Using the Data Hub

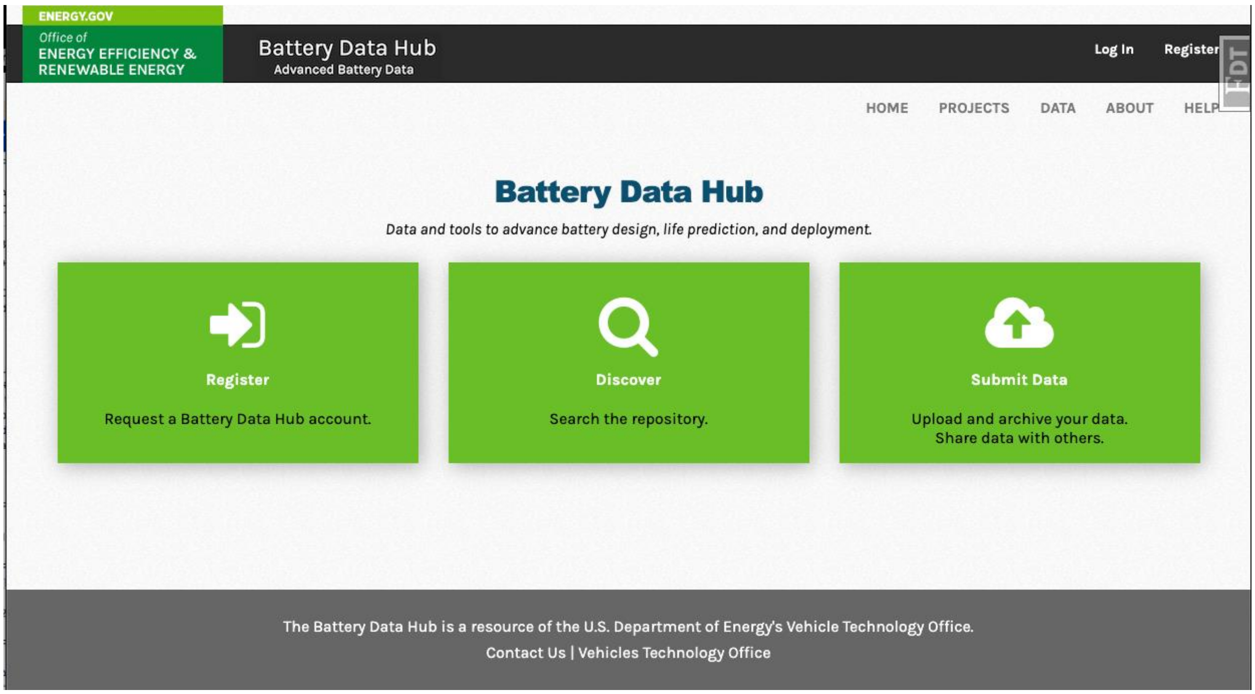
Pictures shown in this document reflect specific Data Hubs, but the key functions remain the same

Table of Contents

Overview

The 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 eventual releasing of data from DOE-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 focus of the data hub is to provide the users with a system that can provide the efficiency and security for collaborative data sharing and public release of data, as specified by DOE data requirements.



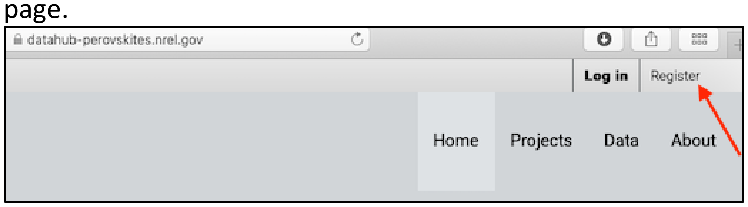
The architecture of the Data Hub can be divided into three main areas

- Registration, Login and Security
- Projects and Data
- Search

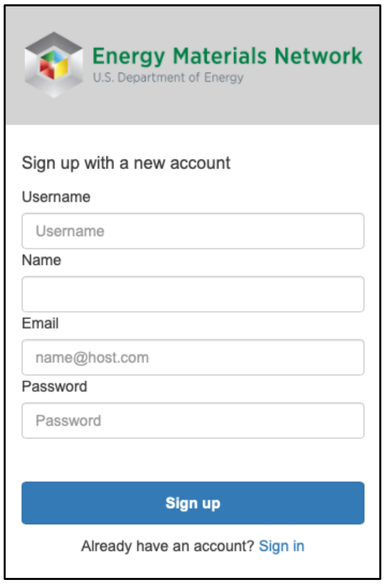
Registration, Login and Security

You must be a registered user of the data hub and a consortium member to be able to access any files that are not deemed **"Public".**

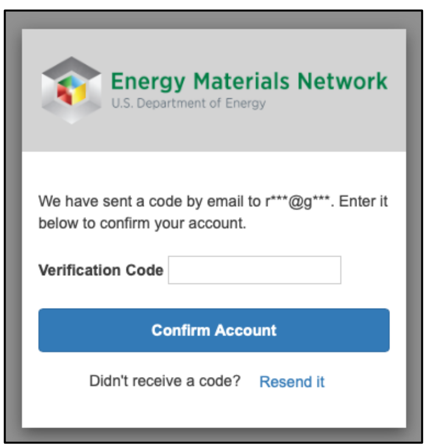
1. The registration process is simple and begins by clicking the large *Register* button on the main page.



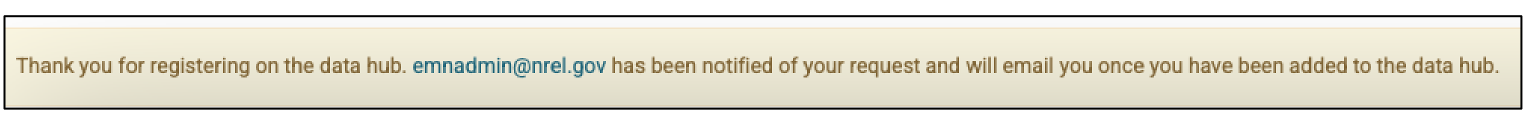
2. Register for an account below and provide the following details and click **Sign Up**.
   **Please note IDs and PWs are case sensitive.*
   - Username: Select a desired username
   - Name: List your first and last name
   - Email: Please use nrel.gov email if applicable
   - PW: Select a strong password



3. This will pop up and you will receive an email with a verification code. Please obtain the code from your email address, enter the code, and click **confirm account**.



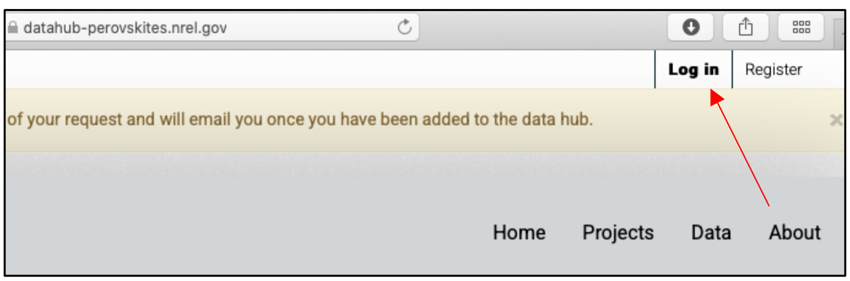
4. If you do not have an existing account, once you click confirm account, you will see a flash pop up message that looks like the below:



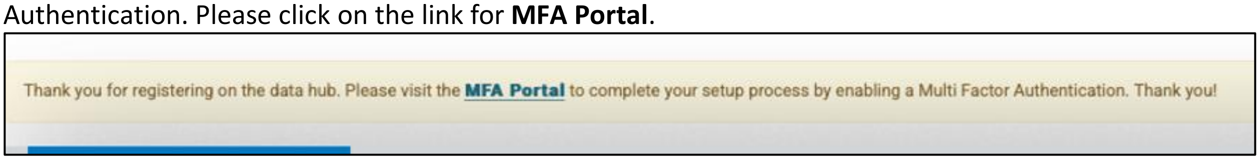
5. Obtaining Access to Specific Data Hubs and Projects.
   An EMN Administrator will need to add your new user to the appropriate access groups.
    a. Upon completion of the previous steps, an automated email will be sent to the EMNAdmin administrator mailbox (emnadmin@nrel.gov).
    b. This mailbox is monitored and when requests are received, we work with the appropriate PIs to validate access.
    c. Once access has been added you will be notified and can proceed with the remaining steps in this document.
    d. *Attempting to move forward with MFA configuration prior to access groups being configured will be unsuccessful.*

6. Configuring Multi Factor Authentication.
   **Please Note: Multi factor authentication requires users to provide a secondary credential in addition to their password to access a system. Cyber and security policies require datahubs that maintain moderate or higher data to require MFA (Multi Factor Authentication)*
   Download an authenticator app (such as Authy or Google Authenticator) available as you complete the remaining steps in this document.
    a. Later steps in this configuration process, as well as future login attempts, will require users to generate and provide a secondary credential to login. This can be done with an authenticator app.

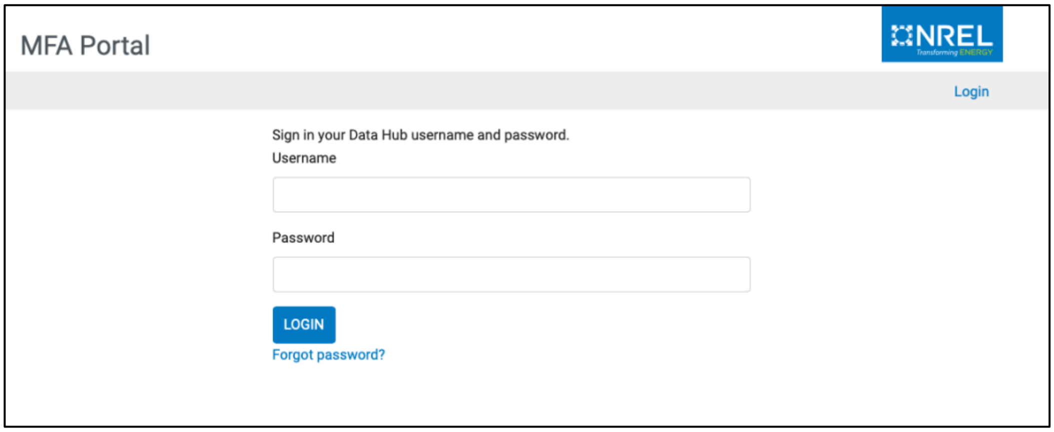
7. Once you've received notice that you have been added to the correct group, click the *login* option at the top right of the datahub page.



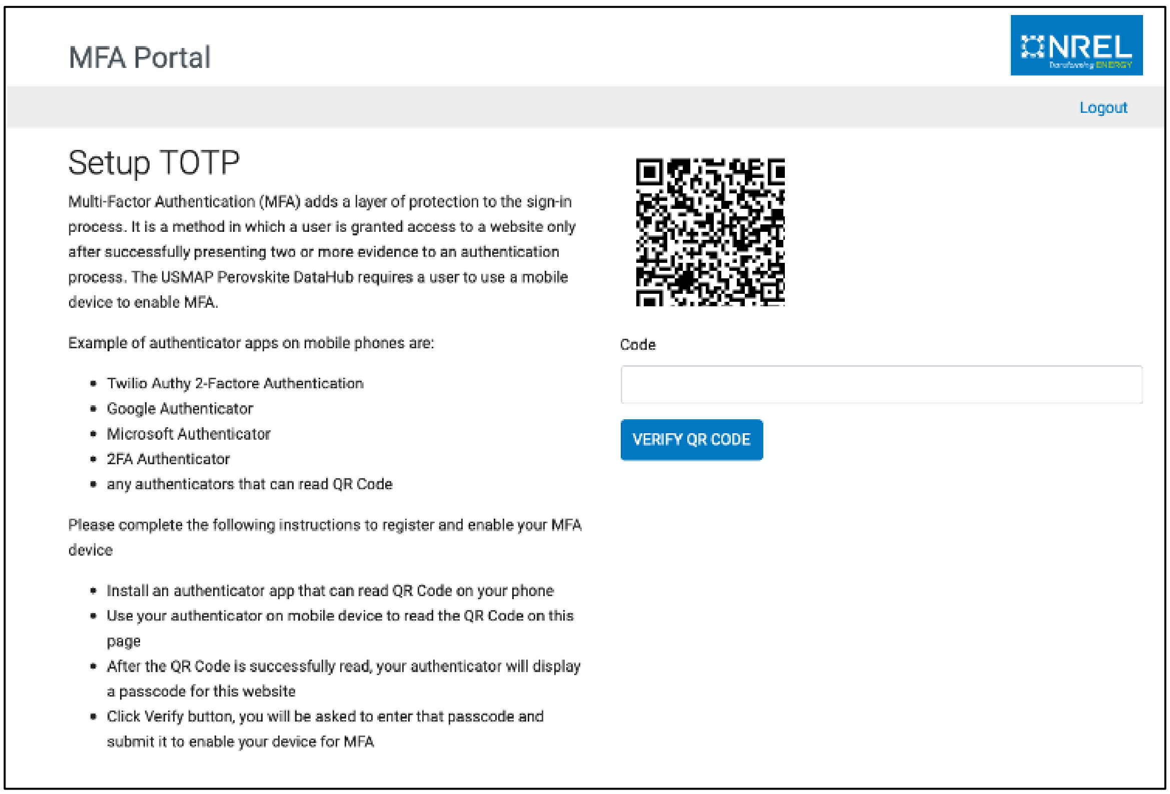
8. The first time you login you'll see the following message requesting you to set up Multi Factor Authentication. Please click on the link for **MFA Portal**.



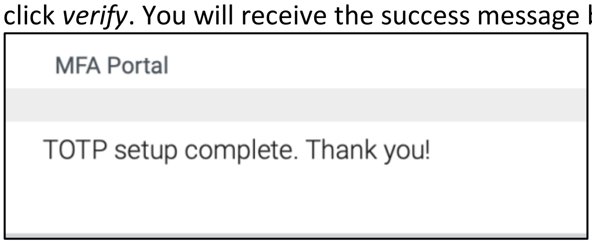
9. On the MFA portal application page, you will see a prompt to login. This step will configure your Multi factor authentication – enter your previously created Data Hub username and password (Step #3) in the fields provided and click *login*.



10. After entering credentials you'll be provided with a QR code like below. Using the previously downloaded authenticator app (Step #6), add your account and scan the QR code on your screen.

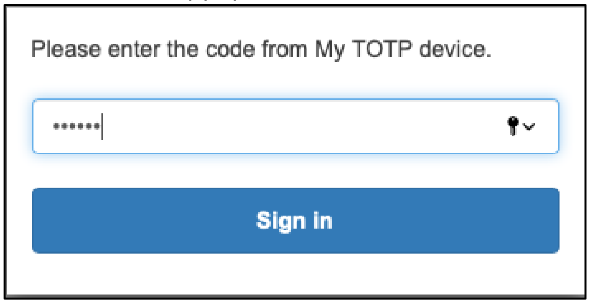


11. After scanning the QR code the app will provide you with an authentication token (a numerical value). Please enter the token value into the **code** box on the NREL MFA Portal Application and click *verify*. You will receive the success message below.



12. Your datahub ID, access user groups, and multi factor authentication should now be configured. Navigate back to the main data hub page, click *login*, and provide your newly configured ID and PW and click *Log In.*

13. You will be prompted for a secondary credential (MFA). Obtain your MFA token from your authenticator app, provide it in the box, and click *Sign In*.

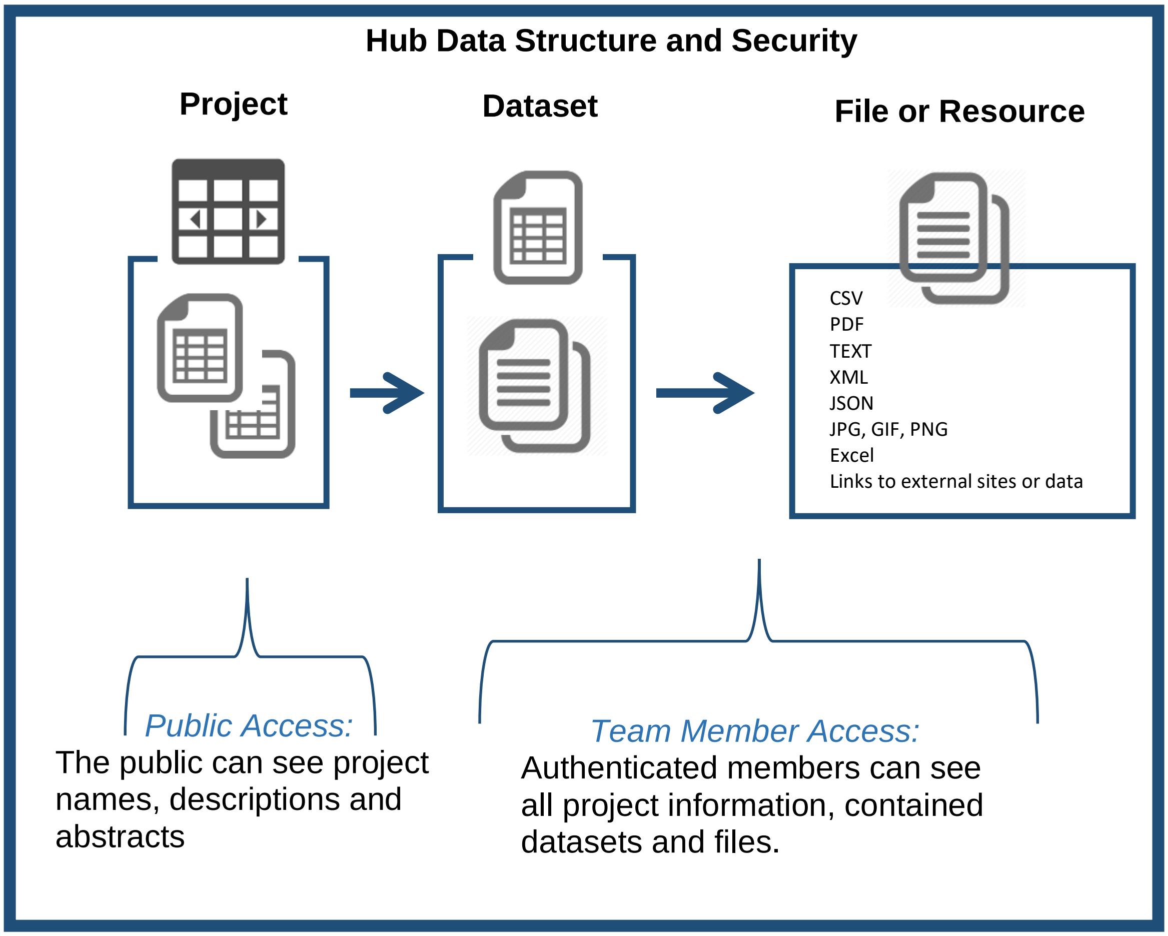


14. You will now be logged into the datahub and can see your username at the top right of the window.

Hub Data Structure and Security

By default, all new datasets 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 details about any project and can view and download any designated public data. However, any private datasets cannot be seen or accessed by the public.



Projects & Data

Projects on the Data Hub are created by admin or PIs and are then used to organize data further by creating and uploading distinct Datasets and Resources (see below).

All projects and sub-projects are created by a system administrator. Contact the administrator (emnadmin@nrel.gov) for any projects or sub-projects you need created or adjusted.

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

Datasets, Files & Resources

In the data hub, data is added to a unit called a “dataset”. A dataset is similar to a folder in most computer systems; a dataset is a parcel of data. Datasets are used to organize data into logical areas and each dataset can contain any number of files or resource links. For example, it could be various runs for a particular type of experiment utilizing the same technology type, it can be data related to a particular DOI, or it could be temperature readings from various weather stations.

A dataset contains two things:

1.) Information or “metadata” about the data. For example, the title and author, date of upload, institution, technology type, data source type, etc.

2.) A number of “resources”, which hold the data itself. A resource can be in any format: a CSV, Excel spreadsheet, XML file, PDF document, image file, linked data in RDF format, etc. A resource can be different characterization data (e.g., SEM, XPS, BET, XRD, RDE) for the same material sample; different resources might contain the data for different years or experiment runs, or they might contain the same data in different formats. A resource can be any file type, or it can simply be a link, the resource itself being elsewhere on the web. A dataset can contain any number of resources.

We have limited the upload to 5GB/file with the following caveats. If uploading a large file (>1GB) the upload will potentially take a significant amount of time and require several browser refreshes as the file completes uploading in the background. However, using the API for uploading large files is recommended. If the situation arises that larger files may need to be stored or if you need help using the API to upload large files, please contact the administrator.

Metadata

Metadata (data that describes the data) is a set of information that describes the file being uploaded. Datasets have associated metadata which the researcher is prompted to enter upon Dataset creation. Each user uploading data into the hub should fill out any associated metadata.

Search

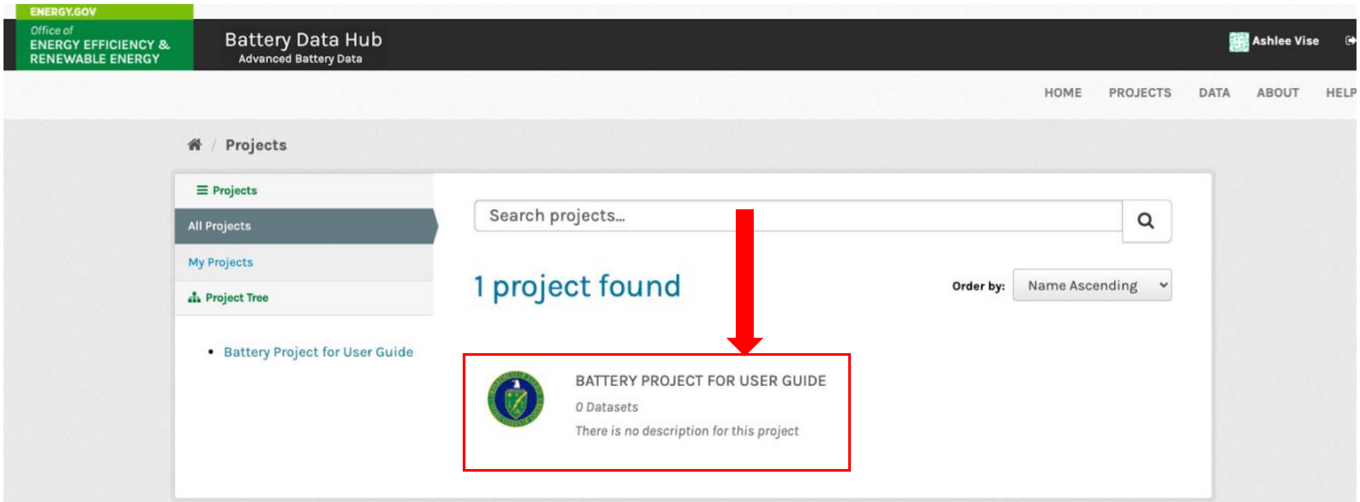
The Data Hub allows you to search on different criteria defined within the metadata for Projects and Datasets. You can search for data from the *“Discover”* button on the “*Home”* tab and from the “*search bar”* on the *“Projects* “or *“Data”* tab. The “*Data”* tab also displays the left-hand “*metadata faceted search”* options. The faceted search shows pieces of metadata that have been identified at the Dataset level with a parenthetical reference of the number of times that metadata is being used. Your results will only return data you have access to.

For All Researchers

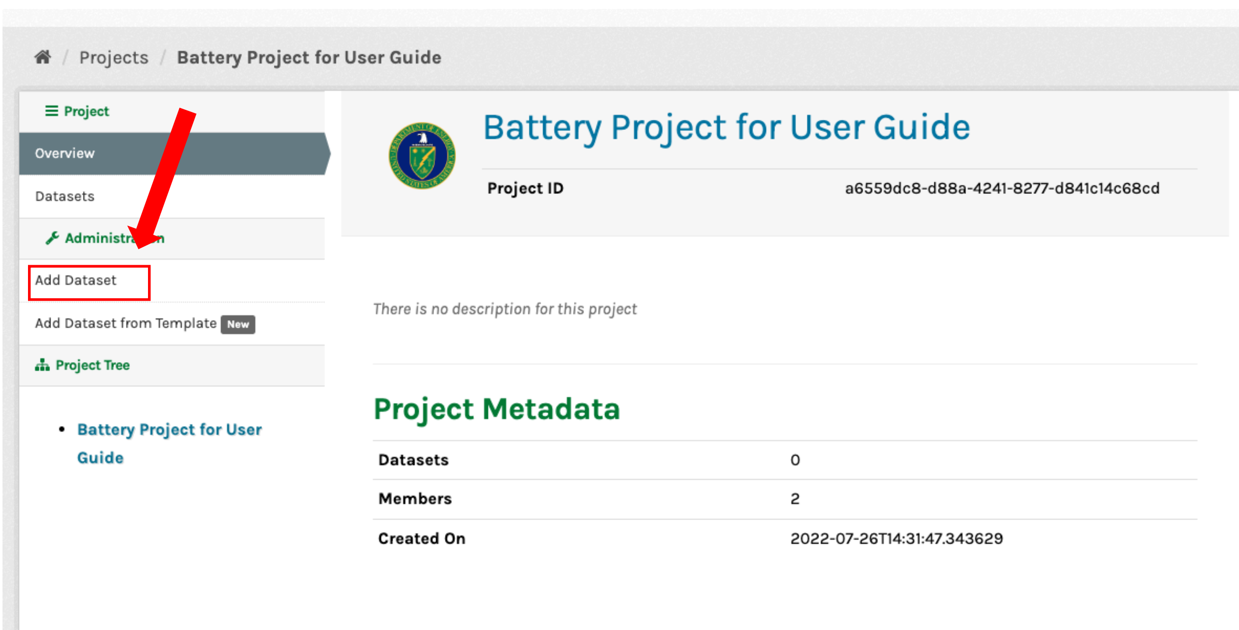
Adding Data

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

1. Click on the *Project*, you want to add data into



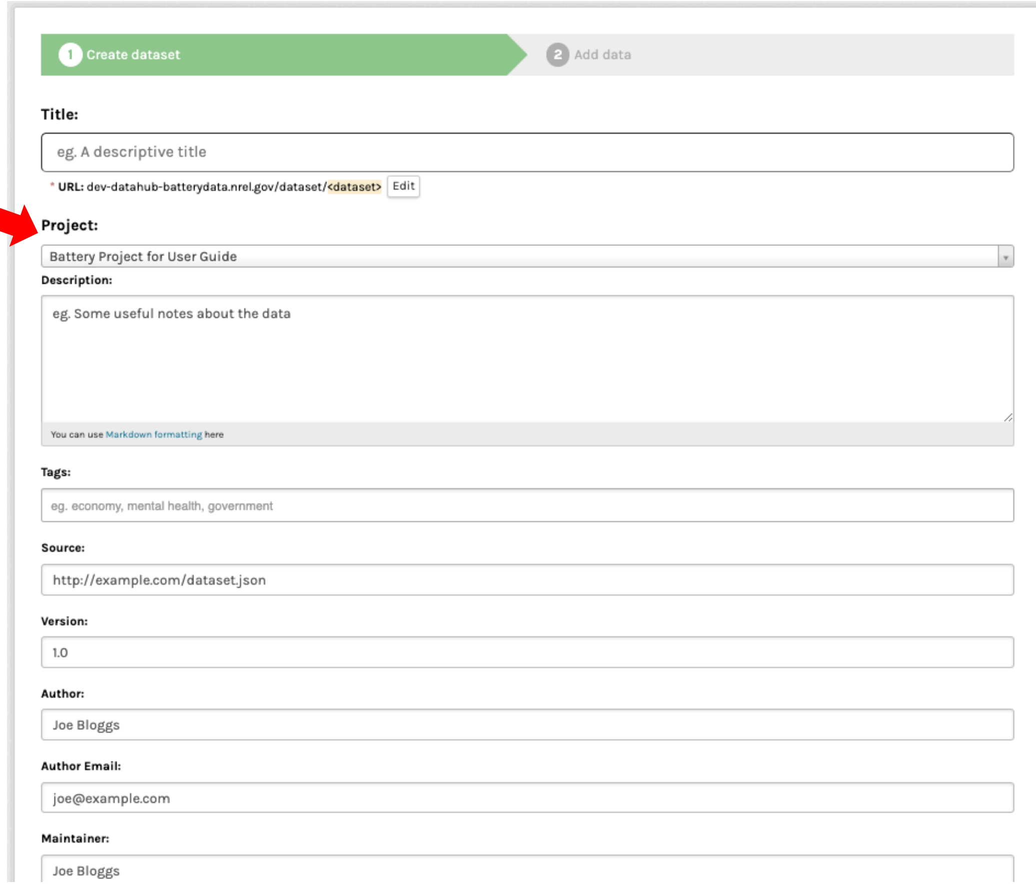
2. Click on *Add Dataset*, left panel



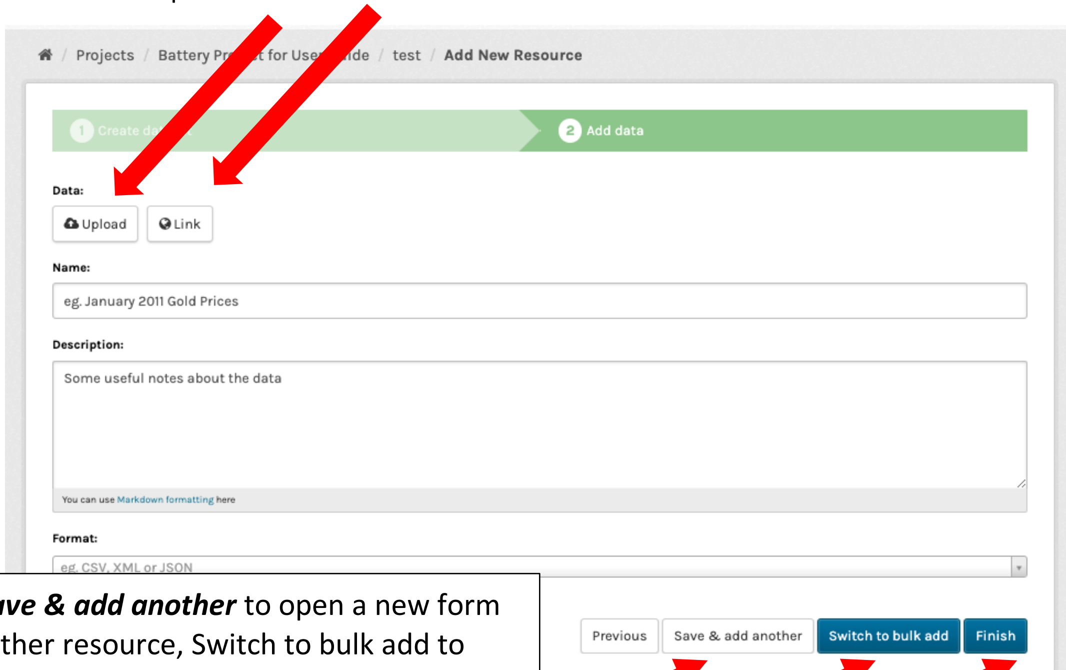
3. Complete all Fields

Verify that the selected *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 and hit Return after each tag



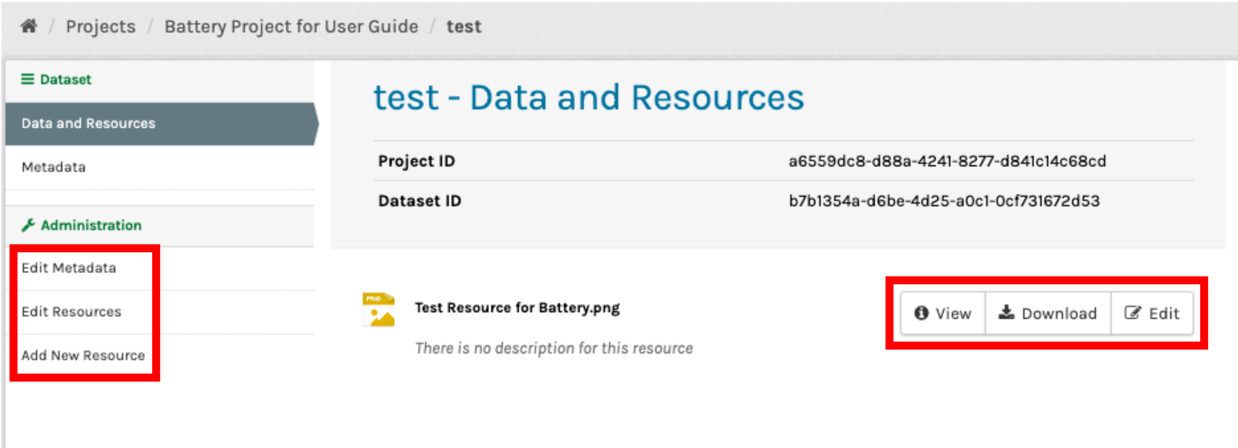
4. Choose to Upload a File or Attach a Link



Click ***Save & add another*** to open a new form for another resource, Switch to bulk add to upload multiple resources at the same time or click ***Finish*** to complete the upload.

5. 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.



The navigation buttons on the left panel:

- *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).

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

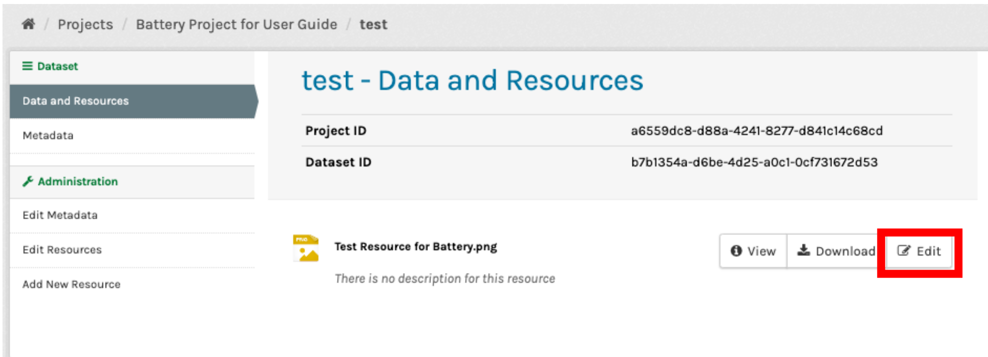
Within some working projects, depending on how the datasets are being used, you may need to continue adding data to an existing dataset. Example: A project could have a dataset for all XRD measurements. The dataset metadata could be the *Lab and Experimental Data Source Type,* and each data file could cover separate measurements for different samples (e.g., prepared by using different conditions), so can show different sample names, synthesis techniques/conditions, and/or pre- and/or post-treatments of a sample.

Deleting a File or Resource

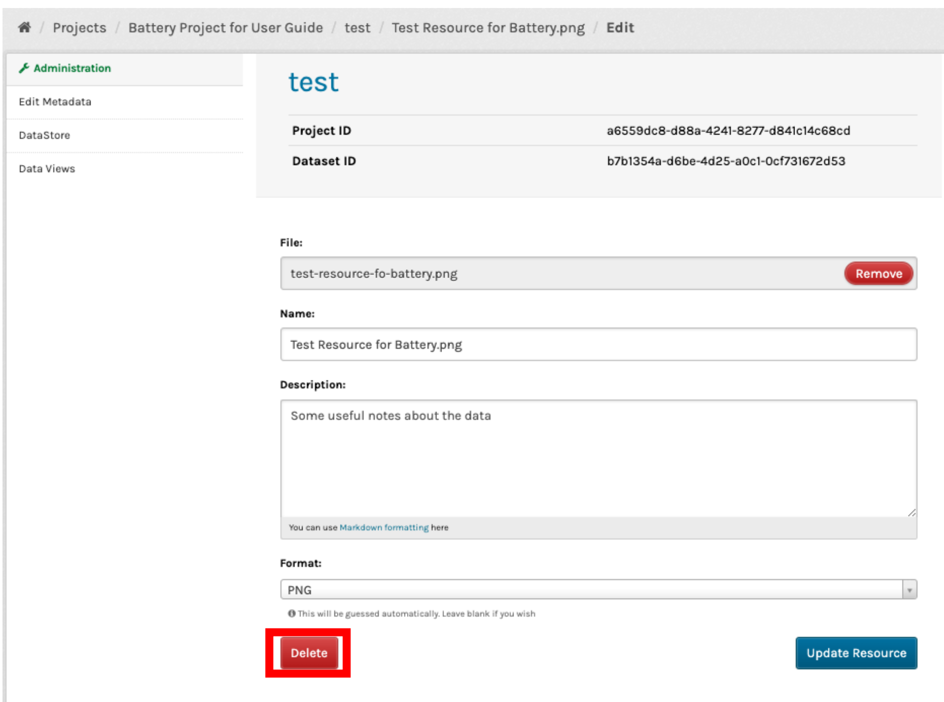
There will be times when a user might need to delete a file or resource that has been added to a dataset.

Starting from the Dataset's page

1. Click *Edit*



2. Click *Delete*



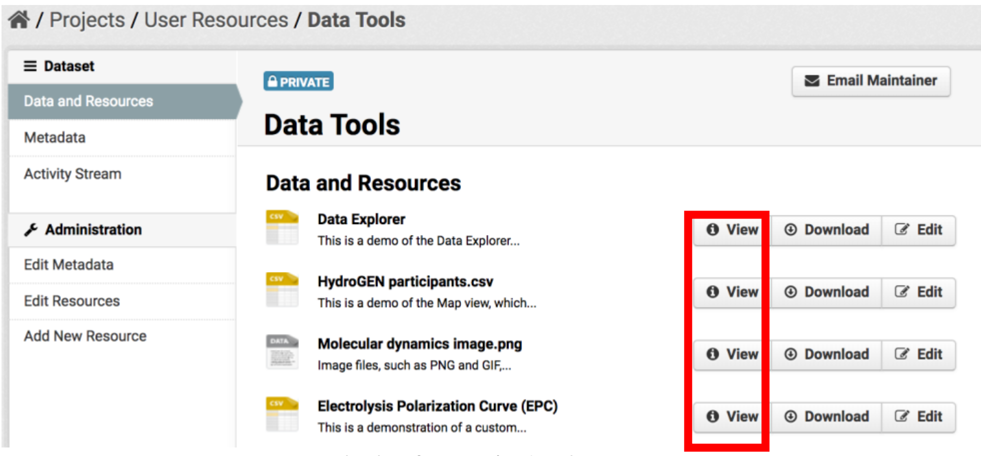
3. After you confirm deleting the file you will be brought back to the dataset's resource list page.

Viewing Data

In many cases the Data Hub has the 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

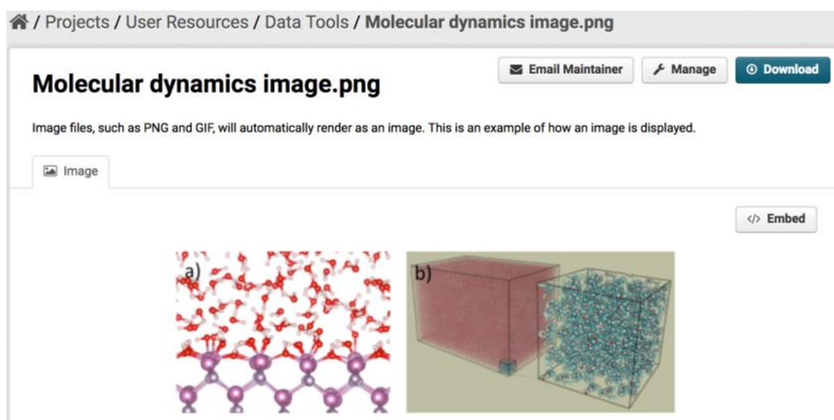
1. Click *View*, to right of the data or resource name.



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.

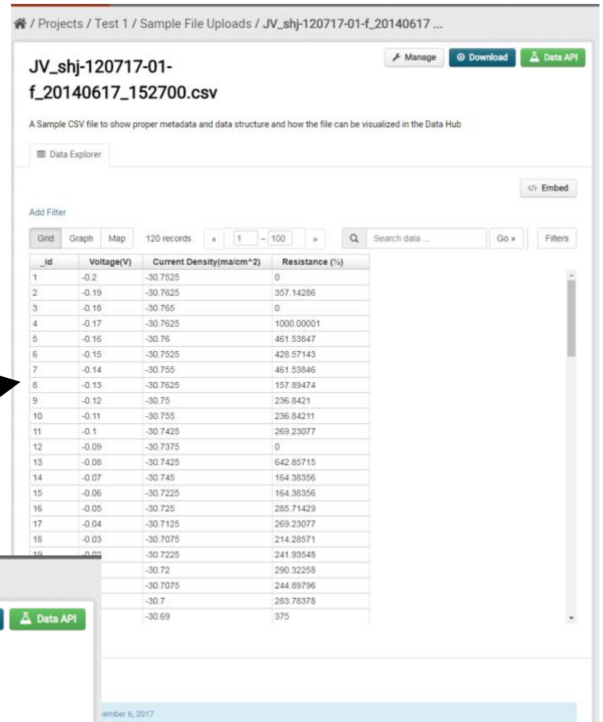
On the right across from the resource name are 2-3 buttons:
Manage – This allows you to change the metadata for the resource.
Download – This will download the file through your browser to your local computer.
Data API – This green button will open another dialog box listing the syntax to access this particular data file programmatically (see below).



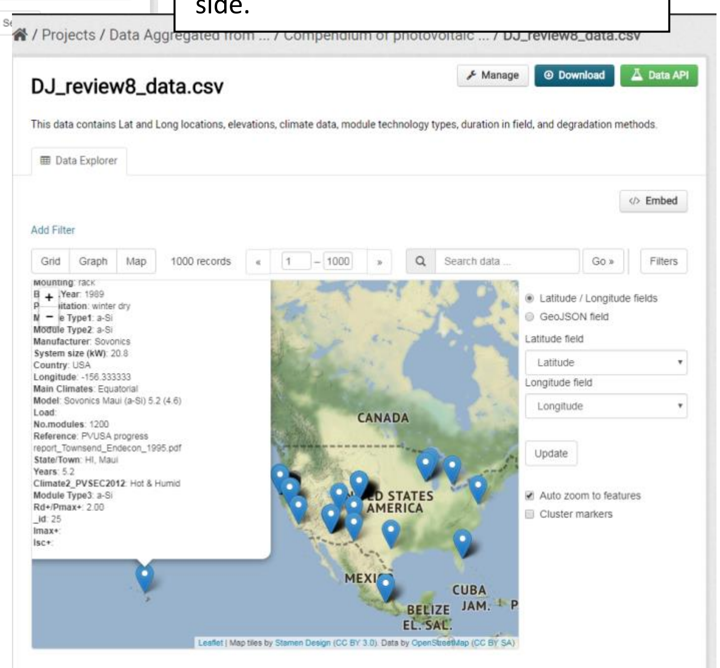
With **CSV files the Data Explorer plugin** will allow for the data to be viewed in three possible ways: **Table, Graph or Map**. To utilize the Graph or Map the data has to lend itself to being viewed in that manner and both will require interaction with the user to display the data as needed. The map 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) that is available as part of the data hub infrastructure. Secured, non-public data will still need to be accessed with credentials, but this can be passed as part of the API call. Using the API can be useful if you need to **access and download multiple files or datasets** on a regular basis in order to be processed through a software pipeline or tool set (e.g. Mathematica, Igor, Origin, etc.)

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 HydroGEN data hub uses the native CKAN API, which is built on a RESTful interface. For further information on using the API and your API key for secure data access please see:
CKAN API Documentation

For Project Principle Investigators (PIs)

Adding new members to a project

Contact the administrator to request to add a new member to a project. If the new member has not registered, the PI may send an email invitation to the new member (Cc: emnadmin@nrel.gov) inviting them to register on the data hub.
There are two main levels of access:

1. **Member** – Read only access to data in the project
2. **Editor** – Read and add datasets or data to existing datasets

Changing permissions for a member of a project

Occasionally a project PI may request to adjust permissions for a member; giving or removing permissions to add or read data. Contact the administrator to adjust permissions for any member of a project.

Removing a project member

Contact the administrator to request to remove a user from a project.

Questions and Feedback?

- If you cannot log on to the website,
- Have problems with your data, and/or
- Have suggestions for improvement

Send an email to the administrator (emnadmin@nrel.gov).