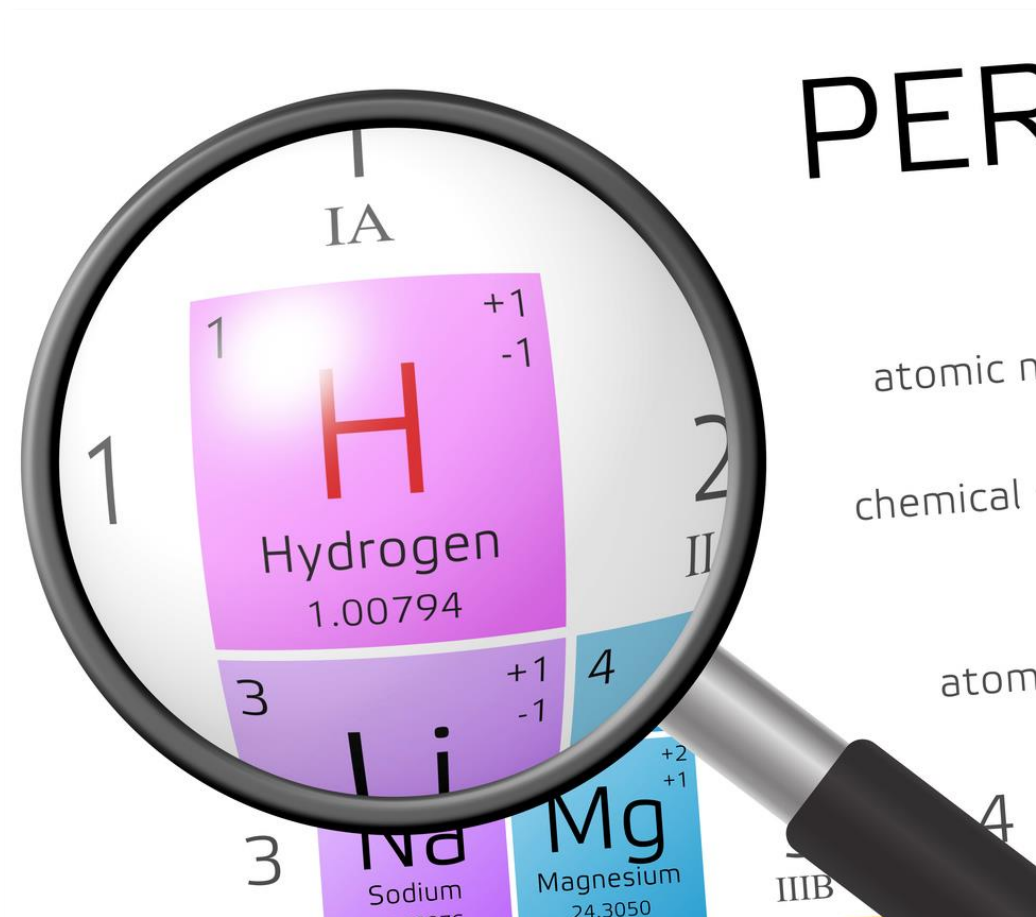


Using the HydroGEN Data Hub



HydroGEN
Advanced Water Splitting Materials



U.S. DEPARTMENT OF
ENERGY

Energy Efficiency &
Renewable Energy



Energy Materials Network
U.S. Department of Energy

November 20, 2017

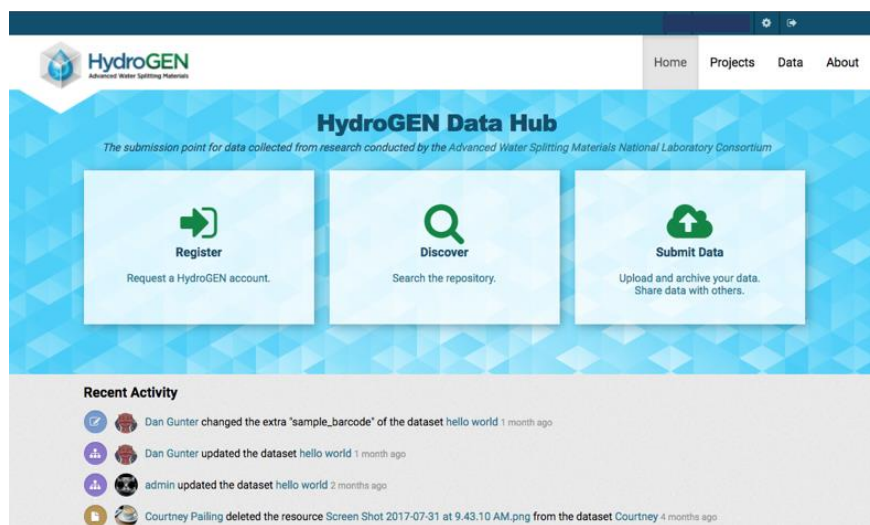
Table of Contents

Using the HydroGEN 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	18
Questions and Feedback?.....	18

Overview

The HydroGEN 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

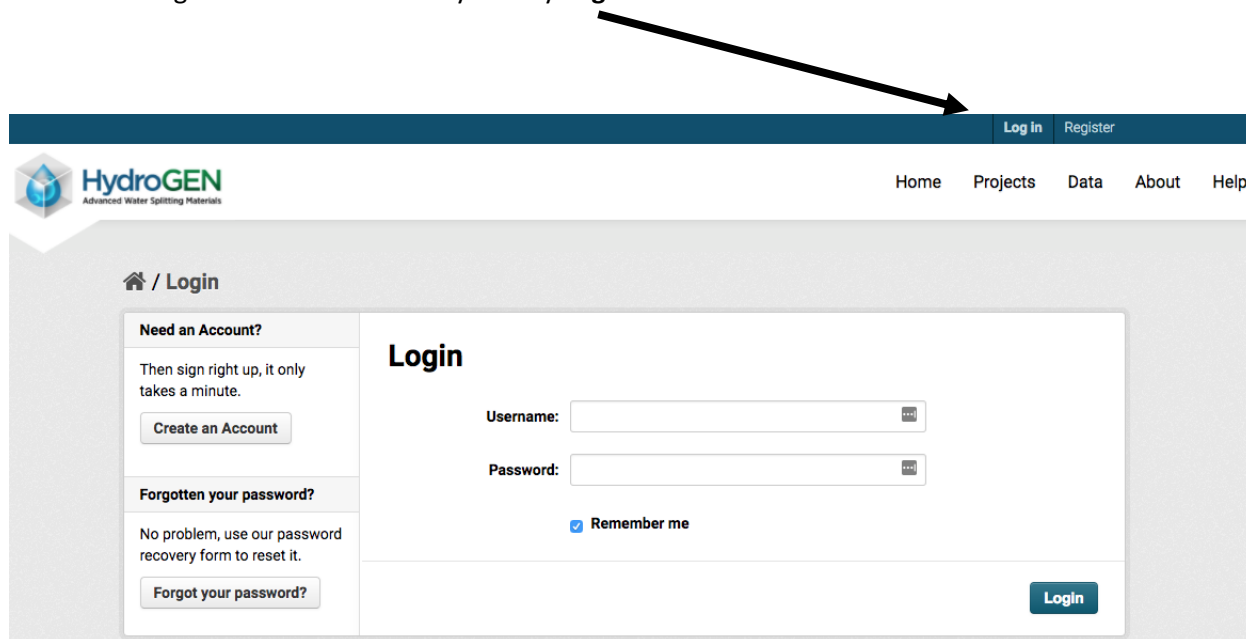
- Registration, Login and Security
- Projects and Data
- Search

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.

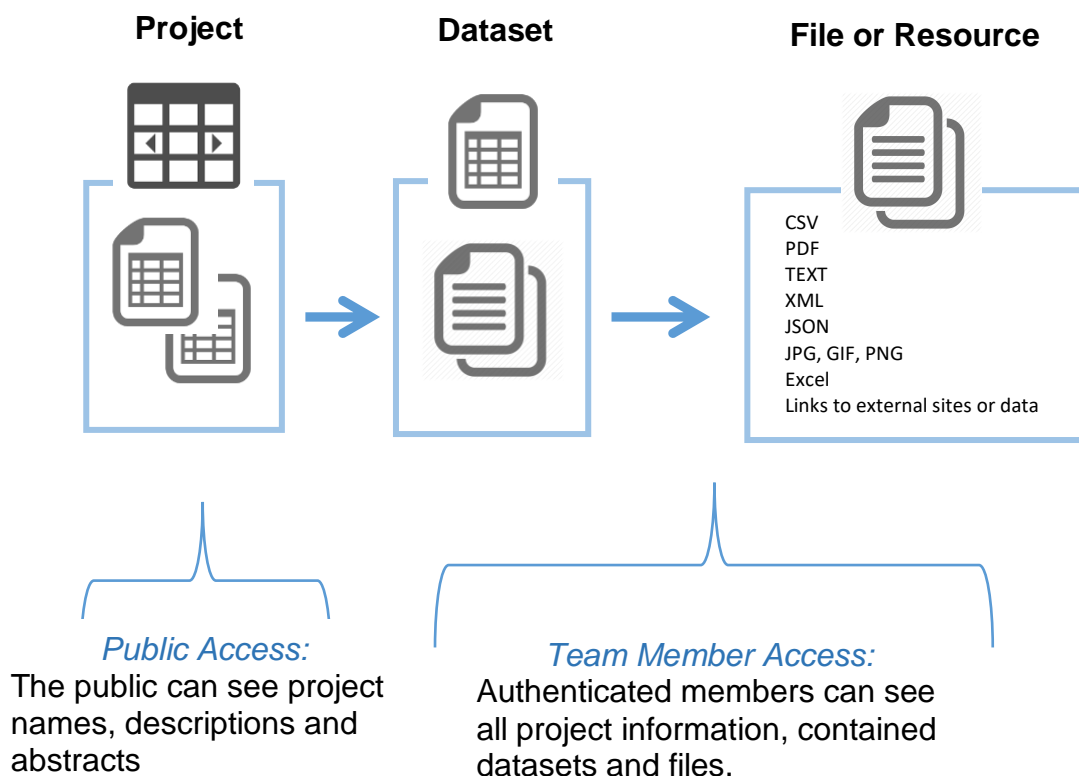


You can request additional projects by contacting the administrator (emnadmin@nrel.gov).

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 HydroGEN 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 HydroGEN 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 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 (data generator), date of upload and update, institution, technology type (HTE, HTE-SOEC, LTE, LTE-AEME, LTE-PEME, PEC, STCH, STCH-Hybrid), 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, Technology Type (HTE, HTE-SOEC, LTE, LTE-AEME, LTE-PEME, PEC, STCH, STCH-Hybrid), 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.h2awsm.org> and accessing the "API 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 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 technology type (HTE-SOEC, LTE, LTE-AEME, LTE-PEME, PEC, STCH, STCH-Hybrid). 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. A blue arrow points from the sidebar to the search bar, labeled 'Metadata Faceted Search'. The right screenshot shows the 'Create Dataset' form, with a blue arrow pointing to the 'Tags' field, labeled 'User-defined Tags entered upon Dataset and Resource creation'. The 'Tags' field contains the text 'eg. economy, mental health, government'. Below the 'Tags' field is the 'Dataset Metadata' section, which includes fields for 'Institution', 'Author', 'Technology Type', '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 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.

HydroGEN
Advanced Water Splitting Materials

Home Projects Data About

/ Projects

1. Click on the Project that will store the data

Project Tree

- ANL PGM Free
- ASU Perovskites
- Benchmarking
 - HTE
 - LTE
 - PEC
 - STCH
- CSM STCH
- GWE Hybrid S
- Hydrogen General
- HydroGEN Media
- LANL Membrane
- LANL PEC
- LANL PGM Free
- NEU PGM Free
- NWU SOEC
- NWU STCH
- Proton LTE
- Rutgers PGM Free
- Stanford PEC

Order by: Name Ascending

Argonne ANL
0 Datasets
PGM Free... Catalysts for PEM Electrolyzer Recipient Argonne National...

ASU
0 Datasets
Materials Characterization and Benchmarking for Advanced Water Splitting Technologies: Best Practices in...

Benchmarking
0 Datasets
Benchmarking Advanced Water Splitting Technologies: Best Practices in...

HTE
0 Datasets
Subsets of Benchmarking

/ Projects / Benchmarking

2. Click on *Add Dataset*, left panel

Project

Overview

Datasets

Activity Stream

Administration

Add Dataset

Project Tree

- Benchmarking
 - HTE
 - LTE
 - PEC
 - STCH

Benchmarking

PROTON ENERGY

ASU
Arizona State University

Caltech

Pacific Northwest NATIONAL LABORATORY

Benchmarking Advanced Water Splitting Technologies: Best Practices in Materials Characterization

Recipient Proton Energy Systems, Inc (PI: Ayers, Kathy)

Subs Caltech (PI: Xiang, C.X.), Arizona State University/ASU (PI: Stechel, Ellen), and Pacific Northwest National Laboratory/PNNL (PI: Holladay, Jamie)

Water Splitting Technology All

Status Awarded

Abstract Topic 2B: HydroGEN - Development of Best Practices in Materials Characterization and Benchmarking for Advanced Water Splitting Technologies

Project Owners

admin

Created On September 25, 2017, 4:24 PM (UTC-06:00)

Datasets 0 Datasets

Users 11

Home / Datasets / Create Dataset

1 Create dataset 2 Add data

Title:

* URL: datahub.h2awsm.org/dataset/<dataset> [Edit](#)

Project:

Description:

You can use [Markdown formatting here](#)

Tags:

Metadata

* **Institution:**

* **Author:**

* **Maintainer Email:**

Technology 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](#).

* Required field

[Next: Add Data](#)

3. Complete all Fields

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.

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 *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.

The screenshot displays the HydroGEN web application interface. At the top left is the HydroGEN logo with the tagline "Advanced Water Splitting Materials". The top navigation bar includes "Home", "Projects", "Data", and "About". The main content area is titled "Projects" and features a search bar, a "Order by" dropdown set to "Name Ascending", and a list of projects. The "ANL PGM Free" project is highlighted, showing "0 Datasets" and a description: "PGM-free OER Catalysts for PEM Electrolyzer Recipient Argonne National...".

Two blue arrows with text annotations indicate the steps for adding data:

1. Click on the Project that will store the data
2. Click on *Datasets*, left panel

The second screenshot shows the "Benchmarking" project page. The left sidebar has a "Project" menu with "Overview", "Datasets", "Activity Stream", "Administration", and "Add Dataset". The "Project Tree" shows "Benchmarking" expanded with sub-items "HTE", "LTE", "PEC", and "STCH". The main content area displays project details for "Benchmarking Advanced Water Splitting Technologies: Best Practices in Materials Characterization", including logos for PROTON, ASU, and Caltech, recipient information, and project owners.

Home / Projects / Benchmarking

Project

- Overview
- Datasets**
- Activity Stream
- Administration
- Add Dataset

Tags

- LTE (2)
- Benchmarking (1)
- original (1)
- PEC (1)

Technology Type

Benchmarking

Search datasets...

Order by: Relevance

PRIVATE Project
1 Resource
External tools list

PRIVATE Background
1 Resource
Overview of Benchmarking origins.

3. Click on the dataset to add data to

Home / Projects / Benchmarking / Background

Dataset

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

PRIVATE

Data and Resources

powerstep31 (1).csv

View Download Edit

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

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.

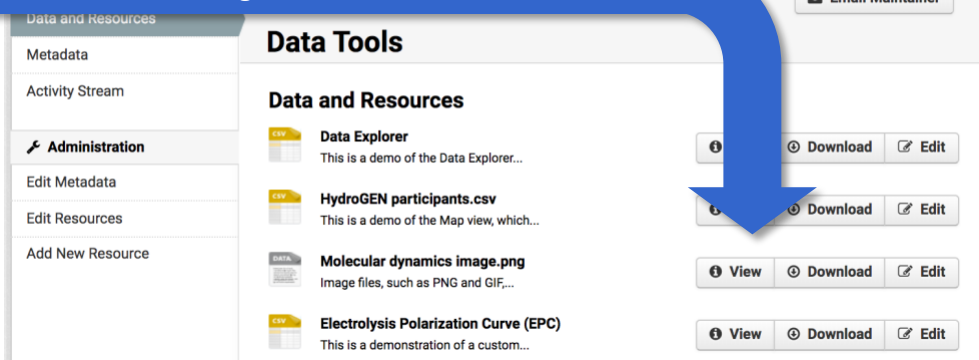
Viewing Data

In many cases the HydroGEN 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

Home / Projects / User Resources / Data Tools

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



The screenshot shows the 'Data Tools' interface. On the left is a sidebar with 'Administration' selected. The main area lists four resources: 'Data Explorer', 'HydroGEN participants.csv', 'Molecular dynamics image.png', and 'Electrolysis Polarization Curve (EPC)'. Each resource has a 'View', 'Download', and 'Edit' button. A blue arrow points from the 'View' button for 'Molecular dynamics image.png' to the next screenshot.

2. View page opens with the format's display

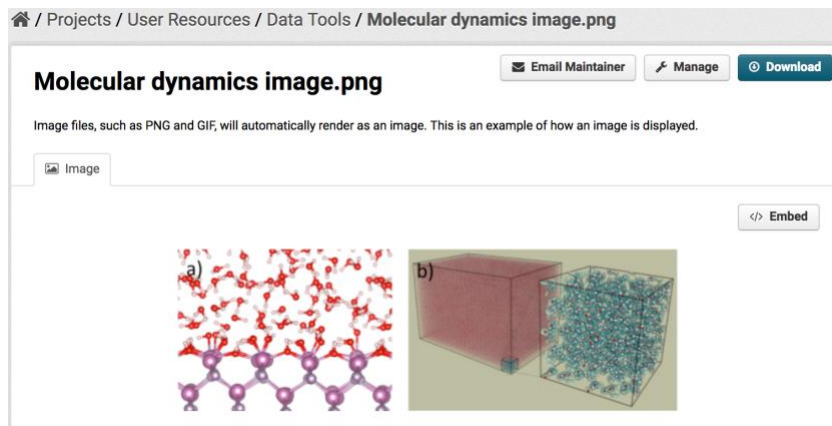
If the format has been predefined it will open that within the viewing engine. Additional view types can be created as plugins and tied into this same system.

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](#)).

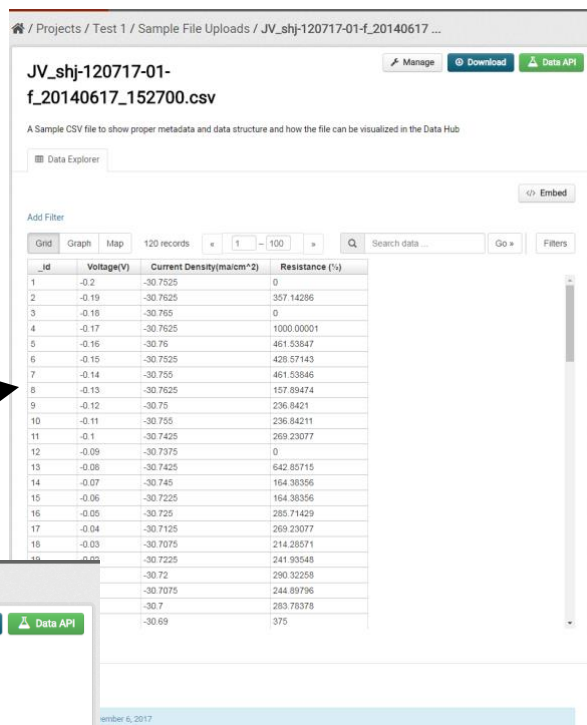


The screenshot shows the 'Molecular dynamics image.png' view page. It features a 3D molecular model of a crystal lattice structure, labeled 'a)' and 'b)'. The model consists of red and purple spheres connected by bonds. A blue arrow from the previous screenshot points to the 'View' button for this resource.

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.

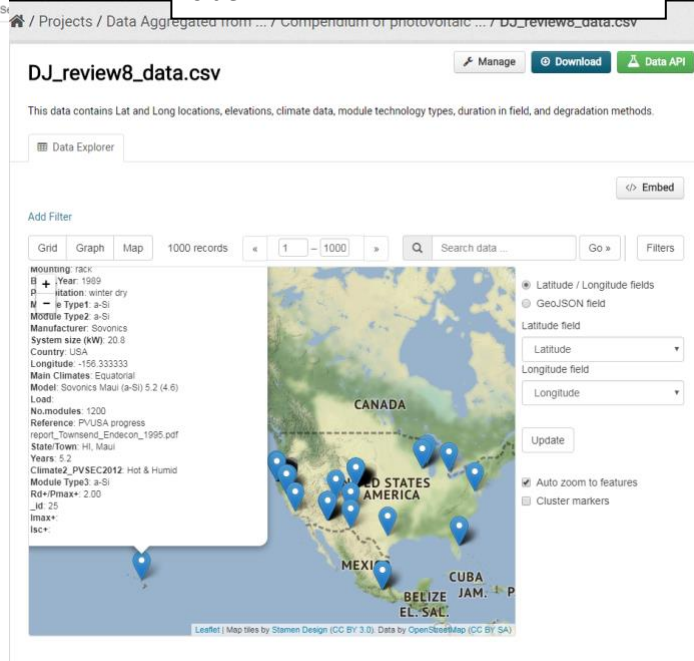
To change the display type for the CSV, click the buttons just above and to the left of the table



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.

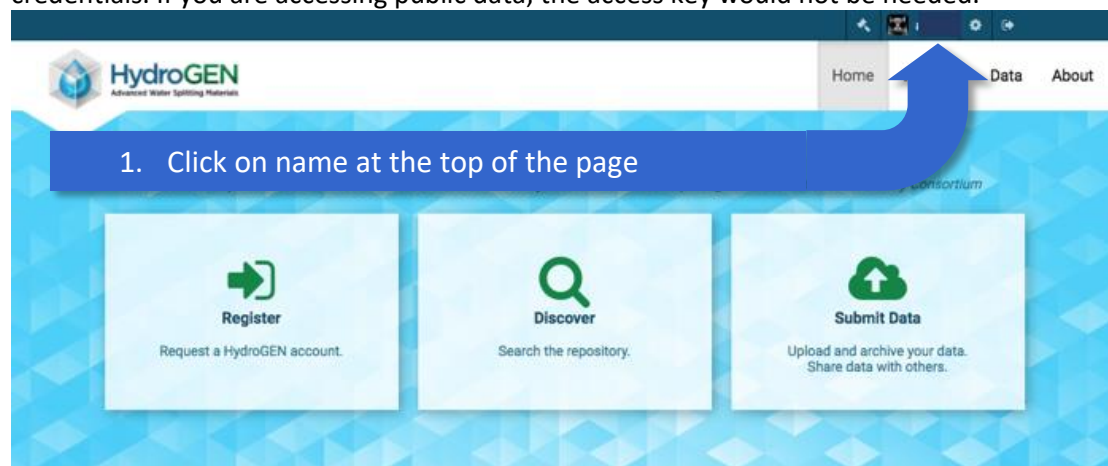


Accessing the Data through the API

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.). 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 data you have access to.

[API Walkthrough Python Notebook](#)

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.



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

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 \(emnadmin@nrel.gov\)](mailto:emnadmin@nrel.gov).