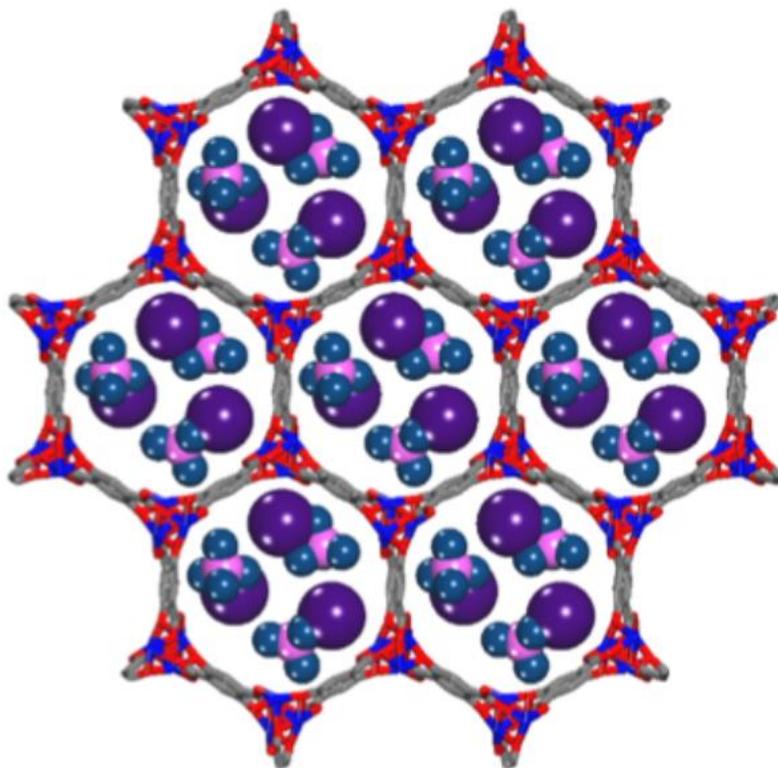


Using the HyMARC Data Hub



U.S. DEPARTMENT OF
ENERGY

Energy Efficiency &
Renewable Energy



Energy Materials Network
U.S. Department of Energy

November 20, 2017

1

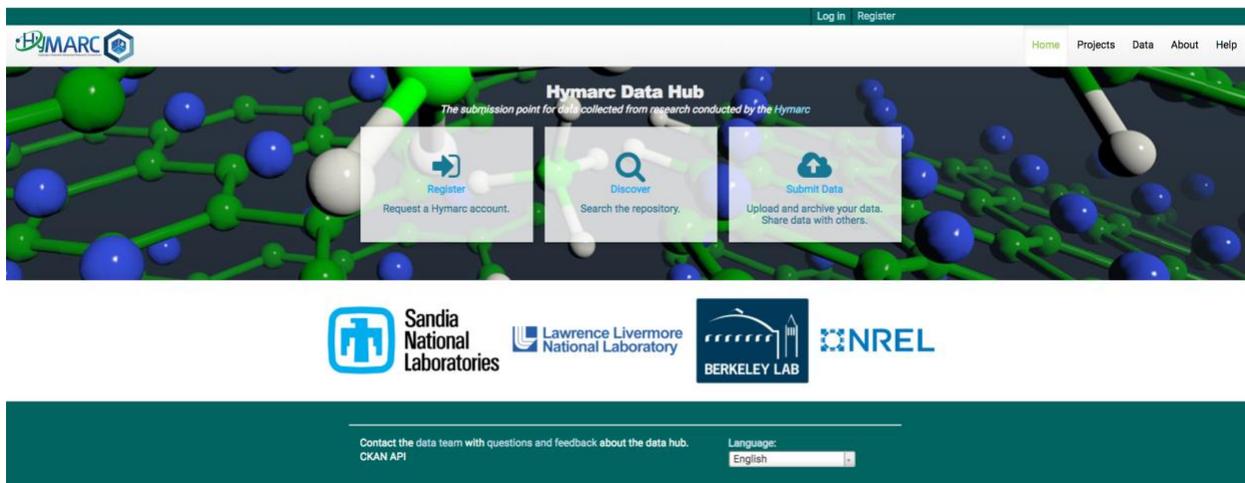
Table of Contents

Using the HyMARC Data Hub	1
Overview	3
Registration, Login and Security	3
Projects & Data	5
Datasets, Files & Resources	6
Metadata	6
Search	7
For All Researchers	7
Adding Data	7
Adding Data to an Existing Dataset	11
Deleting a File or Resource	13
Viewing Data	14
Accessing the Data through the API	16
For Project Principle Investigators (PIs).....	17
Adding new members to a project	17
Changing permissions for a member of a project	17
Removing a project member	17
Questions and Feedback?.....	17

Overview

The HyMARC 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 data hub allows for searching the metadata and data of all resources stored within, providing a method to support discovery.

The data hub is built around the [Comprehensive Knowledge Archive Network](#) or 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 and public release of data, as specified by DOE data requirements.



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

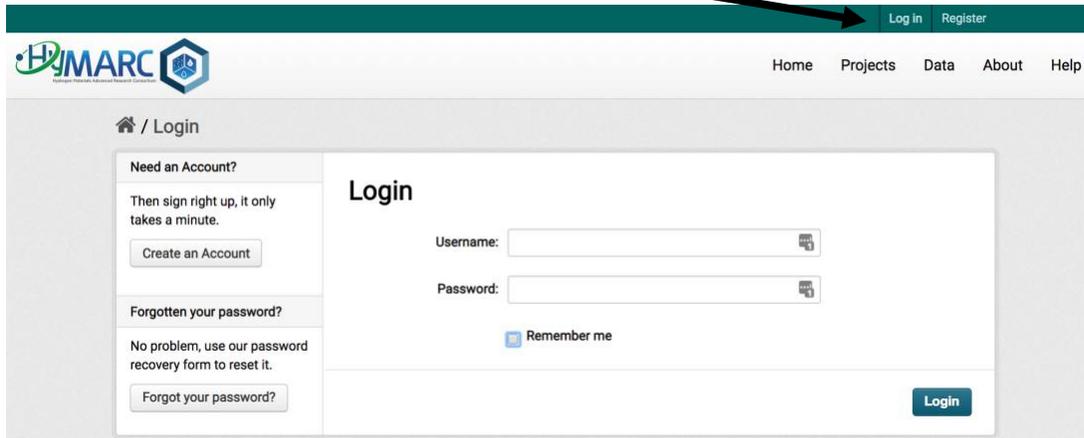
- Registration, Login and Security
- Projects and Data
- Search

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”**. 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**.
 - a. **National Lab users** should use their lab username and lab email address
2. The researcher needs to **email the administrator** (emnadmin@nrel.gov) with the following details:
 - a. The Institution you work for.
 - b. The username you registered as.

- 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.*
3. The administrator will contact the Project PI for **access approval**.
4. Once given access to the site you may **Login** to the data hub.



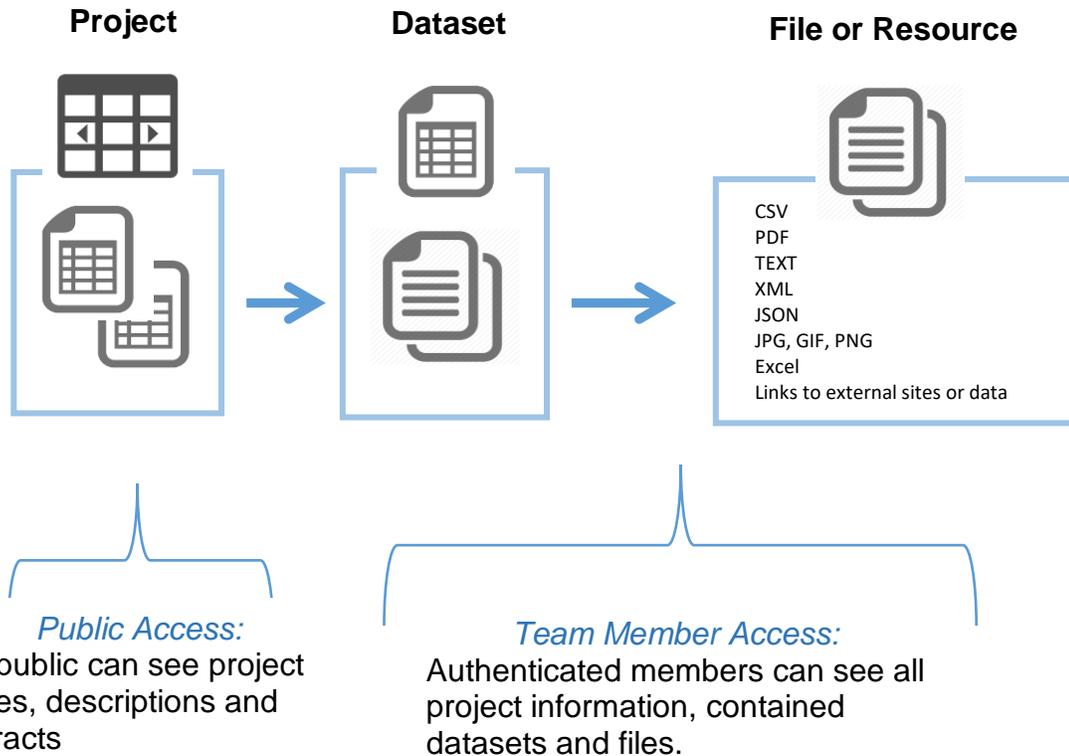
The screenshot shows the EMARC website's login interface. At the top, there is a dark green navigation bar with 'Log in' and 'Register' links. Below this is a white navigation menu with 'Home', 'Projects', 'Data', 'About', and 'Help'. The main content area is titled '/ Login' and contains a login form with fields for 'Username' and 'Password', a 'Remember me' checkbox, and a 'Login' button. To the left of the form are links for 'Need an Account?' and 'Forgotten your password?'.

You can request additional projects by contacting the administrator (emnadmin@nrel.gov).

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 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.

Hub Data Structure and Security



Projects & Data

The awarded projects within the HyMARC Consortium become the Projects on the Data Hub. Projects on the Data Hub are then used to organize data further by creation and upload of distinct Datasets and Resources (see below).

The “User Resources” project showcases data tools and user guides for the Data Hub. All consortium members are given access to the “User Resources” project so they may review the custom data tools and data hub user documents.

All projects and sub-projects are created by a system administrator. 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. 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 project in units called “datasets”. 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, it can be data related to a particular DOI; 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 (data generator), date of upload (and update), institution, data source type (External Data or Tool, Historical or Literature, Lab Experimental, Modeling and Simulation), etc.

2.) A number of “resources”, which hold the data itself. A resource can be in any format: a CSV or 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. Dataset metadata includes Institution, Author, Maintainer E-mail, Sample Barcode, Collection Date, Data Source Type (External Data or Tool, Historical or Literature, Lab Experimental, Modeling and Simulation). Each user uploading data into the hub will need to fill out any associated metadata.

Existing metadata options can be reviewed by visiting <https://datahub.hymarc.org> and accessing the "Data Hub Sandbox" project (click “Projects” in upper right of site). Here, users can create a test dataset to view dataset-level metadata, as well as add a test resource to view resource-level metadata. Note: some metadata is dynamic. For example, if you select "Lab Experimental" or "Historical or Literature" as the dataset Data Source Type, you will see specific resource-level metadata tied to these Data Source Types. Metadata allows users to easily create and refine a search for shared data that will enhance or complete data analysis.

Search

The Data Hub allows you to search on different criteria defined within the metadata, user-defined Tags, as well as within the description of the project, dataset, or resource. You can search for data from the **“Discover”** button on the Home tab and from the **search bar** on the Projects or Data tab. You can search on the file type, to see all CSV files, for example; you can search on keyword or Data Source Type. 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 or Resource 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.**

The image displays two screenshots from the Data Hub interface. The left screenshot shows the 'Datasets' page with a search bar and a faceted search sidebar. The sidebar lists various metadata categories such as Projects, Tags, and Technology Type, each with a count of items. A blue arrow points from the 'Metadata Faceted Search' label to the sidebar. The right screenshot shows the 'Create Dataset' form, which includes fields for Title, Project, Description, and Tags. A blue arrow points from the 'User-defined Tags entered upon Dataset and Resource creation' label to the Tags field, which contains the text 'eg. economy, mental health, government'. Below the Tags field is the 'Dataset Metadata' section with dropdown menus for Institution, Author, and Technology Type, and input fields for Sample Barcode and Collection Date.

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 dataset. During the creation of the dataset and / or adding resource or file, you will be prompted to provide additional information (metadata) that can facilitate understanding, searching, and organizing the data.

Home / Projects

Projects

All Projects

My Projects

1. Click on the Project that will store the data

Add Project

Project Tree

- Advanced Characterization
- Hydrides
- Hydrogen Carriers
- Sorbents
- Admin
- User Resources
 - Data Hub Sandbox



Advanced Characterization

0 Datasets

Research and Development of Advanced Characterization Core Capabilities...



Hydrides

0 Datasets

Research and Development of Advanced Hydrogen Storage Materials: Metal...



Hydrogen Carriers

0 Datasets

Research and Development of Advanced Hydrogen Carrier Materials Recipient...



Sorbents

0 Datasets

Research and Development of Advanced Hydrogen Storage Materials: Sorbents...

Home / Projects / Advanced Characterization

Project

Overview

Datasets

Activity

Administration

Edit Project

Bulk Edit

Add Dataset

Members

Project Tree

- Advanced Characterization

Advanced Characterization

Project ID: eb0ae329-7184-4fa6-9c24-ffe50bd12f65

2. Click on *Add Dataset*, left panel



Research and Development of Advanced Characterization Core Capabilities

Recipient National Renewable Energy Laboratory/NREL (PI: Parilla, Phil) and Lawrence Berkeley National Laboratory/LBNL (PI: Prendergast, David)

Abstract DOE has taken affirmative steps to promote accurate hydrogen sorption measurements by establishing validation facilities and by commissioning the writing of the Recommended Best Practices for the Characterization of Storage Properties of Hydrogen Storage Materials. As HyMARC researchers advance in hydrogen storage and carrier materials development, they will also need to validate their results and qualify their materials and measurement systems, independently. Therefore, there is a continuing need to have world-class materials characterization facilities as the R&D progresses. This project will culminate in a series of new, advanced

1 Create dataset

2 Add data

3. Complete all Fields

Title:

* URL:

Project:

Description:

You can use Markdown formatting here

Tags:

Verify that the set *Project* is correct, you could place the dataset in another project, but make sure you have access to that project.

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

Metadata

Institution:

Author:

* Maintainer Email:

DOI:

Sample Barcode:

Collection Date:

Data Source Type:

Comments:

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

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

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.

Home Projects Data About Help

Home / Projects

Projects

- All Projects
- My Projects
- Administration
- Add Project

Project Tree

- Advanced Characterization
- Hydrides
- Hydrogen Carriers
- Sorbents
- Admin
- User Resources
 - Data Hub Sandbox

Search project

Order by: Name Ascending

1. Click on the Project that will store the data

2. Click on Datasets, left panel

Advanced Characterization
0 Datasets
Research and Development of Advanced Characterization Core Capabilities...

Hydrides
0 Datasets
Research and Development of Advanced Characterization Core Capabilities...

Hydrogen Carriers
0 Datasets
Research and Development of Advanced Characterization Core Capabilities...

Sorbents
0 Datasets
Research and Development of Advanced Characterization Core Capabilities...

Administration
2 Datasets
This project is currently under development.

User Resources
5 Datasets
This project is currently under development.

Advanced Characterization
Project ID: eb0ae329-7184-4fa6-9c24-ffe50bd12f65

Advanced Characterization

NREL Research and Development of Advanced Characterization Core Capabilities

BERKELEY LAB

Recipient National Renewable Energy Laboratory/NREL (PI: Parilla, Phil) and Lawrence Berkeley National Laboratory/LBNL (PI: Prendergast, David)

Abstract DOE has taken affirmative steps to promote accurate hydrogen sorption measurements by establishing validation facilities and by commissioning the writing of the Recommended Best Practices for the Characterization of Storage Properties of Hydrogen Storage Materials. As HyMARC researchers advance in hydrogen storage and carrier materials development, they will also need to validate their results and qualify their materials and measurement systems, independently. Therefore, there is a continuing need to have world-class materials characterization facilities as the R&D progresses. This project will culminate in a series of new, advanced characterization capabilities necessitated by the projected storage and carrier materials in order to establish new validated materials, basic physicochemical properties and applicability of optimal materials for system integration.

Home / Projects / Advanced Characterization

Project ID: 327126aa-76be-4e88-848b-ac91263b08d7

Advanced Characterization

Search datasets...

Order by: Relevance

PUBLIC Sum Frequency Generation Vibrational Spectroscopy (SFG)
 1 Resource
 Sum frequency generation laser spectroscopy method that provides unique opportunities to probe...

PRIVATE Computational Characterization
 2 Resources
 This dataset has no description

There are no Tags that match this search

Lawrence Berkeley N... (1)

1. Click on the dataset to add data to

Home / Projects / Advanced Characterization / Computational Characterization

PRIVATE Email Maintainer

Computational Characterization

Project ID: 327126aa-76be-4e88-848b-ac91263b08d7
 Dataset ID: e2669749-63ec-400c-ac9d-f8ae6d1e70db

Data and Resources

Screening of Porous Sorbents
 Image of Screening of Porous Sorbents
 View Download Edit

Beamline schedule.pdf
 Year in review. 2017 Schedule for the...
 View Download Edit

2. Click *Add New Resource*, left panel

From that point, follow the directions in the [previous section](#), from step 5 and onward.

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. Perhaps you have uploaded the wrong file or noticed a mistake and need to reload the file. Follow the directions below to delete a Resource.

Starting from the Dataset's page

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

2. Click on the target resource 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 HyMARC 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

The screenshot shows the 'Demo: Multi-Spectra Data Tool' dataset page. On the left is a sidebar with 'Data and Resources' selected. The main content area shows a list of resources: 'public-spectra-data1 (2).csv', 'public-spectra-data2 (1).csv', and 'public-data3 (1).csv'. Each resource has 'View', 'Download', and 'Edit' buttons. A blue arrow points to the 'View' button for the first resource.

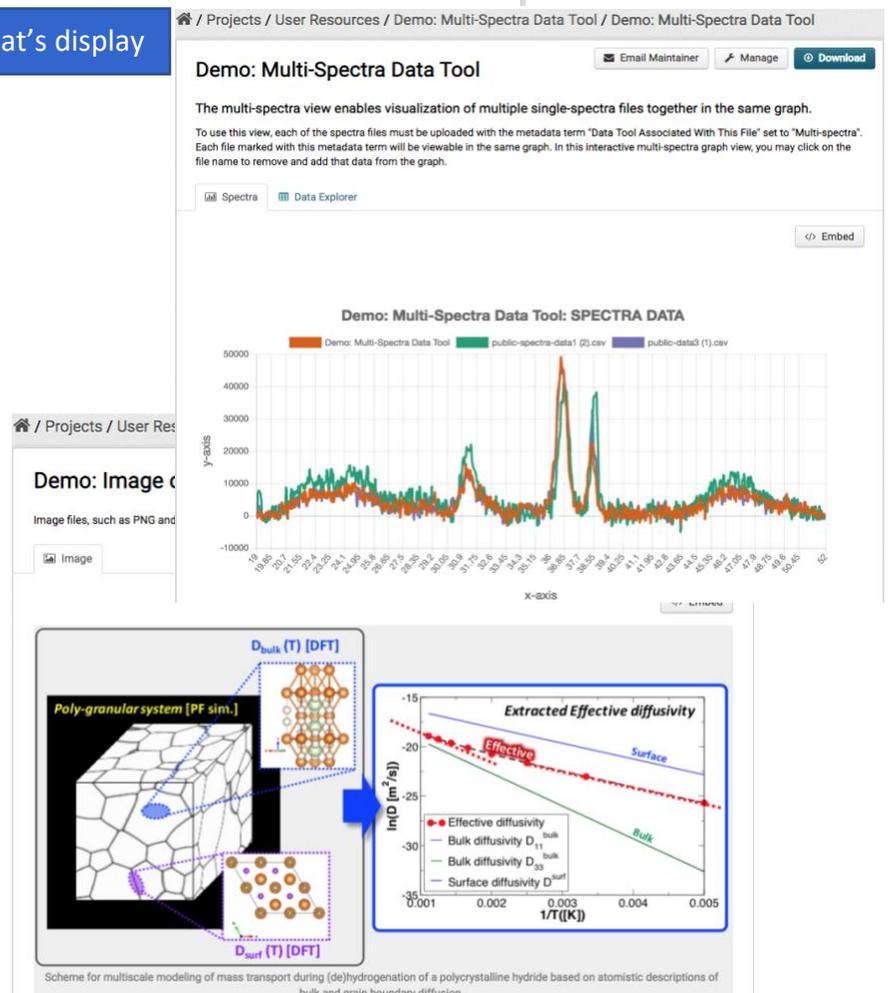
2. View page opens with the format's display

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.

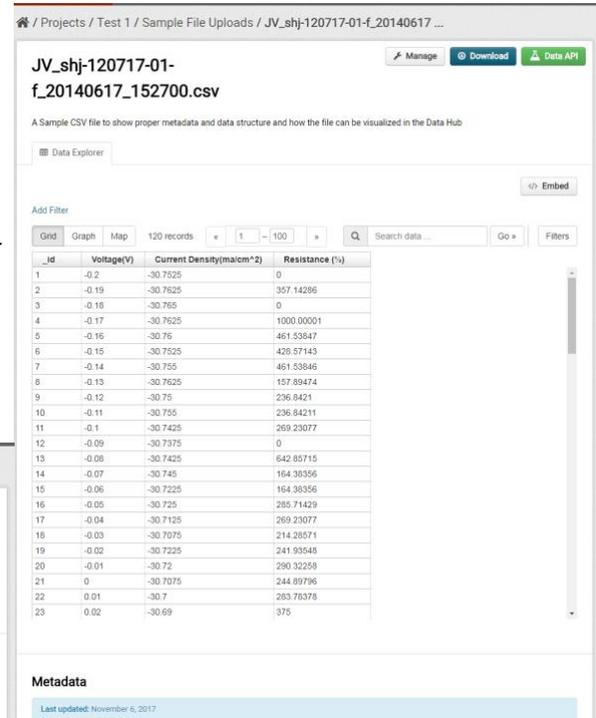
Data API – This green button (not present in picture) will open another dialog box listing the syntax to access this particular data file programmatically ([see below](#)).



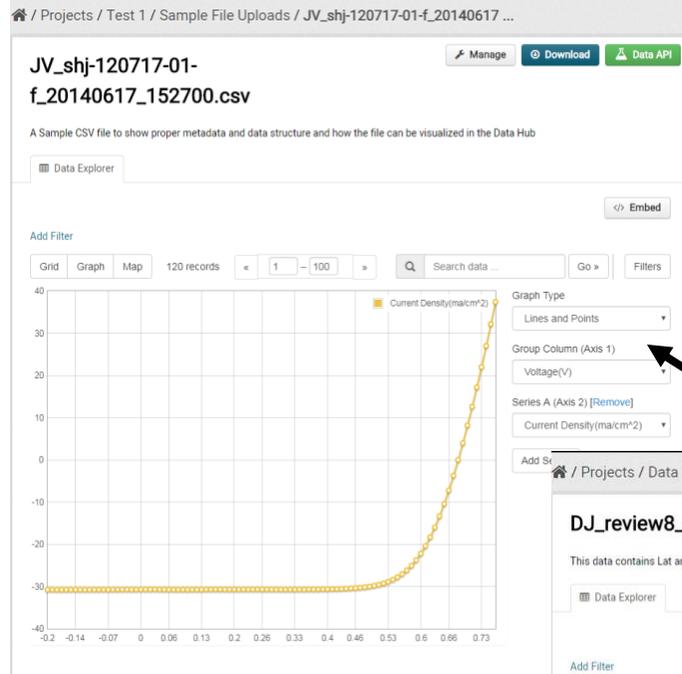
With **CSV files the Data Explorer plugin** will allow for the data to be viewed in three possible ways: **Table, Graph or Map**. To utilize the Graph or Map the data has to lend itself to being viewed in that manner and both will require interaction with the user to display the data as needed. The map 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

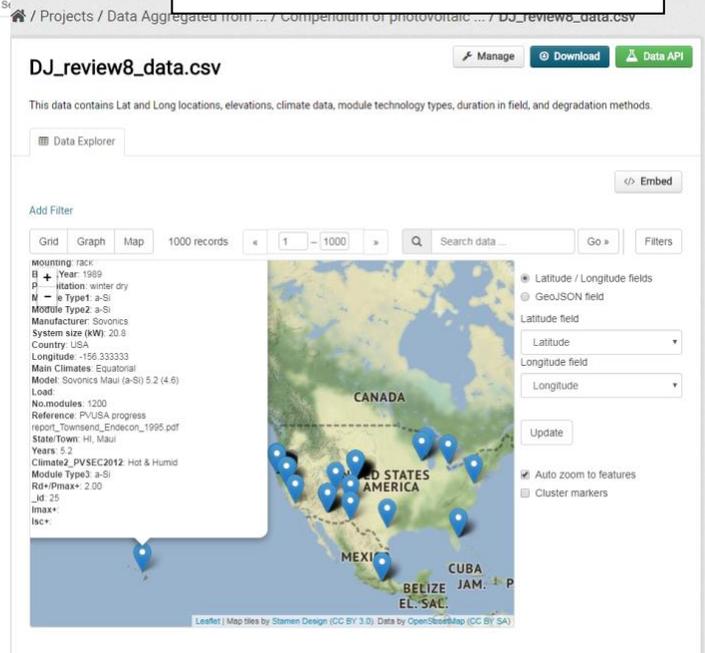


CSV Plot (Graph): the user needs to connect the columns in the table to the axis. This interface is on the right side.



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.

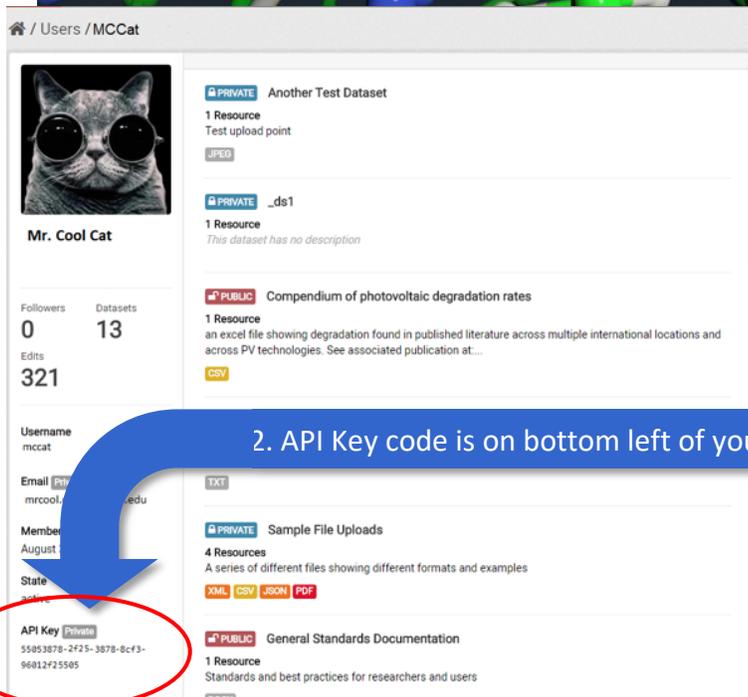
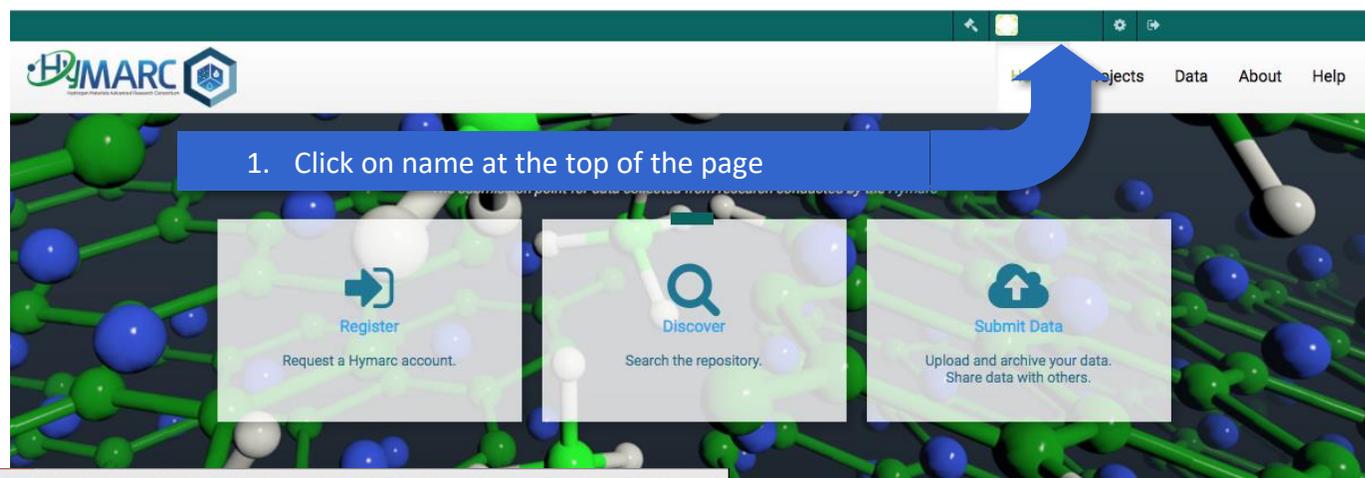
Clicking any of the map markers will display all the data for the record associated with that geographic position.



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. The API can be useful if you need to **upload, 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.). Visit the “User Resources” project, “Help and Tutorial” dataset or the HELP tab on the data hub to review and use a Python notebook that will allow you use the API to make queries against the datasets you have access to or to upload one or more files with metadata. [API Walkthrough Python Notebook](#)

The first step, upon Login, 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 HyMARC 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 \(emnadmin@nrel.gov\)](mailto:emnadmin@nrel.gov).