



HyMARC Data Hub Tutorial



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?

- **Each HyMARC project is required by DOE to use the Data Hub.**
- Secure data sharing among project teams.
- Advanced search across all data using defined metadata.
- Facilitate access to advanced data tools.
- Make selected datasets publicly available.
- Fulfills DOE's requirement for establishing a HyMARC data resource/public repository for data resulting from the projects.
- DOI can be requested for your data and used to reference public data by publication, Ask us to request a DOI for you!

Who does this benefit?

- You, your team, the project.
- **Public HyMARC data will benefit the community into the future!**
- Public, 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/upload link to your paper.
- See the [Data Release Procedure](#) for additional guidance.

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



HyMARC Data Team

Team of lab, data representatives and the Data Hub development team.

- Mark Allendorf, SNL
- Rick Karnesky, SNL
- Chitra Sivaraman, PNNL
- Brandon Wood, LLNL
- Yong Han, LLNL
- Pragya Verma, LBNL
- Tom Gennett, NREL
- Caleb Phillips, NREL
- Courtney Pailing, NREL
- Nick Wunder, NREL
- Nalinrat Guba, NREL
- Nina Prakash, NREL
- Ashlee Vise, NREL

Data Hub success is a condition of robust inter-lab collaboration with developers and research teams to bring new use cases to the broader HyMARC user base.



What is the HyMARC Data Hub?

- Web application + API (Application Programming Interface)
- CKAN Infrastructure shared among all (5) EMN Data Hubs.
 - Open source software that makes it easy to find, publish and share data.
- 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 to 1 Dataset.
- Metadata for Curation
 - Allows you to tag your data and for searching across all data you have access to.
- User Resources - <https://datahub.hymarc.org/project/user-resources>
 - Data Tool demos, API Walkthrough File Type Views, Data Hub User Guide



The Data Hub is a “Virtual Lab” for the EMN Consortium

- Secure Data Sharing by Project for Team Members
- Create Datasets and Upload files
- Search across all data using defined metadata

- Link to other data repositories or databases
- Data Tools for visualization and analyzing data
- Application Programming Interface (API)
- Centralized Authentication
- Make selected datasets public

Log In Register

HyMARC Hydrogen Materials Advanced Research Consortium

Home Projects Data About Help

HyMARC Data Hub

The submission point for data collected from research conducted by the Hydrogen Materials Advanced Research Consortium

Register
Request a HyMARC account.

Discover
Search the repository.

Submit Data
Upload and archive your data.
Share data with others.





Why use the Data Hub?

- All data uploaded is private by default.
- Can make datasets public. Data Release Procedure.
- Can upload data to collaborate with your dispersed project team.
- Curate your data.
- Create meaningful datasets (collection of files), which can be made public.
- Public HyMARC data is beneficial for the community and continuation of this research.
- Search on metadata terms.

The Data Hub was established under the Consortium to host non-proprietary results and data resulting from awarded projects, in a publicly available data portal for the purpose of advancing solid-state hydrogen storage materials.



Data Management Structure

Projects

Search projects...

Order by: Name Ascending

- INREL** Advanced Characterization
3 Datasets
Research and Development of Advanced Characterization Core Capabilities...
- INREL** ALD Synthesis
3 Datasets
ALD (Atomic Layer Deposition) Synthesis of Novel Nanostructured Metal...
- Lixx** Electrolyte Reactions
0 Datasets
Electrolyte Assisted Hydrogen Storage Reactions Recipient Liox Power, Inc...
- INREL** Fluorinated Frameworks
0 Datasets
Fluorinated Covalent Organic Frameworks: A Novel Pathway to Enhance Hydrogen...
- Argonne** Graphene-Wrapped Hydrides
0 Datasets
Graphene-Wrapped Complex Hydrides as High-Capacity, Regenerable Hydrogen...
- Hydrides**
2 Datasets
Research and Development of Advanced Hydrogen Storage Materials: Metal...
- M** Hydrogen Adsorbents
1 Dataset
Optimized Hydrogen Adsorbents via Machine Learning and Crystal Engineering...
- Hydrogen Carriers**
0 Datasets
Research and Development of Advanced Hydrogen Carrier Materials Recipient...
- Hydrogen Sponge**
0 Datasets
Developing a Novel Hydrogen Sponge with Ideal Binding Energy and High Surf...

Project



Private Dataset



Files and Resources



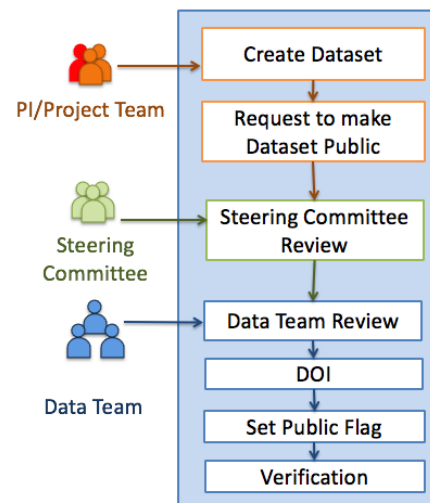
CSV
PDF
TEXT
XML
JSON
JPG, GIF, PNG
Excel

Links to external sites or data

Public Dataset



Data Release Process



Public Access:

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

Team Member Access: Only authenticated and approved Team Members can see Private Datasets and Resources.

All data uploaded is Private by default.

Upon request, data can be made public.



Sharing Data

1. Choose the Project

My Projects

Search projects...

Order by: **Name Ascending**

- NREL** **Advanced Characterization**
4 Datasets
Research and Development of Advanced Characterization Core Capabilities...
- Admin**
4 Datasets
This project is dedicated to administrator production testing.
- MARK** **User Resources**
9 Datasets
This project contains resources for Data Hub users - a data hub user guide...

Project Tree disabled in My Projects view and with search context.

2. Add a Dataset or choose existing Dataset

Project

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

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

4. Upload a file or create a link to external database or other resource

3. Enter metadata

Dataset Metadata

Institution: National Renewable Energy Laboratory

* Author: cpalling

Focus Area: 1-Optimizing of Sorbent Packing

* Maintainer Email: courtney.palling@nrel.gov

Capability Node: Diffuse Reflectance with Cryostat and Gas Adsorption Analyzer

DOI: 10.17042

Sample Barcode:

Sample Name:

Collection Date: 03/03/2020

Data Source Type: Lab Experimental

Comments:

Lab Experimental

Measurement Types: LC NMR DSC

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

Data: Upload Link

Name: eg. January 2011 Gold Prices

Description: Some useful notes about the data

You can use Markdown formatting here



Bulk Upload Feature

The bulk upload feature allows users to upload many files at once through the website (User Interface or UI). You can **access this feature in 2 ways** – when creating a new dataset and when adding additional files to an existing dataset.

I. Creating New Dataset and Bulk Upload. Steps:

1. Click on the project name from the Project menu.
2. Click “Add Dataset”.
3. Complete the Dataset form, select Dataset Metadata terms and add additional “Tags”.
4. Click “Next: Add Data” at the bottom right of form.
5. At the bottom right of the Add New Resource form, click “Switch to bulk add” then “+ Add files” under the Bulk Upload section (you can also Drag and Drop files to the Bulk Upload section on this form).
6. Click “Finish”.


The screenshot illustrates the user interface for creating a new dataset and uploading files in bulk. The interface is divided into several sections:


- Project Menu (Left):** A sidebar menu with options like Overview, Datasets, Activity Stream, Administration, Edit Project, Bulk Edit Datasets, Add Dataset (circled in blue and labeled '2'), Members, and Project Tree.
- Project Overview (Top Right):** Shows the project name 'EMN Data Project', Project ID, and a description field.
- Create Dataset Form (Middle Right):** A form with fields for Title (Bulk II), URL, Project (EMN Data Project), and Description. A green arrow labeled '1' points to the form, and a '2' in a circle is also present.
- Next: Add Data (Bottom Right):** A button labeled '4' that leads to the bulk upload section.
- Switch to bulk add (Bottom Right):** A button labeled '5' that switches the interface to the bulk upload mode.
- Bulk Upload Section (Bottom):** A section with a 'Drag & Drop files here to stage for bulk upload' area, a '+ Add files...' button, and a 'Finish' button labeled '6'.

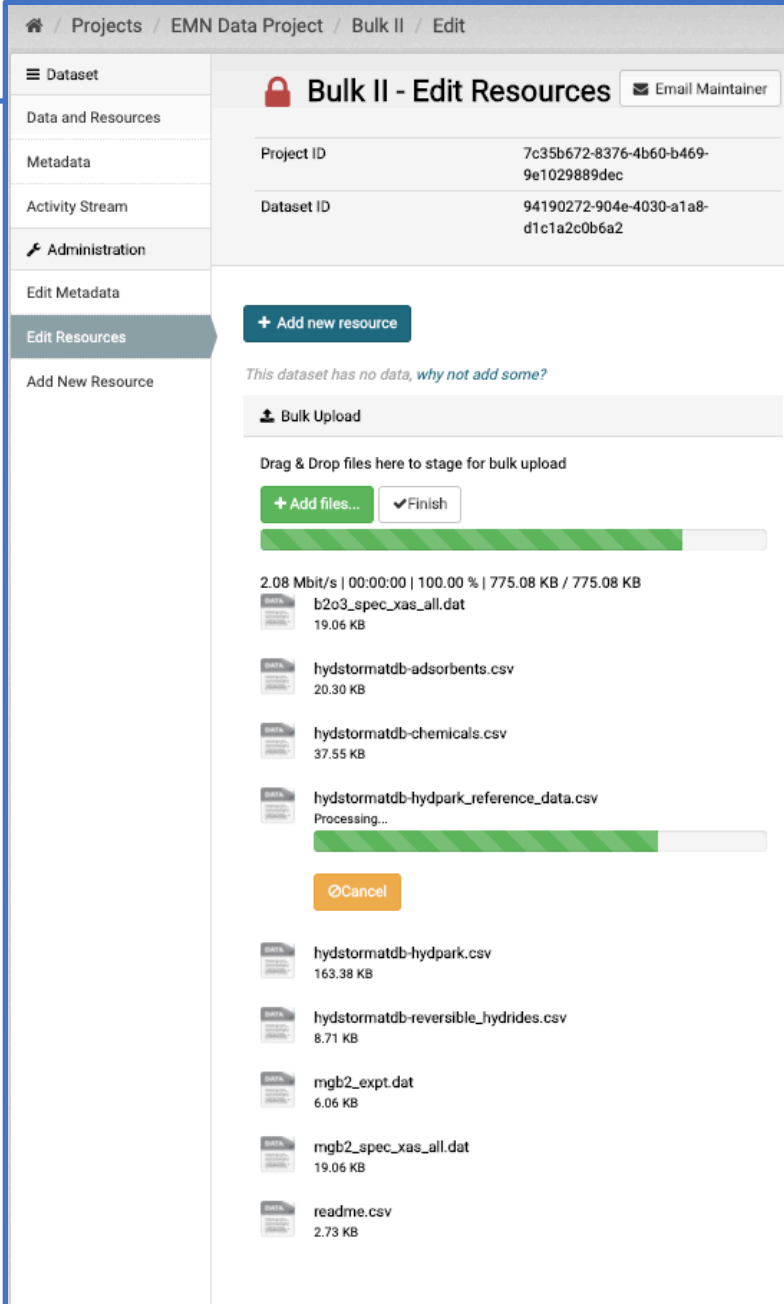


Bulk Upload Feature: Adding data via Bulk Upload to existing dataset

- II. Select Existing Dataset and Bulk Upload Additional Files. Steps:
 1. Select the dataset to which you want to add additional files (Resources).
 2. Click “Edit Resources”, find the Bulk Upload section on the page to “+ Add Files” or Drag and Drop files in the Bulk Upload section.
 3. Click “Finish”

 *The API is recommended to upload large amounts of data (>2GB) as well as it is the most efficient way to access data and metadata for data analysis.*

 *Please contact us to assist with your data hub use case or questions – emnadmin@nrel.gov*



The screenshot shows the 'Bulk II - Edit Resources' interface for the 'EMN Data Project / Bulk II / Edit' dataset. The interface includes a sidebar with navigation options like 'Data and Resources', 'Metadata', 'Activity Stream', 'Administration', 'Edit Metadata', 'Edit Resources', and 'Add New Resource'. The main content area displays the dataset ID (7c35b672-8376-4b60-b469-9e1029889dec) and Dataset ID (94190272-904e-4030-a1a8-d1c1a2c0b6a2). A '+ Add new resource' button is visible. Below it, a message states 'This dataset has no data, why not add some?'. The 'Bulk Upload' section shows a progress bar and a list of files being uploaded, including 'b2o3_spec_xas_all.dat' (19.06 KB), 'hydstormatdb-adsorbents.csv' (20.30 KB), 'hydstormatdb-chemicals.csv' (37.55 KB), 'hydstormatdb-hydpark_reference_data.csv' (Processing...), 'hydstormatdb-hydpark.csv' (163.38 KB), 'hydstormatdb-reversible_hydrides.csv' (8.71 KB), 'mgb2_expt.dat' (6.06 KB), 'mgb2_spec_xas_all.dat' (19.06 KB), and 'readme.csv' (2.73 KB). A '+ Add files...' button and a 'Finish' button are also present.



Sample Tracker (beta release)

- 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).



Metadata – collection of files (Dataset)

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

- **Institution:** choose the HyMARC Institution associated with this Dataset
- **Author:** populated with your username
- **Focus Area:**
 - ✓ -- select an option --
 - 1-Dynamic Sorbent Materials
 - 1-Enthalpy-Entropy
 - 1-Multi H₂s Per Metal
 - 1-Nanoscale Defects in Sorbents
 - 1-Optimizing of Sorbent Packing
 - 1-Optimizing Sorbent Binding Energies
 - 2-Activation of B-B and B-H Bonds
 - 2-Machine Learning for New MH
 - 2-Metal Hydride Thermodynamics
- **Maintainer Email:** responsible for dataset maintenance (author by default)
- **Capability Node:**
 - ✓ -- select an option --
 - Ab Initio Calculation Capabilities
 - Advanced PCT Measurements with Cryogenic Cooler
 - Chemistry of Hydrogen Interactions with Materials
 - Clean Transfer Cell for Air-Free Sample Transfer
- **DOI** – we can request your DOI
- **Sample Barcode** (if applicable)
- **Sample Name**
- **Collection Date:** date this data was collected
- **Data Source Type:**
 - External Data or Tools
 - Historical or Literature
 - Lab Experimental
 - Modeling and Simulation
 - Survey Data
- **Comments**

Dataset Metadata

* Institution:

-- select an option --

* Author:

admin

* Focus Area:

-- select an option --

* Maintainer Email:

emnadmin@nrel.gov

* Capability Node:

-- select an option --

DOI:

10.17042

Sample Barcode:

Sample Name:

Collection Date:

Data Source Type:

-- select an option --

Comments:

Comments



Metadata – data files (Resource)

When uploading data files, you will be prompted to choose metadata to describe the file, such as measurement type and instrument.

* NOTE: These lists will expand as needed.

- **Measurement Type:**

- BET
- DRIFTS
- DSC
- GC
- IPC
- LC
- NMR
- **Other**
- PCT
- Raman
- SEM
- Small Angle Xray Scattering
- Soft Xray Microscopy
- TEM
- TGA
- TGA-DTA
- Thermal Conductivity
- TPD
- UVVis
- XAS-EXAFS
- XPS
- XRD
- XRF

- **Instrument:**

- NREL PCT 1
- NREL PCT 2
- NREL PCT 3
- LBNL NMR-FTIR-PCT
- PNNL High-Low Temp NMR
- LBNL Soft Xray Microscope
- Other



Adding Resources to a Dataset

A dataset can have as many files or links to external resources as desired (1 dataset: many files).

- Dataset metadata pertains to many files.
- You may use the [Data Hub 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 within the project. Follow the prompts (see slide 6).
 - **To add 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.
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.



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

Project - Dataset

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

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

File test file

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

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

File: readme.csv **Remove**

Name: readme

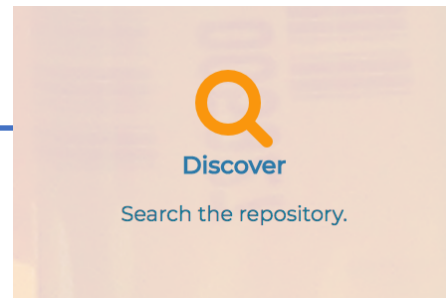
Description: Some useful notes about the data

Format: CSV

Delete **Update Resource**



Searching Data



Login to view all data you have access to – public and private data.

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 has a faceted search and count of matching datasets:

- Drill Down on your projects, tags, institution, capability node, and focus area

The screenshot shows the 'Datasets' page of a data repository. On the left is a navigation menu with sections: Projects (User Resources (5), Advanced Characteri... (2), Hydrides (2)), Tags (demo (3), data tool (2), TEM (2), Al2O3 coating (1), alumina (1), AP-XPS (1), API (1), atomic layer deposi... (1), data tools (1), magnesium borohydride (1)), Institution (National Renewable ... (5), Lawrence Berkeley N... (1), Sandia National Lab... (1)), Capability Node (There are no Capability Node that match this search), and Focus Area (3-Advanc... situ ... (1)). The main content area has an 'Add Dataset' button, a search bar with the text 'Search datasets...', and an 'Order by: Relevance' dropdown. Below the search bar are search results for 'Simulated AI K-edge XANES of NaAlH4' (2 Resources, PRIVATE), 'Help and Tutorial' (2 Resources, PUBLIC), 'TEM Mg(BH4)2 10 cycles Al2O3 ALD' (25 Resources, PRIVATE), 'Synchrotron Proposals' (2 Resources, PRIVATE), and 'TEM MBH 100cycles Al2O3' (19 Resources, PRIVATE). A blue arrow points from the search bar to the text 'Login to view all data you have access to – public and private data.' Another blue arrow points from the 'Focus Area' section of the navigation menu to the text 'Left Navigation has a faceted search and count of matching datasets:'.



Data Tools:

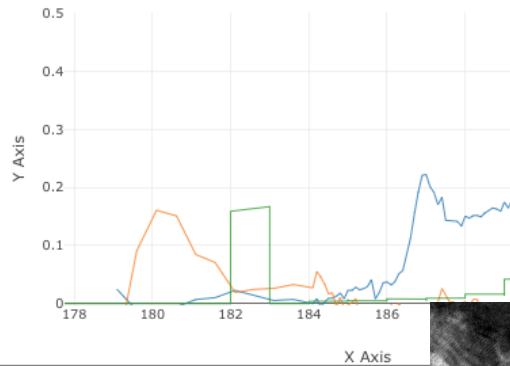
<https://datahub.hymarc.org/project/user-resources>

See User Resources for information on the resources available on the data hub :

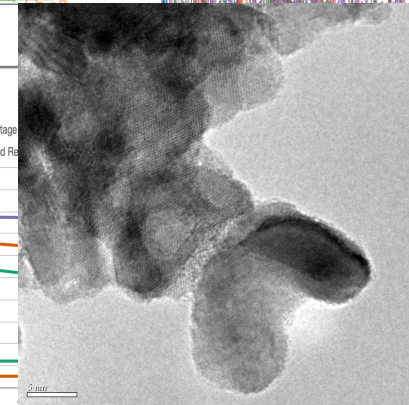
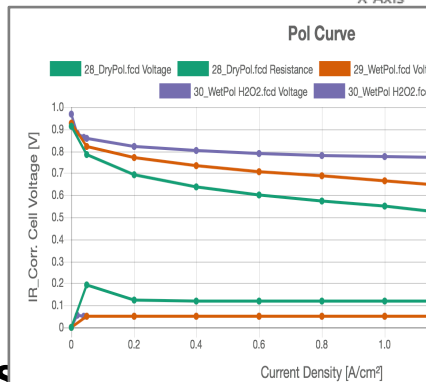
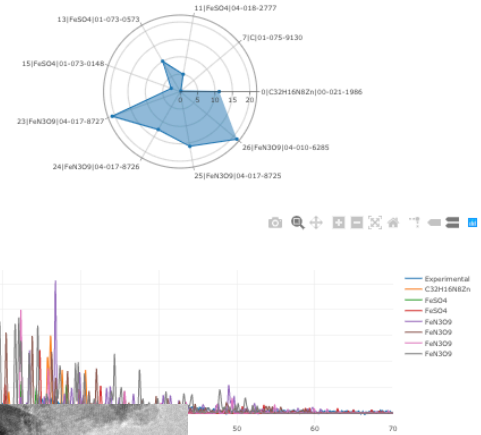
- Multi-spectra data view
- Electrolysis Pol Curve data tool
- Metadata and API helpers
- Data upload tool for image formats and multi-file uploads.
- XRD Unmix data tool
- Ckan data tools and views for csv, pdf, text, JSON and other file formats.

Data tools can be applied to files in all projects. View the demonstrations and descriptions in User Resources.

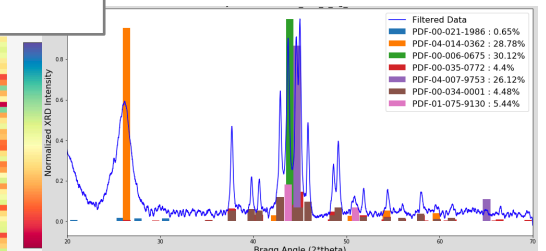
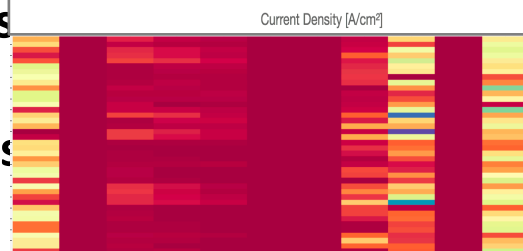
Advanced Multi-spectra Data Tool



Unmix XRD Results



- Acq_date: 3/8/2017
- acq_time: 4:54:31 PM
- Name : FEI Tecnai
- Micro:
- Hv: 300000.0
- Mag: 2950000
- Mode: Imaging
- Operator:
- Specimen "InP_4torr
- Scale: 0.01998 nm/px
- Size (2048 X 2048)





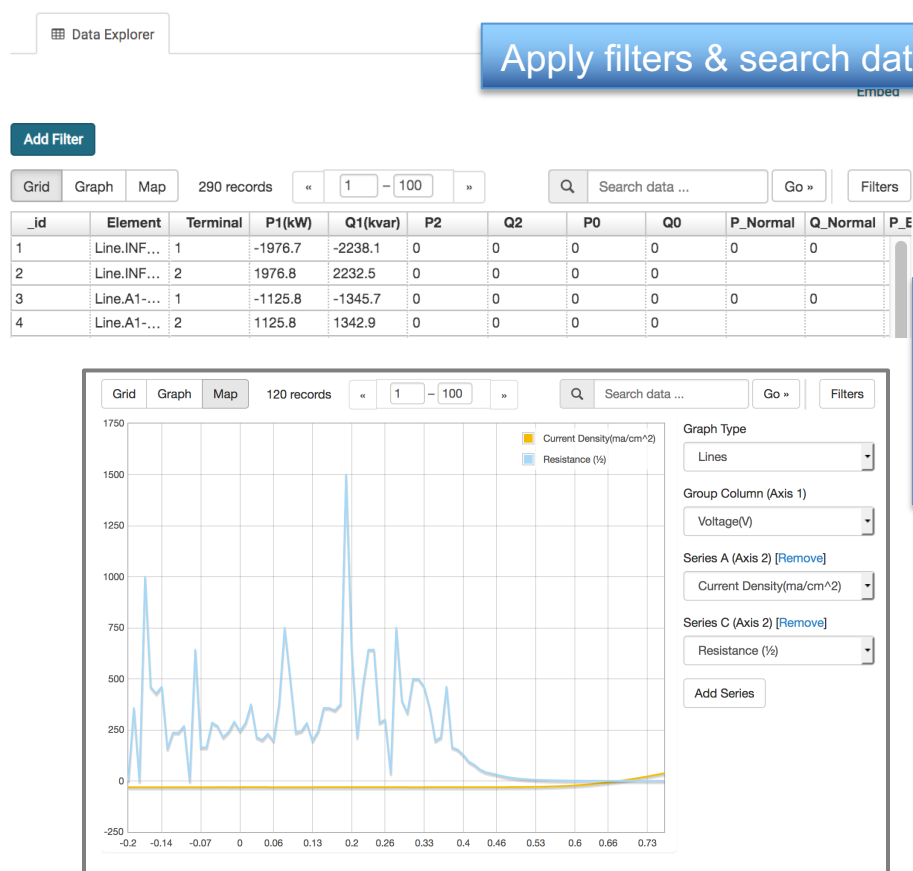
Data Tools continued

Generic Data Tools: The Data Hub comes with a tool for visualizing structured data, such as CSV files:

- View the raw data in grid format
- Filter or Search the data based on column value ranges
- Dynamically plot any of the columns as series
- Example: HyMARC public dataset – Computational Prediction of Hydrogen Storage Capacities in MOFs.

Other CKAN Data Tools:

- Image, pdf, txt file views
- Google Analytics plug-ins



Apply filters & search data within the file

Dynamically create a graph from the CSV file

Shared and Custom Data Tools: The Data Hub supports custom “plug-ins” for HyMARC-specific visualizations

Compare data from multiple files on the same plot



Acknowledgements



Enabling twice the energy density for onboard H₂ storage

We are grateful for the financial support of EERE/HFTO and for technical and programmatic guidance from Ned Stetson, Jesse Adams, and Zeric Hulvey

NREL Data Hub Development Team

- Caleb Phillips
- Nick Wunder
- Nalinrat Guba
- Nina Prakash
- Chris Webber
- Courtney Pailing
- Ashlee Vise

HyMARC Data Team

- Mark Allendorf (SNL)
- Pragya Verma (LBNL)
- Rick Karnesky (SNL)
- Chitra Sivaraman (PNNL)
- Yong Han (LLNL)
- Brandon Wood (LLNL)
- Tom Gennett (NREL)

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Fuel Cell Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.