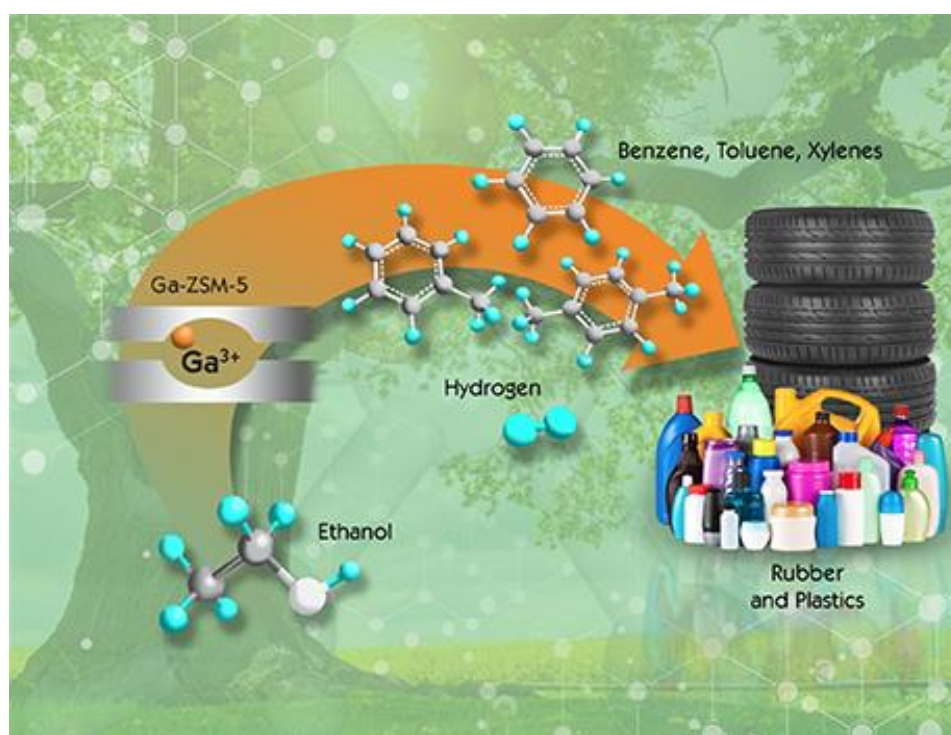


Using the ChemCatBio Data Hub



U.S. DEPARTMENT OF
ENERGY

Energy Efficiency &
Renewable Energy



Energy Materials Network
U.S. Department of Energy

November 20, 2017

1

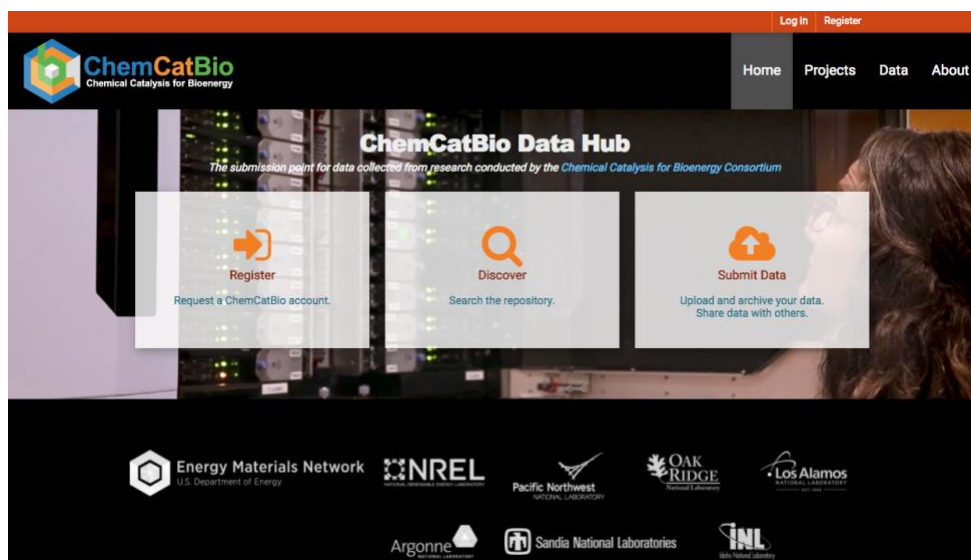
Table of Contents

Using the ChemCatBio Data Hub	1
Overview	3
Registration, Login and Security	3
Projects & Data	5
Datasets, Files & Resources	6
Metadata	6
Search	7
For All Researchers	8
Adding Data	8
Adding Data to an Existing Dataset	12
Deleting a File or Resource	14
Viewing Data	15
Accessing the Data through the API	18
For Project Principle Investigators (PIs).....	19
Adding new members to a project	19
Changing permissions for a member of a project	19
Removing a project member	19
Questions and Feedback?.....	19

Overview

The ChemCatBio 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 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 [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

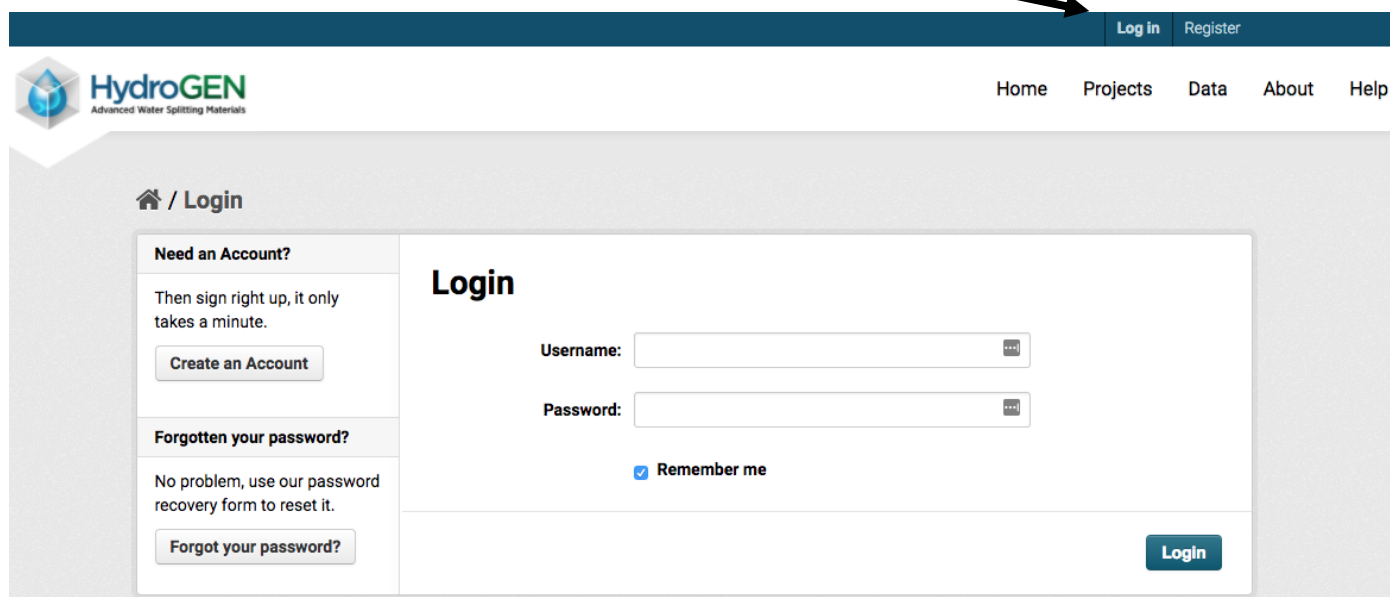
- Registration, Login and Security
- Projects & Data
- Search

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.



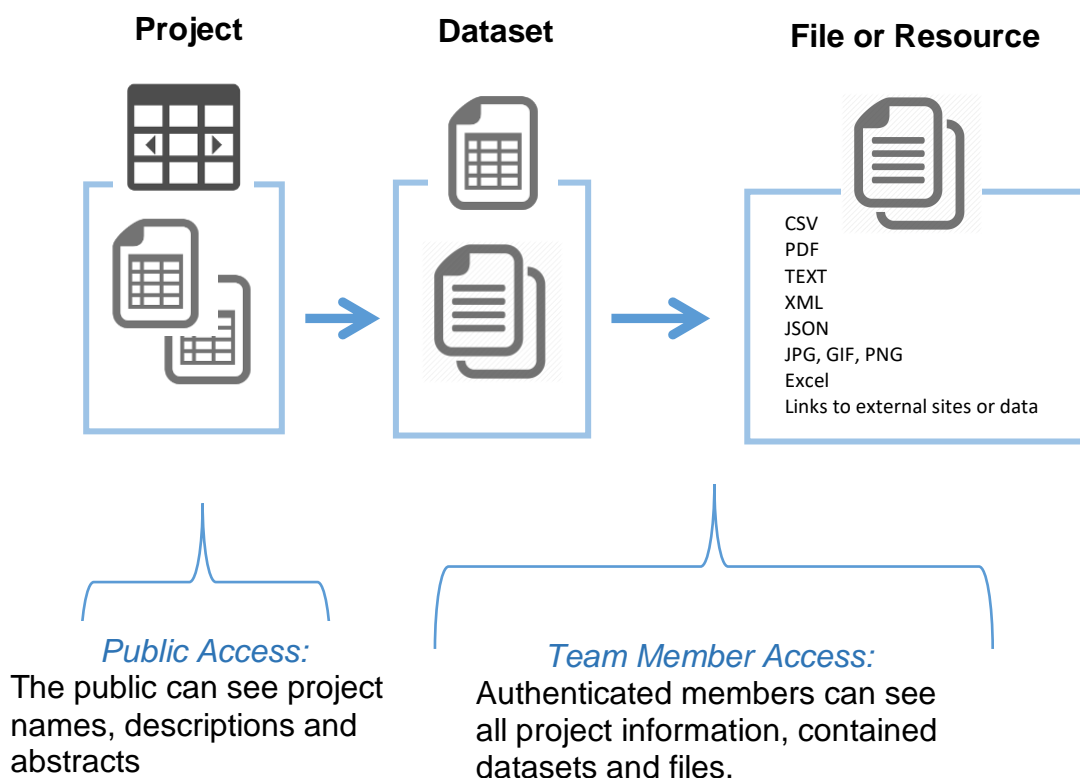
The screenshot shows the HydroGEN website interface. At the top, there is a dark blue navigation bar with 'Log in' and 'Register' buttons. Below this is a white header with the HydroGEN logo (Advanced Water Splitting Materials) on the left and navigation links for 'Home', 'Projects', 'Data', 'About', and 'Help' on the right. The main content area is titled '/ Login'. It contains three sections: 'Need an Account?' with a 'Create an Account' button, 'Forgotten your password?' with a 'Forgot your password?' button, and a 'Login' section with 'Username:' and 'Password:' input fields, a 'Remember me' checkbox, and a 'Login' button.

Once given access to the site, 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 ChemCatBio 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 ChemCatBio specific 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 catalyst 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 (data generator), date of upload, institution, catalyst type, data source type (Advanced Catalyst Characterization, Catalyst Performance Evaluation, Interactive Tools and Databases, Materials Synthesis, Modeling, Techno-Economic and Sustainability Analysis), 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 catalyst type (e.g., Metal Carbide, Metal-modified Zeolite, Metal Oxide - Mixed Metal Oxide, Metal Phosphide, Supported Metal, Zeolite) 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 (see “[Public Tools](#)” project for an example of a public external resource). 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, Catalyst Type (Metal Carbide, Metal-Modified Zeolite, Metal Oxide – Mixed Metal Oxide, Metal Phosphide, Supported Metal, Zeolite), Sample Barcode, Collection Date, Data Source Type (Advanced Catalyst Characterization, Catalyst Performance Evaluation, Interactive Tools and Datasets, Materials Synthesis, Modeling, Techno-Economic and Sustainability Analysis). 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.chemcatbio.org> and accessing the "Data Hub Sandbox" project. 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 "Advanced Catalyst Characterization", "Catalyst Performance Evaluation", or "Materials Synthesis" is selected as the dataset Data Source Type, users 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 catalyst type (Metal Carbide, Metal-Modified Zeolite, Metal Oxide – Mixed Metal Oxide, Metal Phosphide, Supported Metal, Zeolite). 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.**

Metadata Faceted Search

Home / Dataset

Projects

- User Resources (3)
- Test (2)
- Advanced Catalyst (1)
- Public Tools (1)

Tags

- catalyst (1)
- characterization (1)
- CKAN (1)
- graph (1)
- Help (1)
- hydrogenation (1)
- metal carbide (1)
- multiple (1)
- SPE (1)
- Tools (1)

Author

There are no Author that match this search

Institution

- NREL (2)

Catalyst Type

- Metal Carbide (1)

Sample Barcode

Add Dataset

Search datasets...

Order by: Relevance

PRIVATE fod Format View Data Tool
3 Resources
This dataset shows a demonstration of a custom data tool. The data hub has a plug-in architecture, and custom data tools like this one can be developed. In this example, three...

PRIVATE Help and Tutorial
1 Resource
Data hub user guide. The user guide details the functionality of the data hub. To click on the links referenced in the User Guide, you will need...

DOCX

PUBLIC Surface Phase Explorer
1 Resource
Understanding the coadsorption of two species. To facilitate exploration this web tool creates...

PRIVATE Data Tools
2 Resources
This dataset shows out of the box data tools of...

PNG CSV

PRIVATE Metal-modified metal carbide
11 Resources
XRD, TEM

TXT TIFF application/postscript DOCX

PRIVATE datapusher_3

Home / Datasets / Create Dataset

1 Create dataset 2 Add data

Title: eg. A descriptive title
* URL: datahub.h2awsm.org/dataset/<dataset> Edit

Project: Test

Description: eg. ... notes ...
You can use the following formatting here

Tags: eg. economy, mental health, government

Dataset Metadata

* Institution: -- select an option --

* Author: admin

Technology Type: -- select an option --

Sample Barcode:

Collection Date: 0/29/2018

User-defined Tags entered upon Dataset and Resource creation

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

From the Home page, the middle large button, **“Discover”**, will take you directly to the Datasets you have access to. If you are not logged in, you will only see Public Datasets. By default each Dataset is established as private, giving only project members access to the data. At a future date, we will set up procedures for public release of data. If you have a current dataset that needs to be made public, please contact the [administrator](#) and we will help you.

The third large button, **“Submit Data”** will prompt you to login with the credentials you established at the time you registered on the Data Hub. Once you are logged in, you can either navigate to the Project to which you have access to “Add New Resource” or click on the “Submit Data” button from the Home tab to directly create a Dataset under the appropriate Project to which you have access.

/ Projects

☰ Projects

All Projects

My Projects

✎ Administration







Add Project

👤 Project Tree

- Advanced Catalyst
- Atomic Layer Deposition
- Aviation Fuels
- Biochemical Intermediates
- Bioproduct Upgrading
- Bio-Renewable Surfactants
- Catalytic Fast Pyrolysis
- CCM
- Computational Models
- Ethanol Upgrading
- Indirect Liquefaction Intermediates
- Olefins Processes
- Public Tools
- Reduction of CO₂
- Steering Committee
- Terephthalic Acid
- Test
- User Resources
- Value of Gasoline and Fuel Oil

Search project

Order by: Name Ascending

- 
Advanced Catalyst
 1 Dataset
 Advanced Catalyst Synthesis and Characterization Lead PI: Susan Habas,...
- 
Atomic Layer Deposition
 0 Datasets
 Enhanced Catalyst Durability and Sulfur Tolerance by Atomic Layer Deposition...
- 
Aviation Fuels
 0 Datasets
 Tactical Aviation Fuels Lead PI: Andy Sutton, Los Alamos National Lab/LANL...
- 
Biochemical Intermediates
 0 Datasets
 Catalytic Upgrading of Biochemical Intermediates Lead PI: Rick Elander,...
- 
Bioproduct Upgrading
 0 Datasets
 Low Pressure Hydrogenolysis Catalysts for Bioproduct Upgrading Lead PI:...
- 
Bio-Renewable Surfactants

1. Click on the Project that will store the data

/ Projects / Advanced Catalyst

☰ Project

Overview

Datasets

Activity S

✎ Admin

Edit Proj


Bulk Edit Datasets

Add Dataset

Members

👤 Project Tree

- **Advanced Catalyst**



Advanced Catalyst Synthesis and Characterization


Lead PI: Susan Habas, National Renewable Energy Lab/NREL

Co-PI's:

- Ted Krause, Argonne National Lab/ANL
- Kinga Unocic, Oakridge National Lab/ORNL

Abstract The ACSC project addresses key research challenges central to the core catalysis projects by leveraging the unique synthesis expertise and advanced characterization capabilities at multiple DOE National Laboratories and coupling them with computational modeling through the CCPC to facilitate the rational design of catalysts for the core catalysis projects and shorten the cycle of catalyst development across the entire BETO program, leading to innovative catalytic materials for commercial bioenergy applications.

Project Owners

 admin

Created On October 18, 2017, 3:46 PM (UTC-06:00)

Datasets 1

Users 5

 2. Click on *Add Dataset*, left panel

Home / Datasets / Create Dataset

1 Create dataset 2 Add data

Title: Edit
* URL: datahub.chemcatbio.org/dataset/<dataset>

Project:

Description:
You can use Markdown formatting here

Tags:

Dataset Metadata

* Institution:

* Author:

Maintainer Email:

Catalyst Type:

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.

Next: Add Data

3. Complete all Fields on form

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.

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

5. Choose to upload a file or create a link.

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.

Cost Modeling / Add New Resource

1 Create dataset 2 Add data

Data:

Name:

Description:
You can use Markdown formatting here

Format:

This will be guessed automatically. Leave blank if you wish

6. Complete all Fields on the form.

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

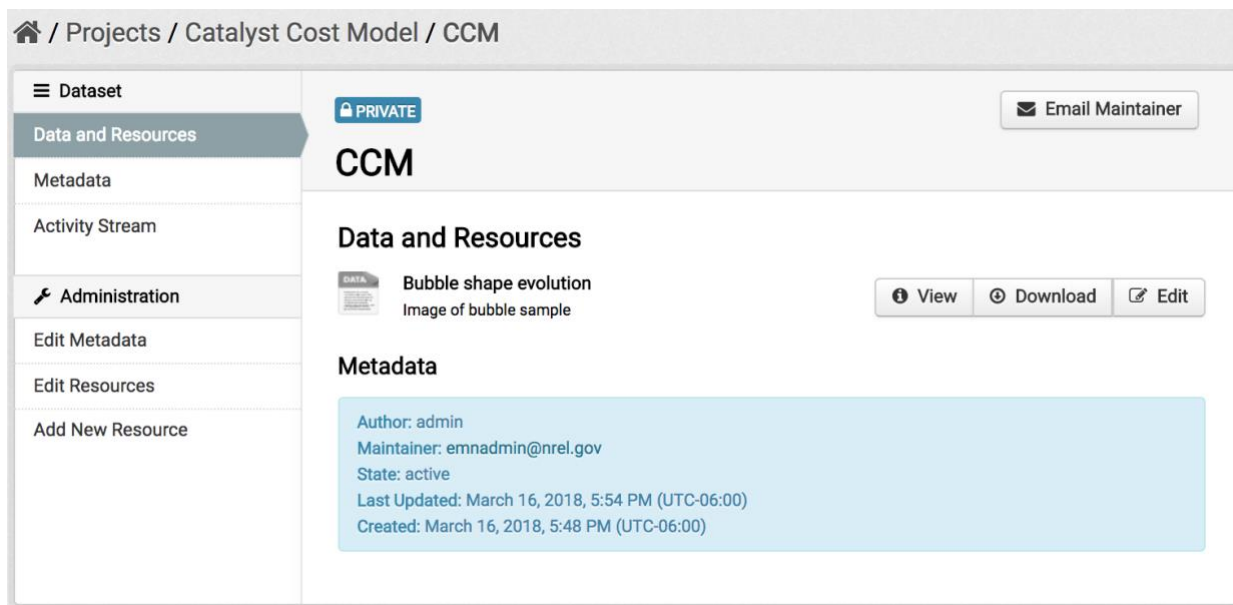
Once everything has been completed using the Finish button, a page for the dataset will appear showing all resources currently within it.

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 *Materials Synthesis 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.

The screenshot displays the ChemCatBio web application interface. At the top, there is a navigation bar with the ChemCatBio logo and the text "Chemical Catalysis for Bioenergy". The main navigation menu includes "Home", "Projects", "Data", and "About". The current page is "Projects".

The left sidebar contains a navigation menu with the following items:

- Projects
- All Projects
- My Projects
- Administration
- Add Project

The main content area shows a search bar for projects and a dropdown menu for "Order by" set to "Name Ascending". Below this, a list of projects is displayed. A blue arrow points to the "Datasets" option in the left panel of a project page, and another blue arrow points to a specific project in the list.

The project page for "Catalyst Cost Model Development" is shown. The left sidebar of this page includes:

- Overview
- Datasets
- Activity Stream
- Administration
- Edit Project
- Bulk Edit Datasets
- Add Dataset
- Members
- Project Tree

The main content area of the project page includes the NREL logo, the project title "Catalyst Cost Model Development", the lead PI "Fred Baddour, National Renewable Energy Lab/ NREL", the co-PI "Lesley Snowden-Swan, Pacific Northwest National Lab/PNNL", and an abstract describing the project's objective. The "Project Owners" section lists "admin" as the owner, with creation details: "Created On October 18, 2017, 4:00 PM (UTC-06:00)", "Datasets 0 Datasets", and "Users 3".

Home / Projects / Catalyst Cost Model

Project

- Overview
- Datasets**
- Activity Stream
- Administration
- Edit Project
- Bulk Edit Datasets
- Add Dataset
- Members
- Tags
- CCM (1)
- library (1)
- Author

Catalyst Cost Model

Order by: Relevance

PUBLIC **Materials Library**

2 Resources

This dataset has no description

PRIVATE **CCM**

1 Resource

This dataset has no description

3. Click on the dataset to add data to

Home / Projects / Catalyst Cost Model / Materials Library

Dataset

- Data and Resources**
- Metadata
- Activity Stream
- Administration
- Edit Metadata
- Edit Resources
- Add New Resource

PUBLIC Email Maintainer

Data and Resources

demo.materials-library.v1.json View Download Edit

A set of three materials.

Demo-Test Materials Library View Download Edit

A set of three materials. Same...

Metadata

Author: admin
 State: active
 Last Updated: March 19, 2018, 9:45 AM (UTC-06:00)
 Created: December 5, 2017, 9:50 AM (UTC-07:00)

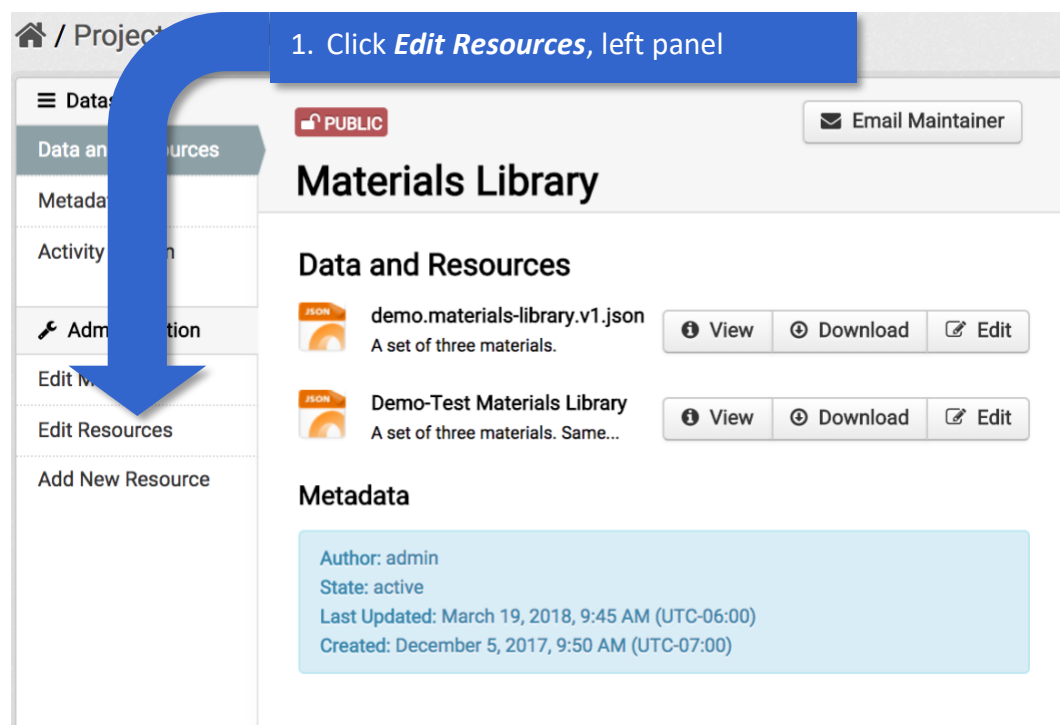
4. Click **Add New Resource**, left panel

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

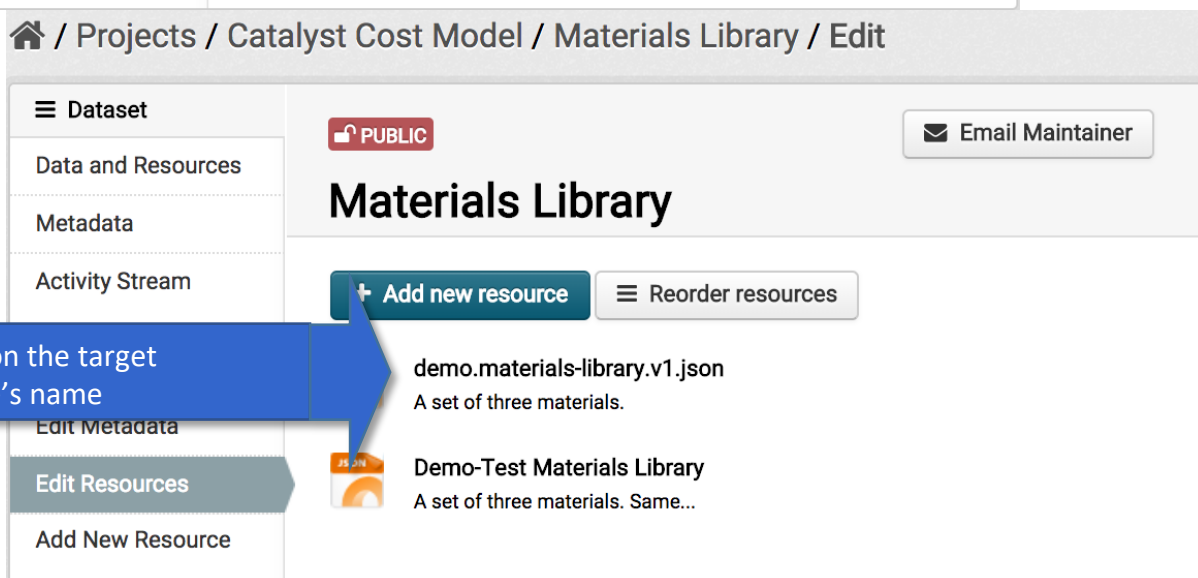
Materials Library

Data and Resources

- demo.materials-library.v1.json
A set of three materials. [View] [Download] [Edit]
- Demo-Test Materials Library
A set of three materials. Same... [View] [Download] [Edit]

Metadata

Author: admin
State: active
Last Updated: March 19, 2018, 9:45 AM (UTC-06:00)
Created: December 5, 2017, 9:50 AM (UTC-07:00)

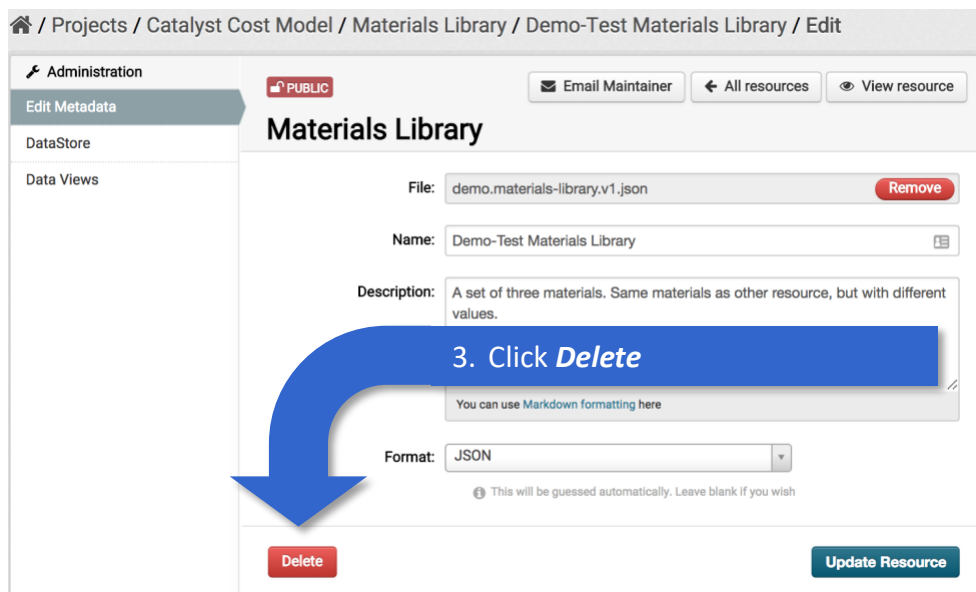


2. Click on the target resource's name

Materials Library

+ Add new resource [Reorder resources]

- demo.materials-library.v1.json
A set of three materials.
- Demo-Test Materials Library
A set of three materials. Same...



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.

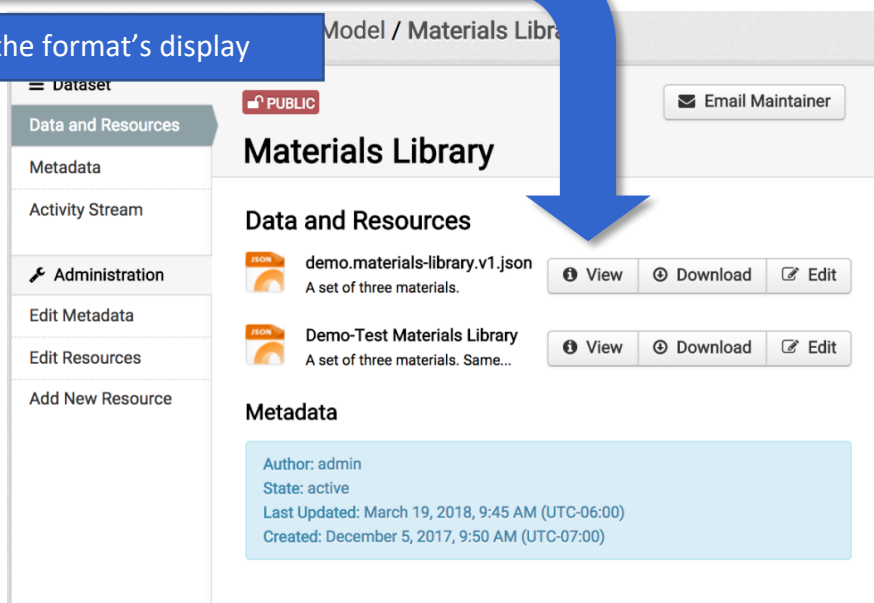
Viewing Data

In many cases the ChemCatBio 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. This is an example of an image file type.

On the right across from the resource name are 3-4 buttons:

Email Maintainer – Opens an email to the person listed as the maintainer of the dataset.

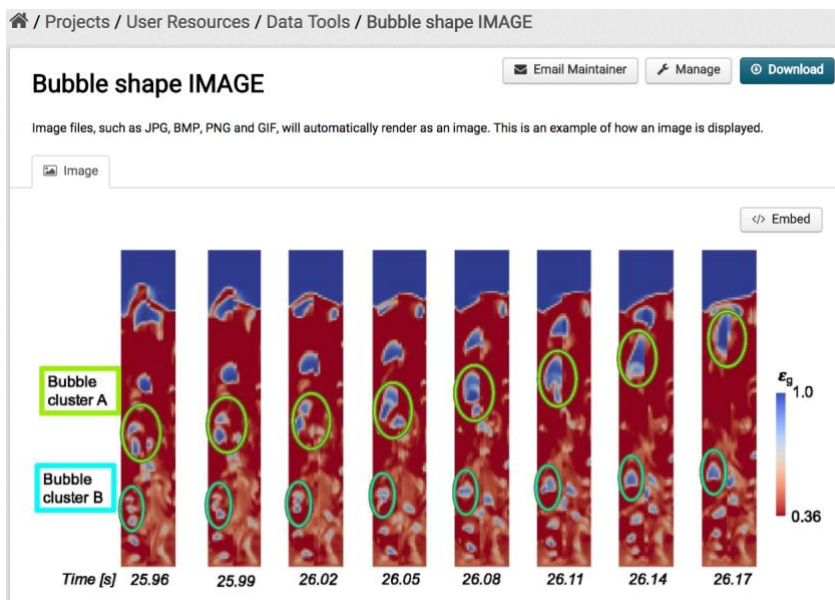
Manage – This allows you to change the metadata for the resource.

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](#)).

CSV Table (Grid): 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.

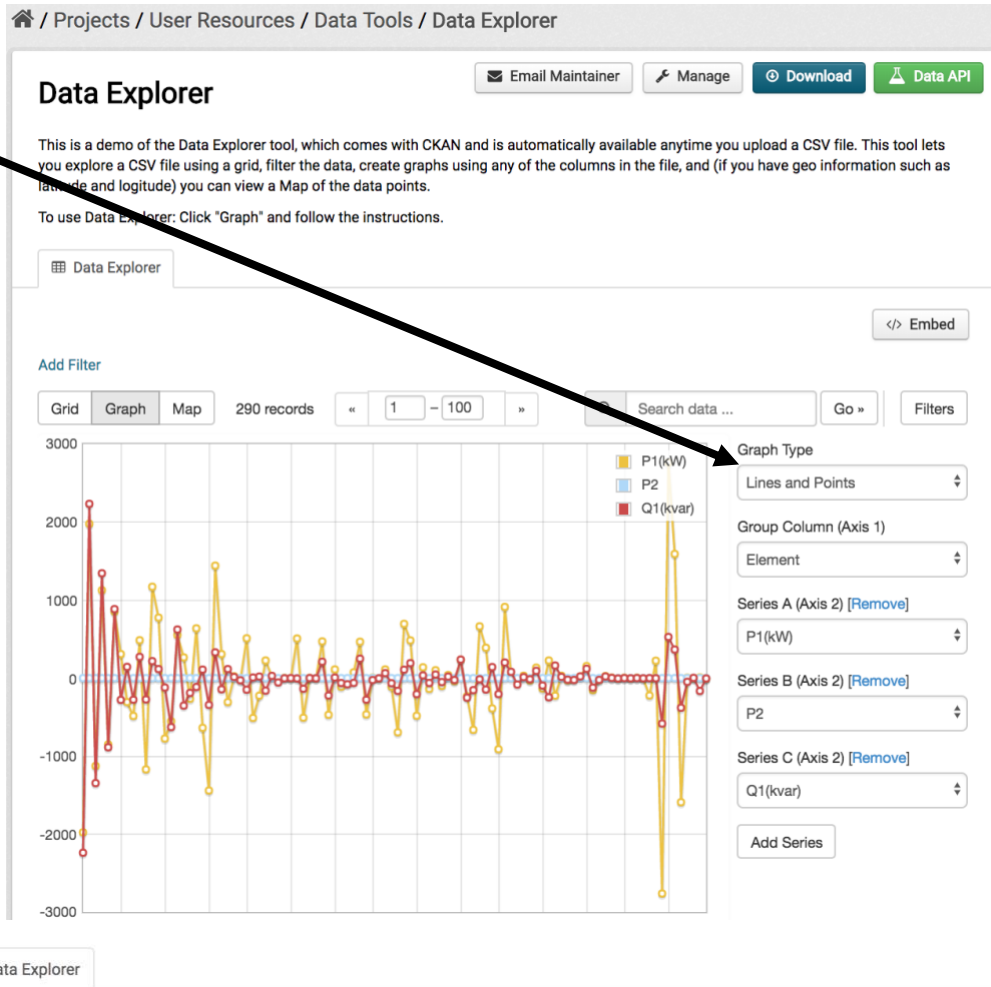
To change the display type for the CSV, click the buttons just above and to the left of the table



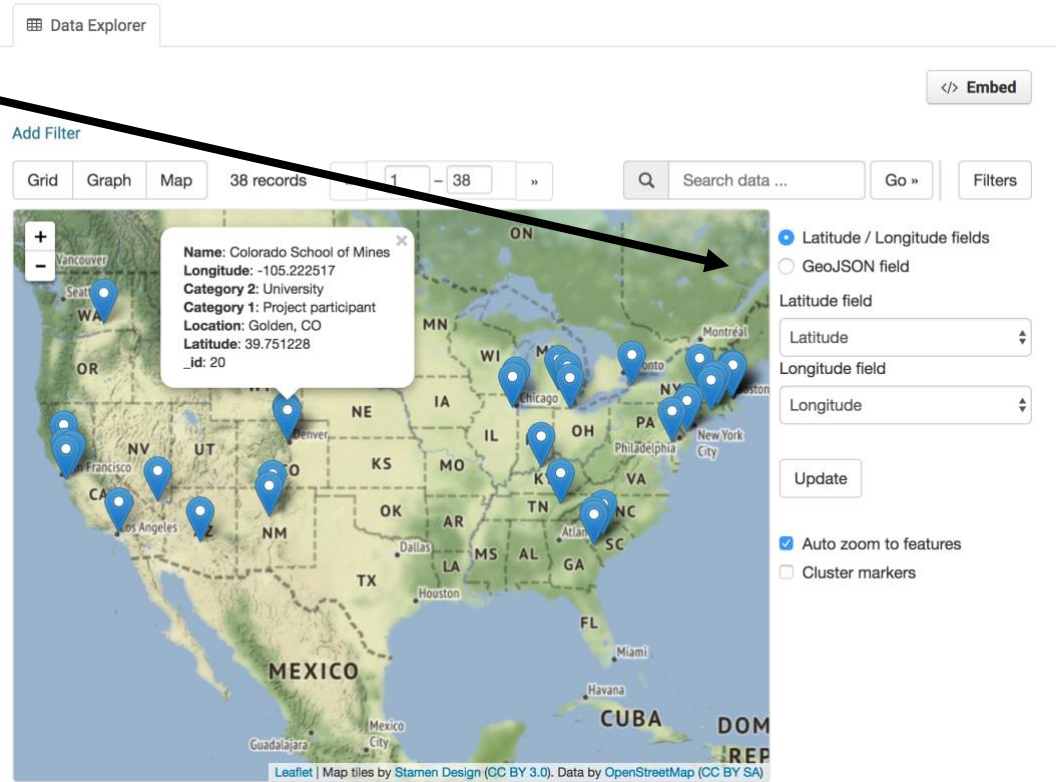
With **CSV files** the **Data Explorer plugin** will allow for the data to be viewed in three possible ways: **Table, Graph or Map**. To utilize the Graph or Map the data has to lend itself to being viewed in that manner and both will require interaction with the user to display the data as needed. The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

_id	Voltage(V)	Current Density(ma/cm ²)	Resistance (%)
1	-0.2	-30.7525	0
2	-0.19	-30.7625	357.14286
3	-0.18	-30.765	0
4	-0.17	-30.7625	1000.00001
5	-0.16	-30.76	461.53847
6	-0.15	-30.7525	428.57143
7	-0.14	-30.755	461.53846
8	-0.13	-30.7625	157.89474
9	-0.12	-30.75	236.8421
10	-0.11	-30.755	236.84211
11	-0.1	-30.7425	269.23077
12	-0.09	-30.7375	0
13	-0.08	-30.7425	642.85715
14	-0.07	-30.745	164.38356
15	-0.06	-30.7225	164.38356
16	-0.05	-30.725	285.71429
17	-0.04	-30.7125	269.23077
18	-0.03	-30.7075	214.28571
19	-0.02	-30.7225	241.93548
20	-0.01	-30.72	290.32258
21	0	-30.7075	244.89796
22	0.01	-30.7	283.76378
23	0.02	-30.69	375

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



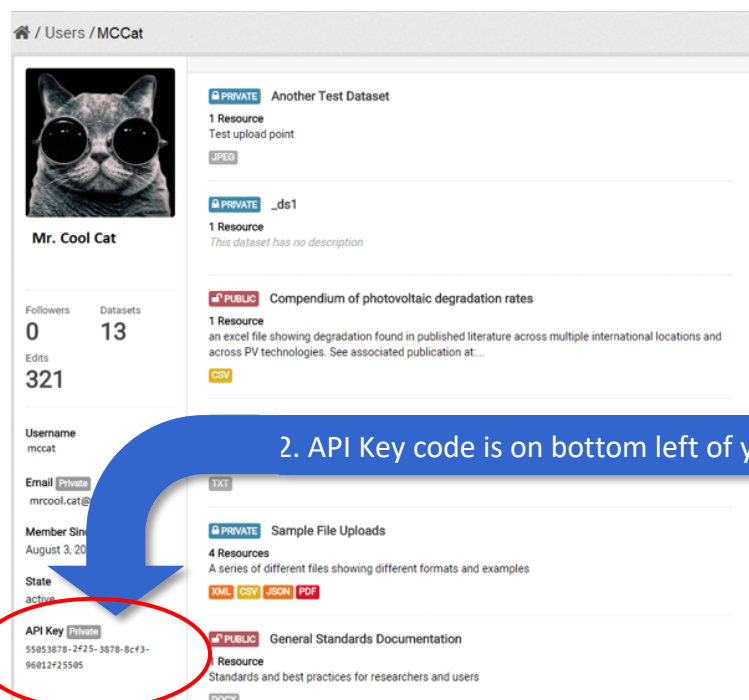
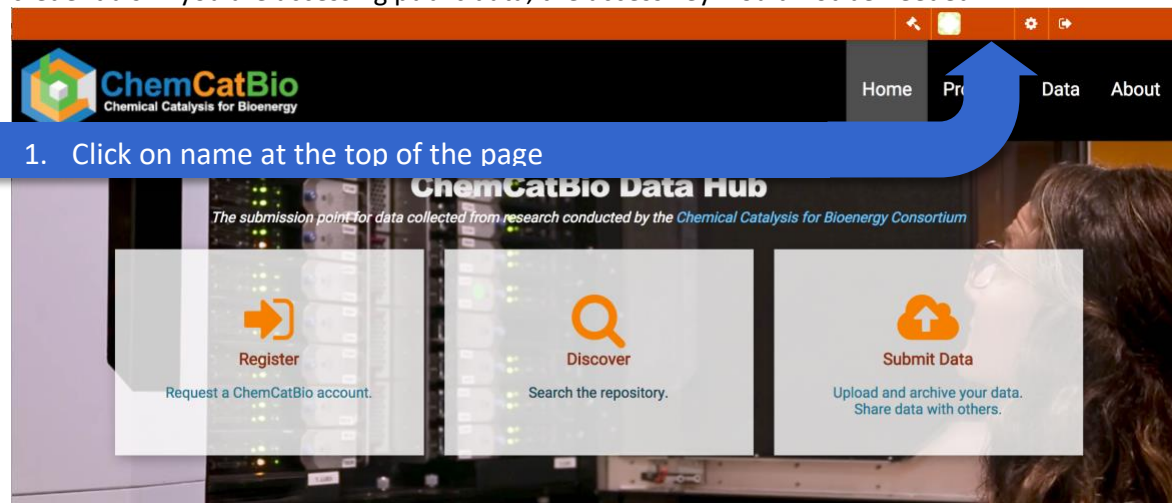
CSV Map: The Data Explorer will automatically detect if the file has Latitude and Longitude header names and use those by default for the geographic coordinates. Clicking any of the map markers will display all the data for the record associated with that geographic position.



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 ChemCatBio 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.



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](#).