



Data Hub Tutorial

<https://datahub.h2awsm.org>

emnadmin@nrel.gov



Agenda

1. Intro
2. What is the Data Hub and Why are we using it?
3. Data Hub Basics - slides
4. How to Upload Data - Tutorial
5. User Resources and Data Tools
6. Making Data Public
7. Q&A



HydroGEN Data Team

The Data Team meets monthly on data and Data Hub topics. Topics that intersect with your research, your data management and how the Data Hub can help.

Rick Karnesky – SNL

Tony McDaniel – SNL

Dan Gunter – LBL

Nem Danilovic – LBL

Adam Weber – LBL

Yong Han – LLNL

Tadashi Ogitsu – LLNL

Hector Colon-Mercado – SRNL

James Vickers – DOE HQ

Gary Groenewold – INL

Robert Kinoshita – INL

Richard Boardman – INL

Hanping Ding - INL

Huyen Dinh – NREL

Kristin Munch – NREL

Nick Wunder - NREL

Courtney Pailing- NREL

Nalinrat Guba – NREL

Ashlee Vise - NREL



What is the Data Hub?

- Secure data sharing for team members
- Create datasets and upload files
- Search across all data using defined metadata
- Make selected datasets public
- Centralized Authentication

- Link to other repositories or databases
- HydroGEN Data Team
- Data plug-ins for visualization and analyzing data
- Application Programming Interface (API)
- Data Hubs are in active development

[Log in](#) [Register](#)



[Home](#) [Projects](#) [Data](#) [About](#) [Help](#)

HydroGEN Data Hub

The submission point for data collected from research conducted by the Advanced Water Splitting Materials National Laboratory Consortium



Register

Request a HydroGEN account.



Discover

Search the repository.



Submit Data

Upload and archive your data.
Share data with others.



Why use the Data Hub?

The Data Hub was established under the Consortium to host non-proprietary results and data resulting from awarded projects, in a publicly available HydroGEN data portal for the purpose of advancing a broader understanding of material systems for sustained high performance water splitting.

- All data uploaded is private by default.
- Can make datasets public. Data Release Procedure.
- Can upload data to collaborate with your dispersed project team.
- DOE requires each HydroGEN project to use the Data Hub.



What is the Data Hub

- Web application + API (Application Programming Interface)
- CKAN Infrastructure shared among all (5) EMN Data Hubs.
- Data Management Structure
 - **Project:** Permissions applied at the Project; Public project description
 - **Datasets:** Can be made public; create as many Datasets as you need to.
 - **Resources:** File or link to external dataset; upload many files in a Dataset.
- Metadata for Curation
 - Allows you to tag your data and for searching across all data you have access to.
- User Resources https://datahub.h2awsm.org/project/about/user_resources
 - Data Tool demos, API Walkthrough File Type Views, Data Hub User Guide



Tutorial

Tutorial

<https://datahub.h2awsm.org>



Data Management Structure

Projects

Search projects...

Order by: Name Ascending

Project Tree

- ANL PGM Free
- ASU Perovskites
- Benchmarking
 - PEC
 - STCH
 - LTE
 - HTE
- CSM STCH
- GWE Hybrid S
- HTE Supermode
- HydroGEN Media
- LANL Membrane
- LANL PEC
- LANL PGM Free
- LTE Supermode
- NEU PGM Free
- NSF DMREF CSM STCH
- NSF DMREF PSU LTE
- NSF DMREF PSU PEC
- NSF DMREF UB PEC
- NWU o-SOEC
- NWU STCH
- OER Supermode
- PEC Supermode
- Proton LTE
- Rutgers PGM Free
- Stanford PEC
- STCH Supermode
- St. Gobain o-SOEC
- Supplementary Data
- Test
- UCB STCH
- UCORNN p-SOEC
- UH Chalcopyrites
- UM PEC
- User Resources
 - API Sandbox
 - UTRC p-SOEC
 - WVU p-SOEC

ANL PGM Free
PGM-free OER Catalysts for PEM Electrolyzer Recipient Argonne National...

ASU Perovskites
Mixed Ionic Electronic Conducting Quaternary Perovskites: Materials by Design...

Benchmarking
Benchmarking Advanced Water Splitting Technologies: Best Practices in...
Sub-projects:
PEC
STCH
LTE
HTE

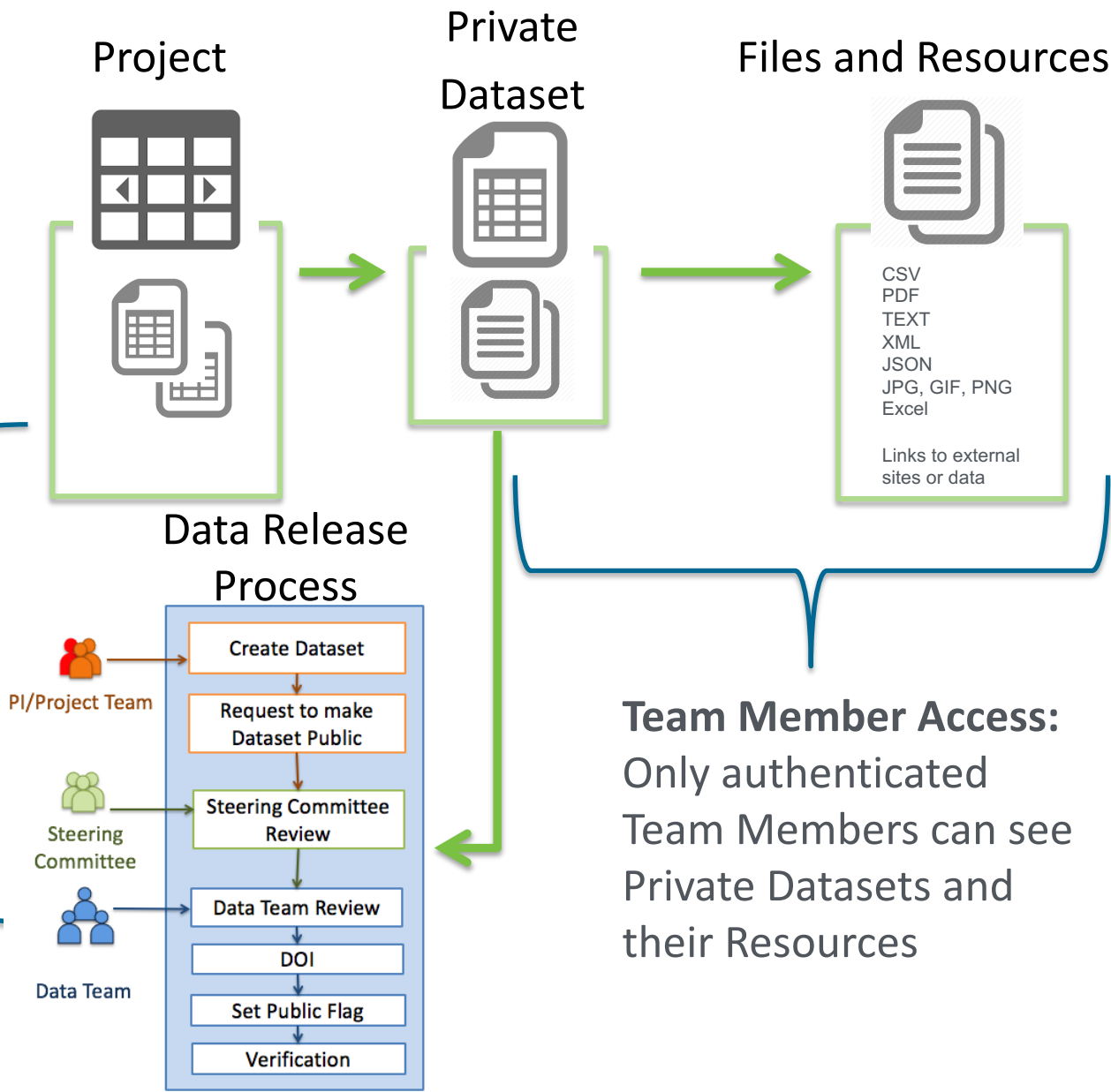
CSM STCH
Accelerated Discovery of Solar Thermochemical Hydrogen Production Materials...

GWE Hybrid S
High Temperature Reactor Catalyst Material Development for Low Cost and...

HTE Supermode
Supermode IDEA: Supermode Capability to Characterize HTE Electrode...

HydroGEN Media
This project hold various media files cross-cutting to the HydroGEN consortium.

Public Access:
The public can see Project names, descriptions and abstracts. **Public Datasets** are also accessible.





Sharing Data

3. Enter Metadata

1. Choose the Project

Home / Projects

Projects

- All Projects
- My Projects**

Project Tree

- CMU Cathode
- User Resources

My Projects

Search projects...

Order by: [Dropdown]

CMU Cathode
0 Datasets
Advanced PGM-free Cathode Engineering for High Power Density and Durability...

User Resources
1 Dataset
A collection of resources for registered data hub users. Examples of site...

2. Add a Dataset, or choose existing Dataset

Project

- Overview
- Datasets
- Activity Stream
- Administration**
- Edit Project
- Bulk Edit Datasets
- Add Dataset

Dataset Metadata

Institution: Idaho National Laboratory

Capability Node

Capability Nodes:

- INL Electrode and Electrolyte Elevated Temperature Water
- INL Hydrided Material Performance
- NREL Multi-Scale Thermochemical and Electrochemical Modeling

Author: ndanilovic

Maintainer Email: robert.white@nrel.gov

DOI: 10.11532/1495059

Technology Type: HTE-oSOEC

Sample Barcode:

Sample Name:

Collection Date: 4/5/2018

Data Source Type: Lab Experimental

Comments:

Lab Experimental

Measurement Types:

4. Upload a file, or create a Link

Data: Upload Link

Name: eg. January 2011 Gold Prices

Description: Some useful notes about the data

You can use Markdown formatting here

- All data is private, shared only with the project team by default.
- Datasets can be made public by request.



Dataset Metadata

When creating a new Dataset, you will be prompted for metadata:

- **Tag:** user defined metadata, searchable and faceted
- **Institution:** choose institution associated with the Dataset
- **Author:** populated with your username
- **Maintainer E-mail:** as contact for questions
- **Capability Node:** multi-select
- **Technology Type**
- **Sample metadata** (if applicable)
- **Collection Date:** date this data was collected
- **Data Source Type:**
 - External Data or Tools
 - Historical or Literature
 - Lab Experimental
 - Modeling and Simulation

Dataset Metadata

Author	slany
Maintainer Email	Stephan.Lany@nrel.gov
DOI	10.17025/1532370
Institution	National Renewable Energy Laboratory
Capability Node	NREL FPMT for Advanced Water Splitting Pathways
Technology Type	STCH
Data Source Type	Modeling and Simulation
Sample Barcode	
Sample Name	
Collection Date	06/25/2019
Comments	
Measurement Types	
Measurement Type Other	None
Author	slany
Updated	July 10, 2019, 9:50 AM (UTC-06:00)
Created	May 21, 2019, 1:04 PM (UTC-06:00)



Adding Resources to a Dataset

A dataset can have as many files or link to external resources as makes sense to the uploader.

- Dataset metadata can pertain to many files.
- You may use the [API Sandbox](#) to test the functionality of the Data Hub with non-sensitive data.

There are 3 ways to upload data to a dataset

1. Through the web interface you will upload one file at a time or “Save and Add Another” when you are creating the dataset and uploading files at the same time.
 - First identify or create the dataset.
 - **To add more files to an existing dataset, click on the dataset then “Add New Resource”**
 - There are 2 levels of metadata: 1 at the dataset and 1 at the individual file or link you upload.
2. Through the Command Line Tool, “EMN Multiple File Uploader” – for Windows users only
 - Download the .exe file and follow the prompts.
 - [Review the details for this tool](#) and the various functions it performs.
3. Through the API or application programming interface.
 - If you are an API user, someone with a little programming experience, please review the [API walkthrough](#) with all of the information you will need to access data with the API.



Searching Data

🏠 / Datasets

Projects

- Benchmarking (1)
- HTE (1)
- LTE (1)
- PEC (1)
- STCH (1)

Tags

- STCH (2)
- Benchmarking (1)
- HTE (1)
- LTE (1)
- PEC (1)
- Questionnaire (1)
- Survey (1)
- workshop (1)

Institution

- Proton Energy Systems Inc. (2)
- Arizona State University (1)
- CalTech (1)
- Pacific Northwest National Laboratory (1)

Capability Node

There are no Capability Node that match this search

Technology Type

- HTE (1)
- LTE (1)
- PEC (1)
- STCH (1)

benchmarking

5 datasets found for "benchmarking" Order by: Relevance

2018 STCH Benchmarking Questionnaire Summary

2 Resources
EMN STCH Benchmarking Questionnaire Summary: Includes background and motivation, respondent demographics and summary of STCH questionnaire responses.

PDF

2018 LTE Benchmarking Questionnaire Summary

2 Resources
EMN LTE Benchmarking Questionnaire Summary: Includes background and motivation, respondent demographics and summary of LTE questionnaire responses.

PDF

2018 HTE Benchmarking Questionnaire Summary

2 Resources
EMN HTE Benchmarking Questionnaire Summary: Includes background and motivation, respondent demographics and summary of HTE questionnaire responses. The survey questionnaire was...

PDF

2018 PEC Benchmarking Questionnaire Summary

2 Resources
EMN PEC Benchmarking Questionnaire Summary: Includes background, motivation and summary of PEC questionnaire responses.

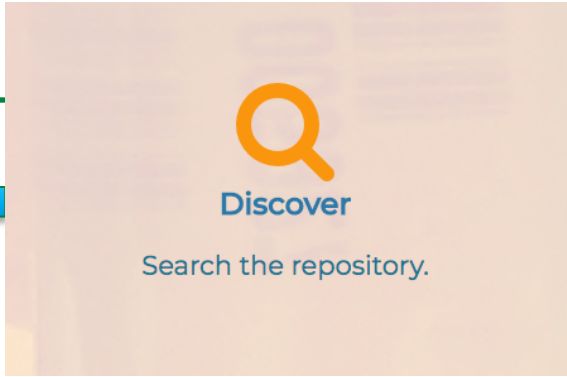
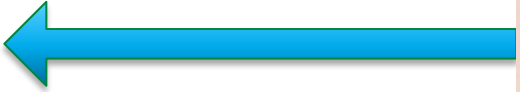
PDF

2018 Water Splitting Technologies Benchmarking and Protocols Workshop

7 Resources
The benchmarking team held a workshop for the advanced water splitting technologies within the EMN on October 24-25 at Arizona State University, in Tempe, AZ. Several breakout...

PDF

You can also access this registry using the API (see [API Docs](#)).



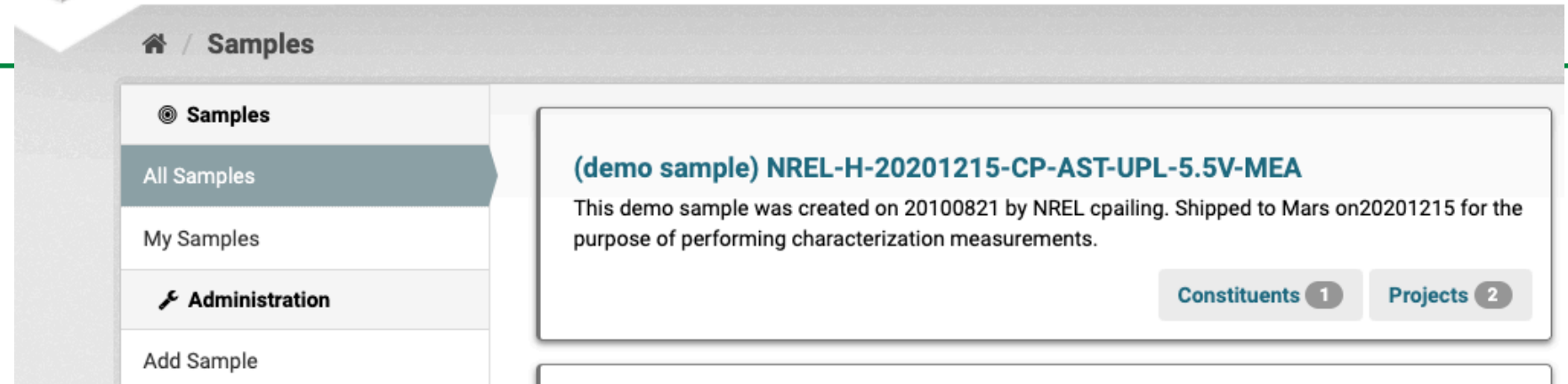
Search across all the data that you have access to

Search Bar – search for any word in the Dataset name, description, tags or metadata

- Order results – e.g. Last Modified
- Facets update for further narrowing

Left Navigation is a faceted search and count of matching datasets:

- Your Projects
- Tags
- Metadata - each metadata term shows a count of the matching datasets



- Uniquely identify samples of different types, and associate these samples with uploaded resources on the Data Hub.
- By creating a sample and completing the form, there is a new sample record to associate to a project and to add to data.
- Constituents are a subset of other samples; the other samples that make up the new sample.
- Sample Tracker is a collection of metadata specific to samples that can be associated with a project/dataset/resource (PDR) such that sample metadata is defined only once rather than duplicated as metadata in each PDR.
- Users may create new samples, view a list of all samples, and edit samples they have created. Access to the sample tracker UI and API is controlled using the same authentication as the data hubs.
- A sample record consists of a name, description (markdown format), type, constituents (other sample records in the database), project, and a field for user defined metadata (JSON format).



Updating or Deleting a File within a Dataset

With proper access, a team member can replace an existing file within a dataset or delete the file entry entirely. To **update a file**, “**Remove**” a file, upload the new file and “**Update Resource**” –if the new file name or format has changed, you would need to update those details in the form (type a new name and or type in a new file format) before you “Update Resource”.

To **delete a file entirely**, you would need to “**Remove**” the file and also “**Delete**” the file entry (3rd blue square in step #2).

Step 2: Update the file (or delete the file entry entirely):

Step 1: Find the project and the dataset, then the particular file you want to update with the dataset. Click “Edit”:

Projects / Admin / test file / readme / Edit

Administration

Edit Metadata

DataStore

Data Views

test file - Data and Resources

Project ID: 58156f9e-4abb-4ad9-aa10-13b006fbac21

Dataset ID: 95a31b25-12c5-46e9-99f2-d3072991b29e

readme File

View Download Edit

There is no description for this resource

File test file

Email Maintainer

Project ID: 58156f9e-4abb-4ad9-aa10-13b006fbac21

Dataset ID: 95a31b25-12c5-46e9-99f2-d3072991b29e

File: readme.csv Remove

Name: readme

Description: Some useful notes about the data

You can use [Markdown formatting here](#)

Format: CSV

This will be guessed automatically. Leave blank if you wish

Resource Metadata

Data Tool Associated With This File: -- select an option --

Data Source: -- select an option --

Data Source Other: Data Source Other

Measurement Type: -- select an option --

Measurement Type Other: Measurement Type Other

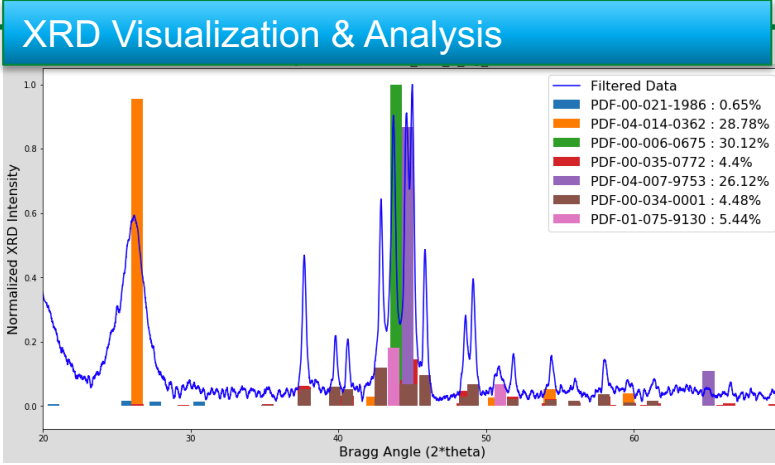
Resource Comments: Resource Comments

Delete Update Resource



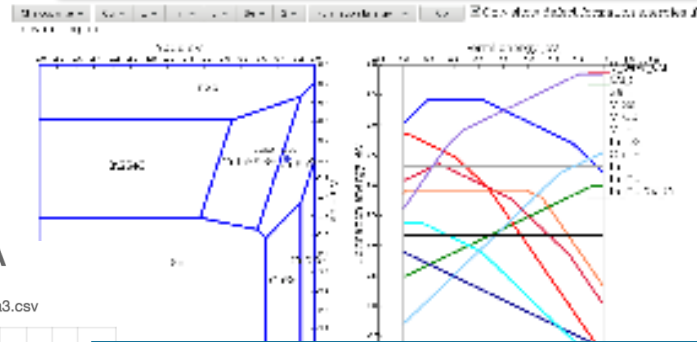
Data Tools

Generic Data Tools: The Data Hub comes with a tools for visualizing common data files and structured data, such as CSV files, pdfs, images.

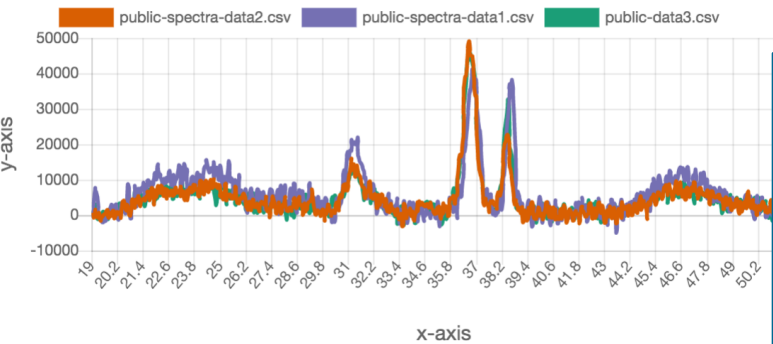


Custom Data Tools: The Data Hub supports custom “plug-ins” for HydroGEN-specific visualizations.

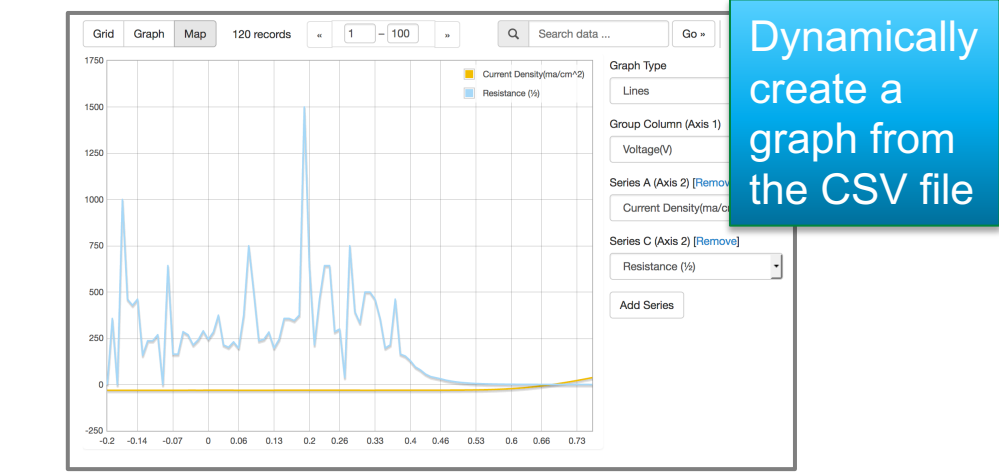
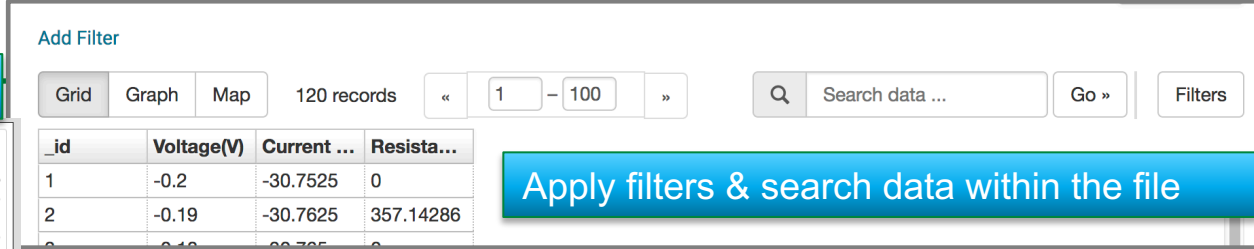
Phase Stability & Defect Properties



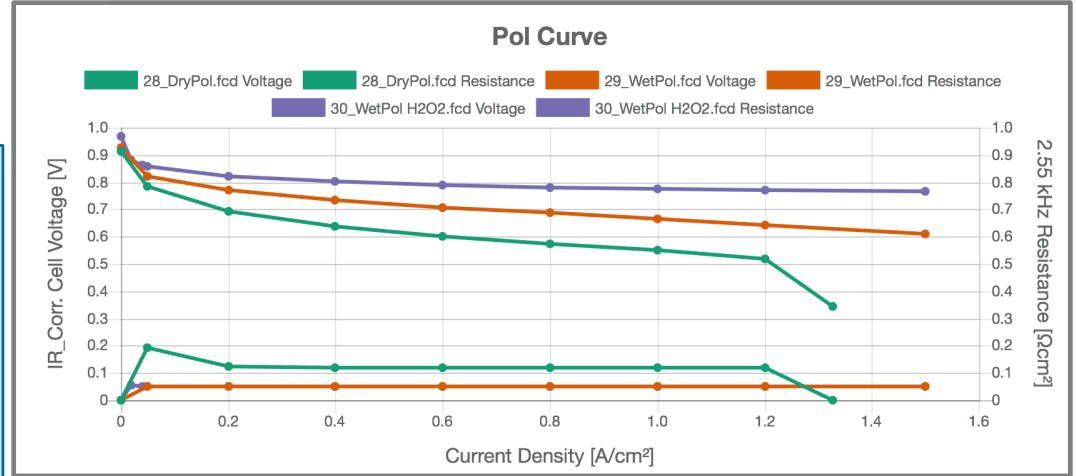
Demo: Multi-Spectra Data Tool: SPECTRA DATA



Visit
“User Resources”
 project for demos of
 existing tools.



Compare data from multiple files on the same plot





EMN Data Hub FAQ

When do I add data to the data hub?

- When you have data related to a Consortium project.
- When you need to share data with your team.
- A phase of your work is complete, and data can be made public.
- New publication – create a dataset and a DOI for your publication.

Why add data?

- Secure data sharing among project team members.
- Advanced search across all data (you have permissions to) using defined metadata.
- Facilitate access to advanced data tools for analysis.
- Make selected datasets publicly available.
- Fulfills DOE's requirement for establishing a HydroGEN data resource/public repository for data resulting from the projects.
- DOI can be used to reference public data by publication, even if the data is not included in the published paper.

Ask us to request a DOI for you!

Who does this benefit?

- You, your team, the project.
- Public, Researchers - Future use – i.e. Machine Learning

How to make it useful for my team (or for public).

- Raw data with description of what the data is and how to use it.
- Ensure the best data tool/data view is applied (if applicable).
- Add a summary of the data, processed data.
- Add an image of the visualization.

What are the data tools that make sense?

- This depends on your data. Do you have image files, csv, data for an X, Y plot?
- Data from multiple materials can be plotted and compared.
- View demos of current [data tools](#) and discuss your data needs with us.



Summary

- **We are here to help you use the Data Hub.**
- **Let's make public data contributions to the community.**
- **Data is private by default.**
- **Upload many files into 1 dataset.**
- **Data Release Procedure.**

Q&A