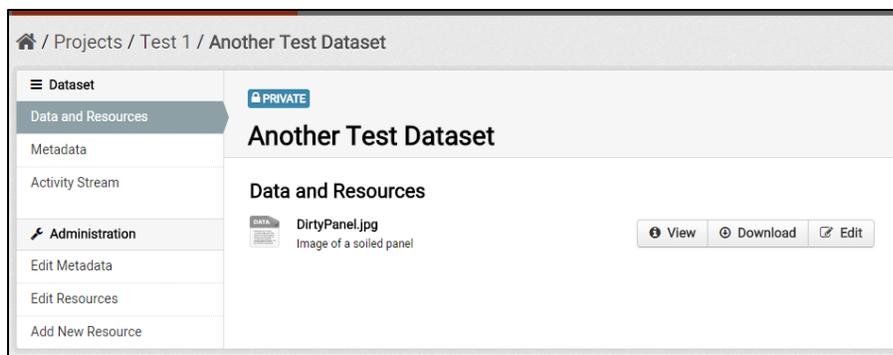
Another method is to also upload a CSV containing all the metadata values for the data files when you are doing the Bulk Upload

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



### The navigation buttons on the left panel:

*Edit Metadata* - Edit the dataset's current metadata. Only the owner/creator of the dataset can adjust the 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).

### The buttons to the right of the resources:

*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.

## 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 can show changing measurement conditions.

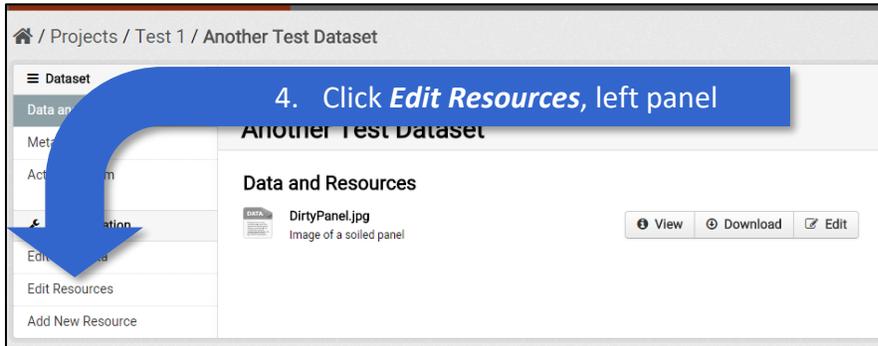
The image illustrates the process of adding data to an existing dataset in the DuraMAT system through four sequential screenshots:

- 1. Click on the Project that will store the data:** The first screenshot shows the 'Projects' page. A blue arrow points to the 'Test 1' project card in the main content area.
- 2. Click on *Datasets*, left panel:** The second screenshot shows the 'Test 1' project page. A blue arrow points to the 'Datasets' tab in the left-hand navigation menu.
- 3. Click on the dataset to add data to:** The third screenshot shows the 'Datasets' page for 'Test 1'. A blue arrow points to the 'Another Test Dataset' entry in the list.
- 4. Click *Add New Resource*, left panel:** The fourth screenshot shows the 'Another Test Dataset' page. A blue arrow points to the 'Add New Resource' button in the left-hand navigation menu.

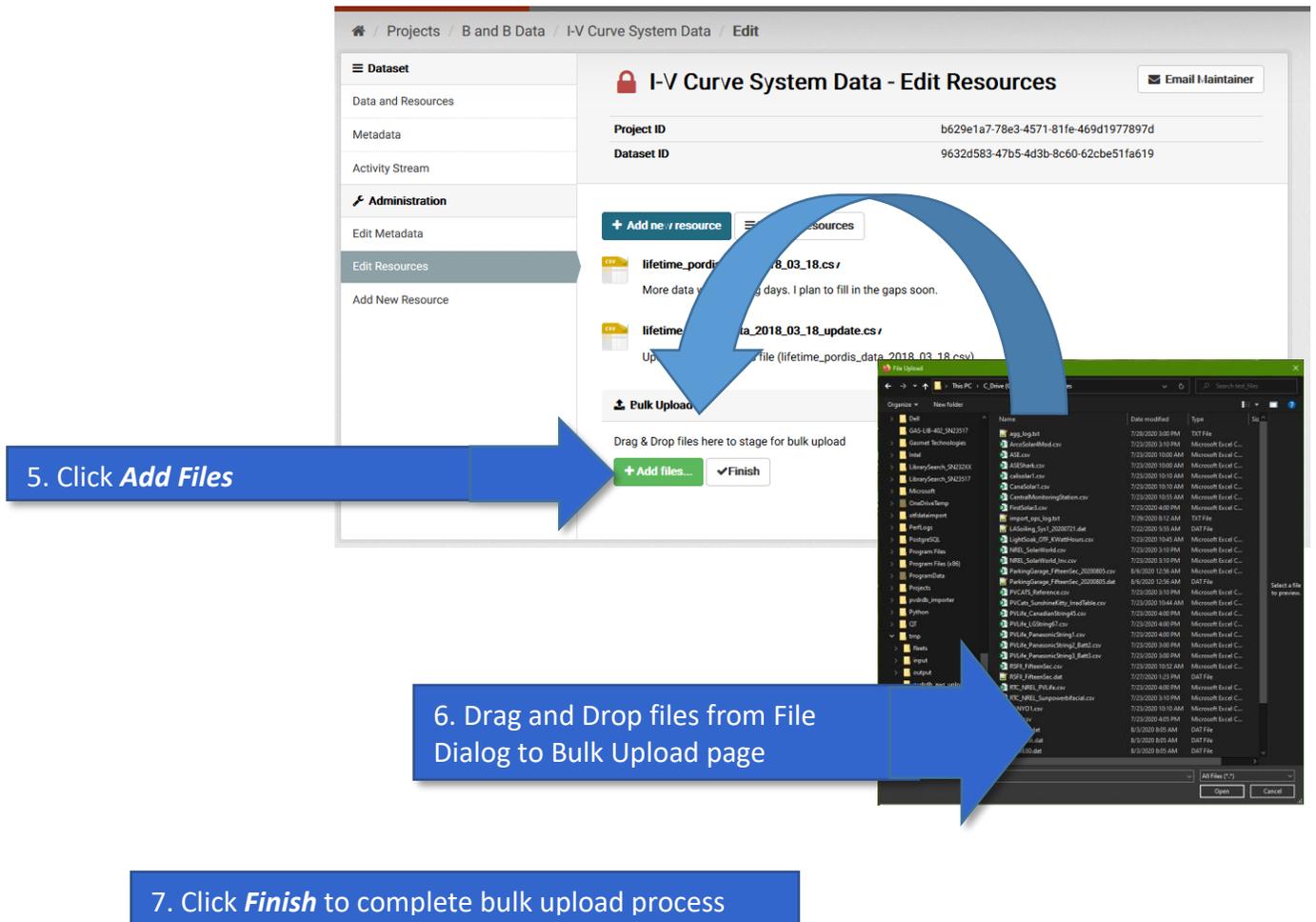
From that point follow the directions in the [previous section](#) from number 5.

## 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 the owner 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. Normally only the researcher who uploaded the file can delete it, but there could be times that the system administrator will need to do it and you can contact [the administrator](#) if it is important.

*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

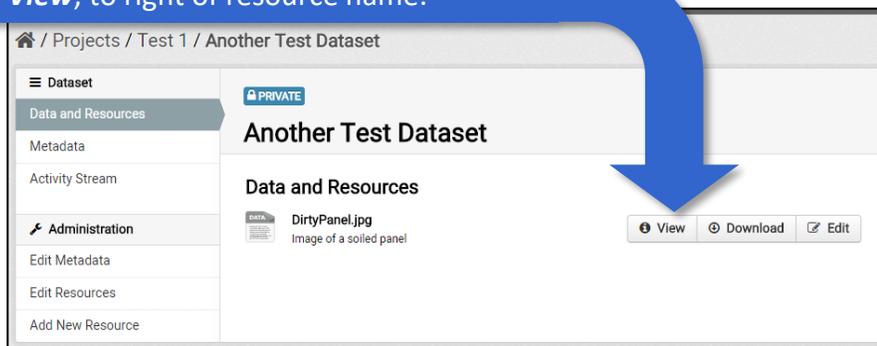
After you confirm deleting the file you will be brought back to the dataset's resource list page.

## Viewing Data

In many cases the DuraMat data hub has the inherent 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.

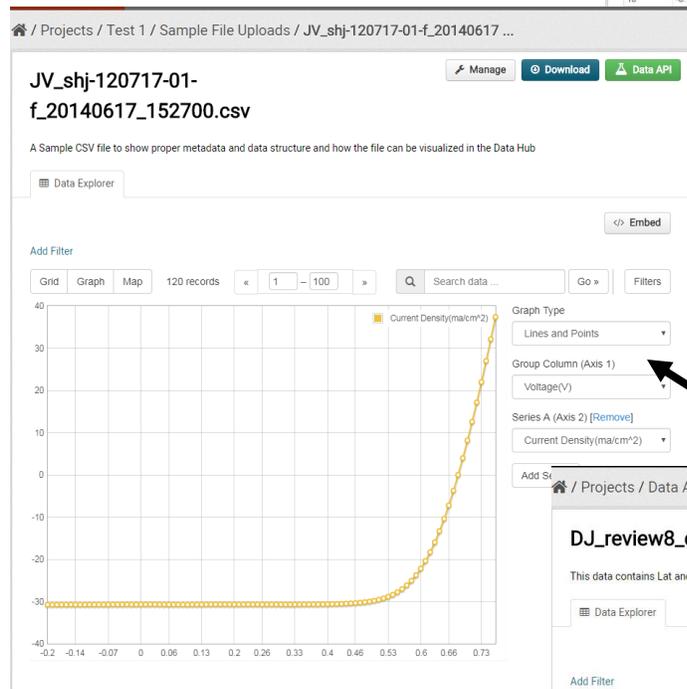
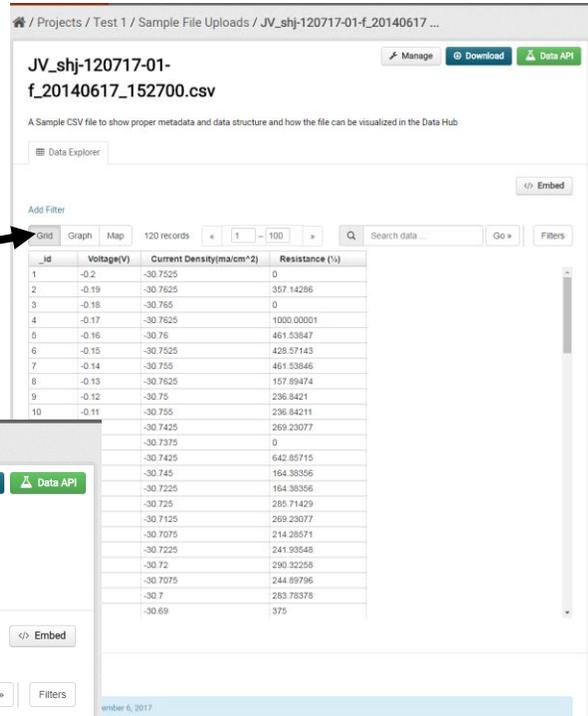
*Data API* (not shown) – This green button will open another dialog box listing the syntax to access this particular data file programmatically ([see below](#)).



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 the data has to lend itself to being viewed in that manner and both will require interaction with the user to display the data as needed. The map function requires either be Latitude and Longitude or GeoJSON included in the file for it to function correctly

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.

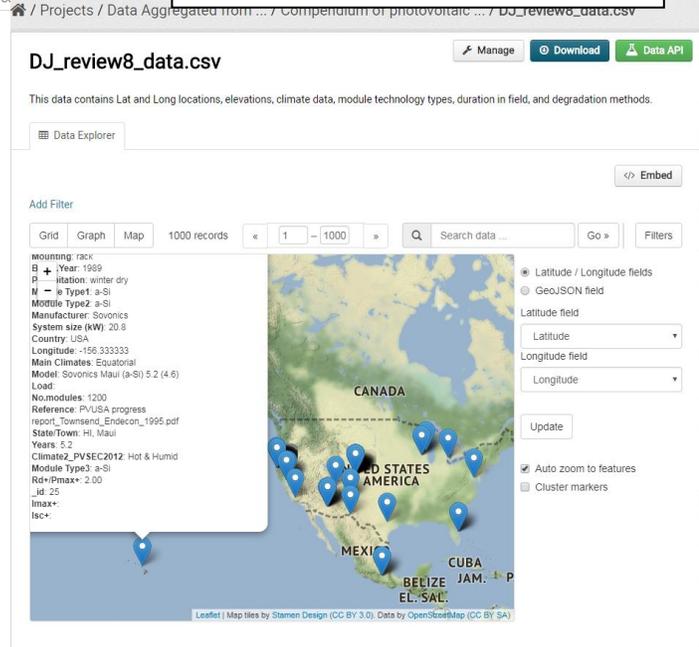
To change the display type for the CSV, Click the buttons just above and to the left of the table



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 access and download several datasets on a regular basis that need to be processed through a software pipeline or tool set (e.g. Mathematica, Igor, Origin, etc.). The DuraMat 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 screenshot of the DuraMAT website. At the top, there is a navigation bar with the DuraMAT logo and the text 'Durable Module Materials Consortium'. The user 'Robert White' is logged in. A blue arrow points to the user's name in the top right corner. Below this, a blue box contains the text '1. Click on name at the top of the page'. The main content area shows four cards: 'Register', 'Discover', 'Submit Data', and 'Data Tools'. Below this is a user profile for 'Mr. Cool Cat'. The profile includes a cat wearing sunglasses, the username 'mccat', and a list of datasets. A blue arrow points to the 'API Key' field at the bottom left of the profile, which is circled in red. A blue box contains the text '2. API Key code is on bottom left of your User page. Write it down!'. The API Key is shown as a long alphanumeric string.

1. Click on name at the top of the page

2. API Key code is on bottom left of your User page. Write it down!

API Key Private  
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

_id	Model	Manufac...	Module ...	Module ...	Module ...	System/...	System ...	Years	Begin.Y...	Ref year
1			a-Si	a-Si	a-Si	Module		1	1983	1987
2			a-Si	a-Si	a-Si	Module		0.5	1983	1987
								1	1984	1989

CKAN Data API

- Update / Insert: [https://datahub.duramat.org/api/action/datastore\\_upsert](https://datahub.duramat.org/api/action/datastore_upsert)
- Query: [https://datahub.duramat.org/api/action/datastore\\_search](https://datahub.duramat.org/api/action/datastore_search)
- Query (via SQL): [https://datahub.duramat.org/api/action/datastore\\_search\\_sql](https://datahub.duramat.org/api/action/datastore_search_sql)

Querying »

Example: Javascript »

Example: Python »

```
import urllib
url = 'https://datahub.duramat.org/api/action/datastore_search?resource_id=80269ad4-e83d-4793-a2fd-1b7802a21372&limit=5&q=title:jones'
fileobj = urllib.urlopen(url)
print fileobj.read()
```

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

3. Use example code in 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.duramat.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 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**

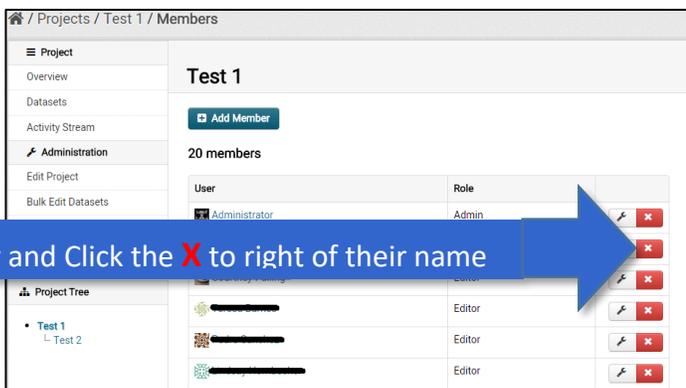
The screenshots show the DuraMAT interface with the following elements:

- Screenshot 1:** The 'Projects' page with a search bar and a list of projects. A blue arrow points to 'Test 1'.
- Screenshot 2:** The 'Test 1' project page with a 'Members' link in the left sidebar. A blue arrow points to the 'Members' link.
- Screenshot 3:** The 'Members' page showing a table of 20 members. A blue arrow points to the 'Wrench' icon in the 'Action' column.
- Screenshot 4:** The 'Edit Member' page showing a dropdown menu for selecting permissions (Member, Admin, Editor, Member). A blue arrow points to the dropdown.
- Screenshot 5:** The 'Edit Member' page with the 'Update Member' button highlighted. A blue arrow points to the button.

## 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.

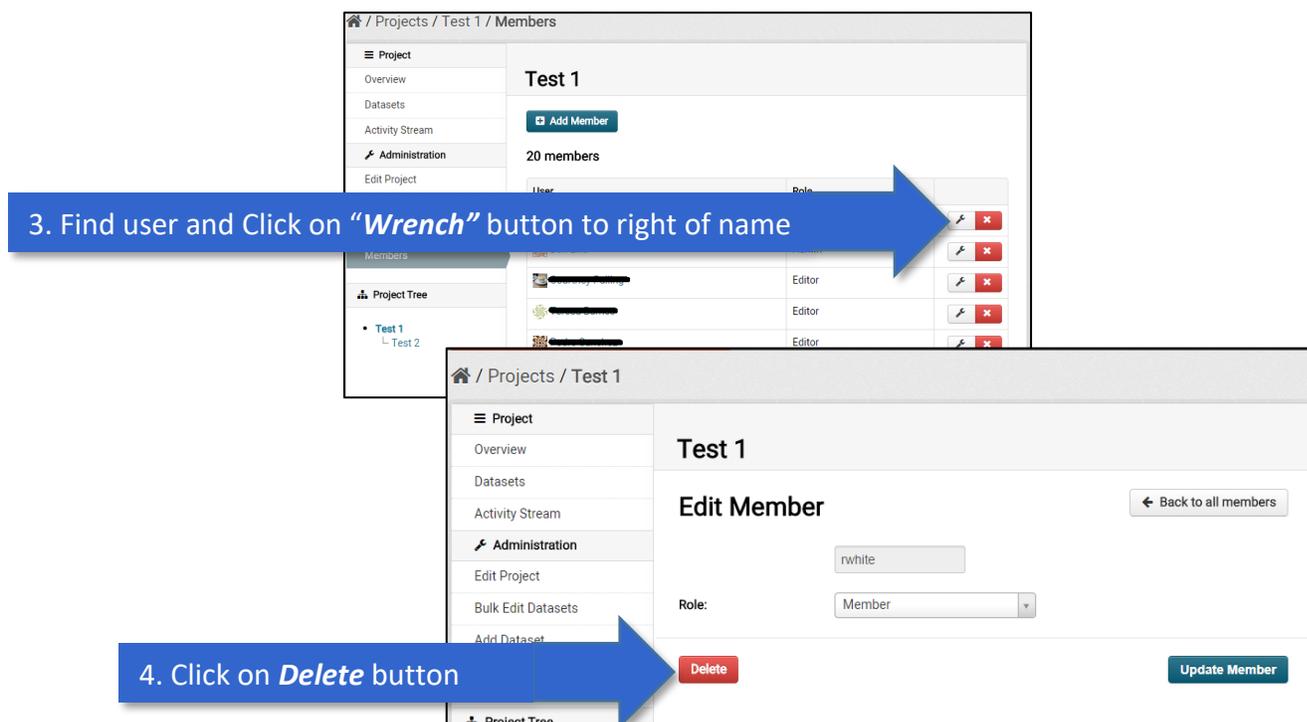
### A. Deleting a user from the Members Page



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

User	Role	
[Redacted]	Editor	 
[Redacted]	Editor	 
[Redacted]	Editor	 

**Edit Member**

Role:

**Delete** **Update Member**

## 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](mailto:robert.white@nrel.gov) and we will take care of the problem as soon as we can.