



A RESEARCHER'S GUIDE TO THE DURAMAT DATA HUB:

HOW TO MAKE IT WORK FOR YOUR PROJECT AND RESEARCH

Robert R. White

Aug 10, 2020

What is a Data Hub?

Simply,

A software platform that provides a virtual laboratory with a robust method of secure data management, data exploration, and dissemination, allowing the remote researchers to work efficiently and effectively.

Organizationally,

The Data Hub is part of Capability one: Central Data Resource.

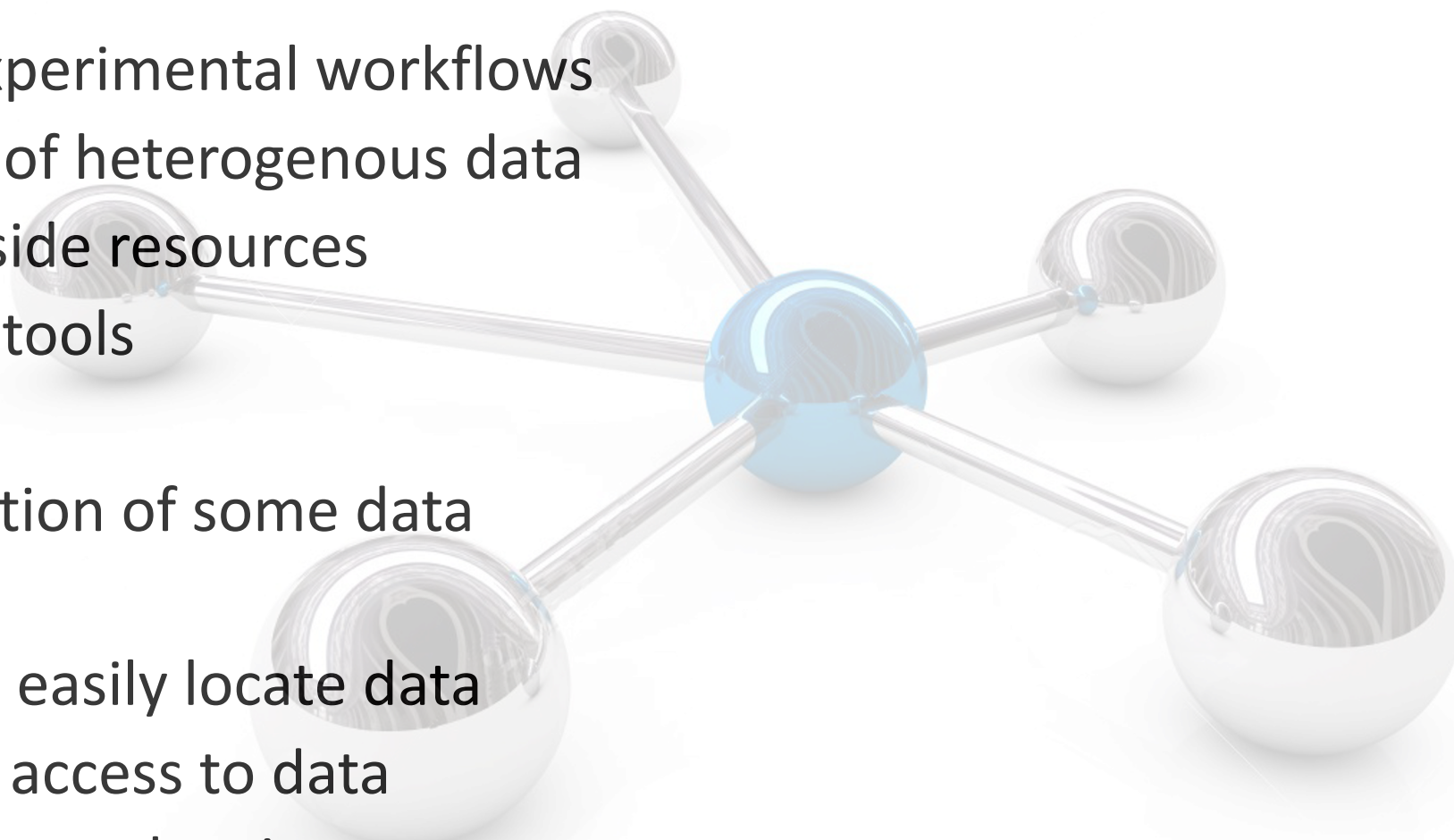
Technically,

The Data Hub is series of customizable software elements integrated to provide a data repository and website, deployed in the cloud.

Follow along or more help at: <https://datahub.duramat.org/dataset/help-and-tutorial>

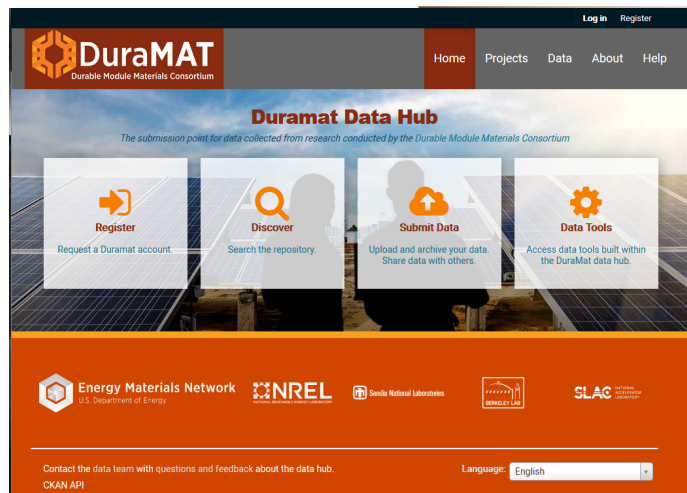
What Can the Data Hub Do?

- Support a variety of experimental workflows
- Archive a wide variety of heterogeneous data
- Support linking to outside resources
- Archive or link to data tools
- Capture metadata
- Provide quick visualization of some data
- Secure web portal
- Queryable interface to easily locate data
- Provide programmatic access to data
- Allow customization as needs arise



Where is the Data Hub?

<https://datahub.duramat.org>



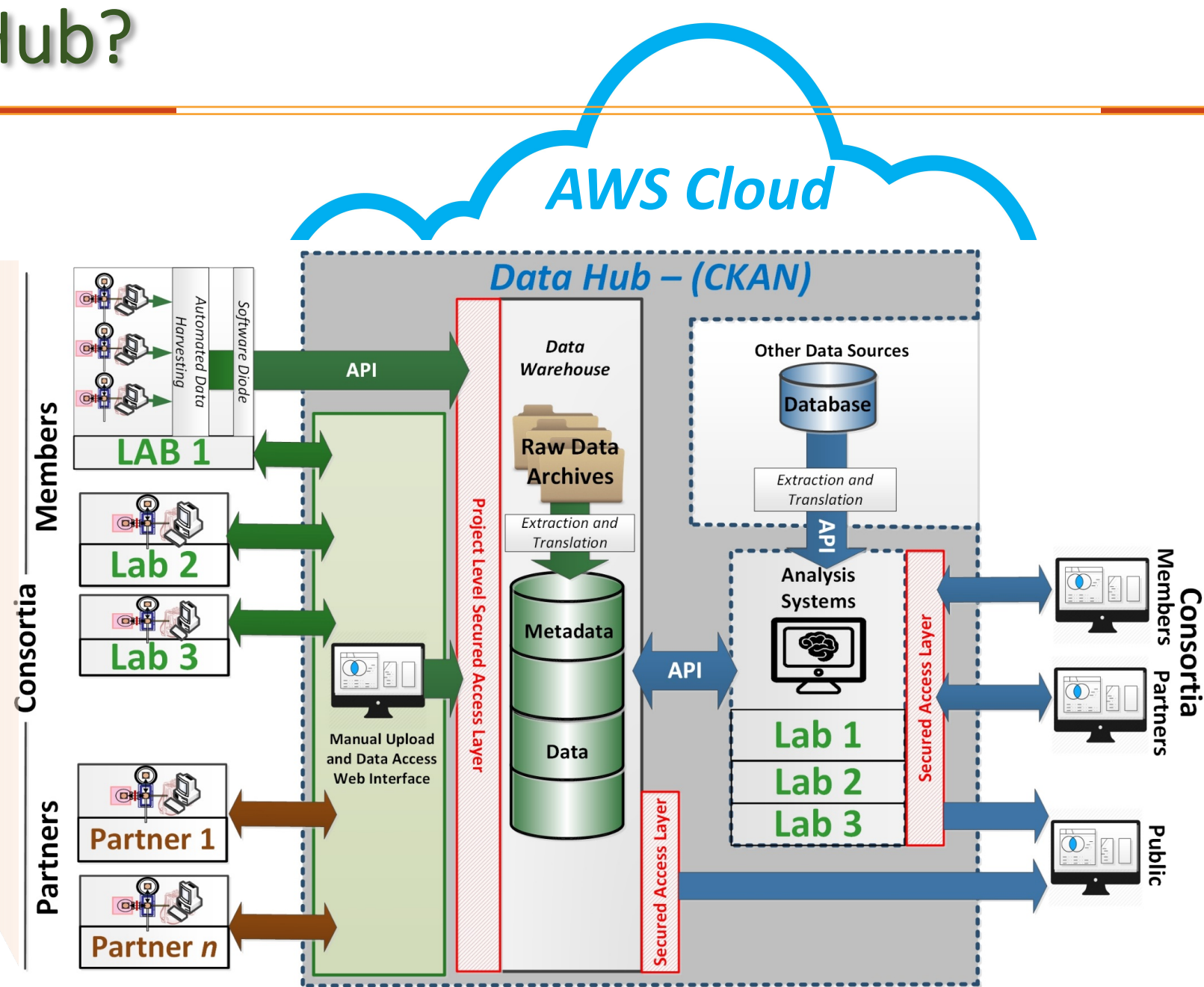
Ancillary sites

<https://pvtools.lbl.gov>

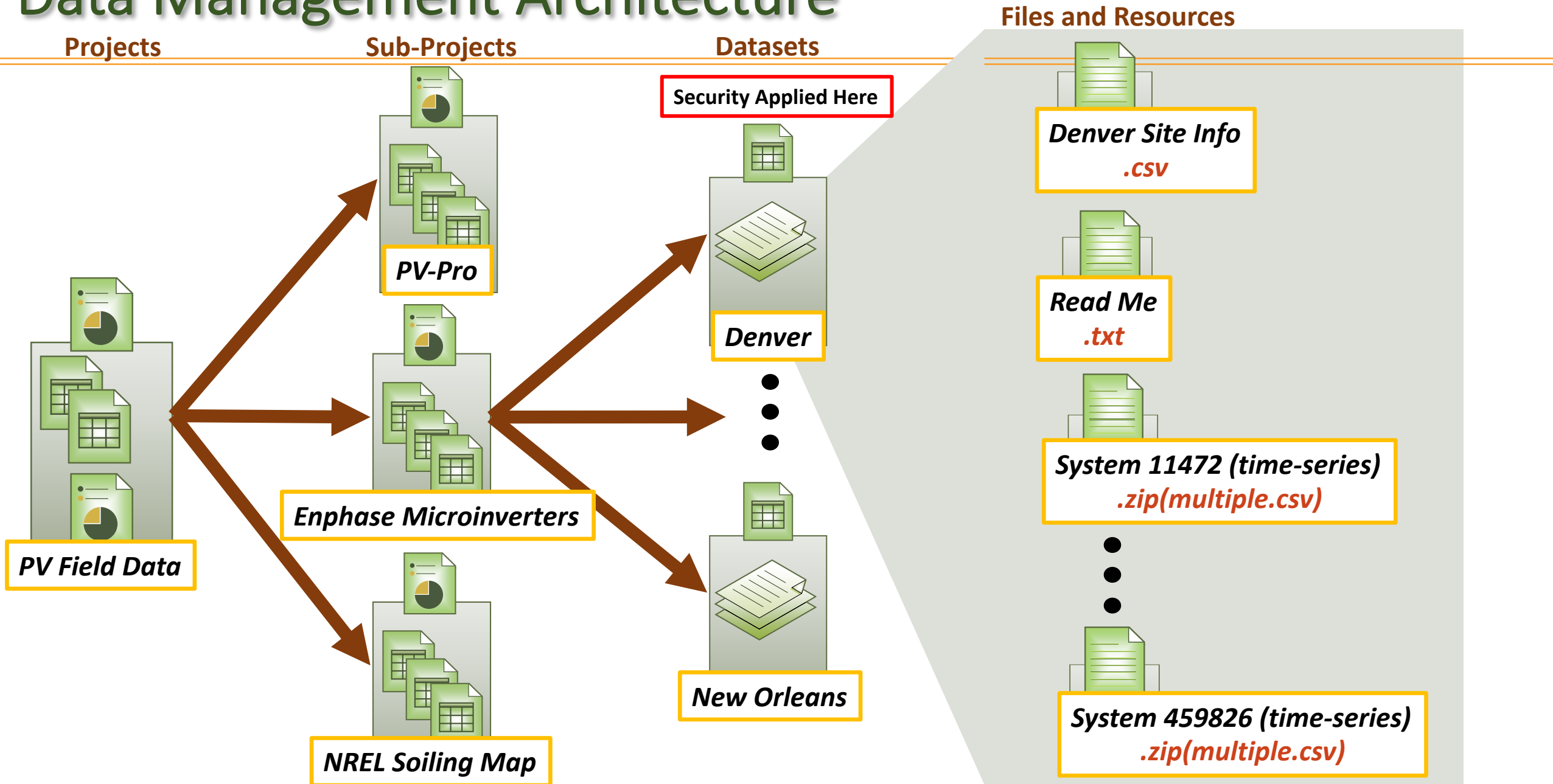
<https://pvdata.duramat.org>

<https://github.com/DuraMAT>

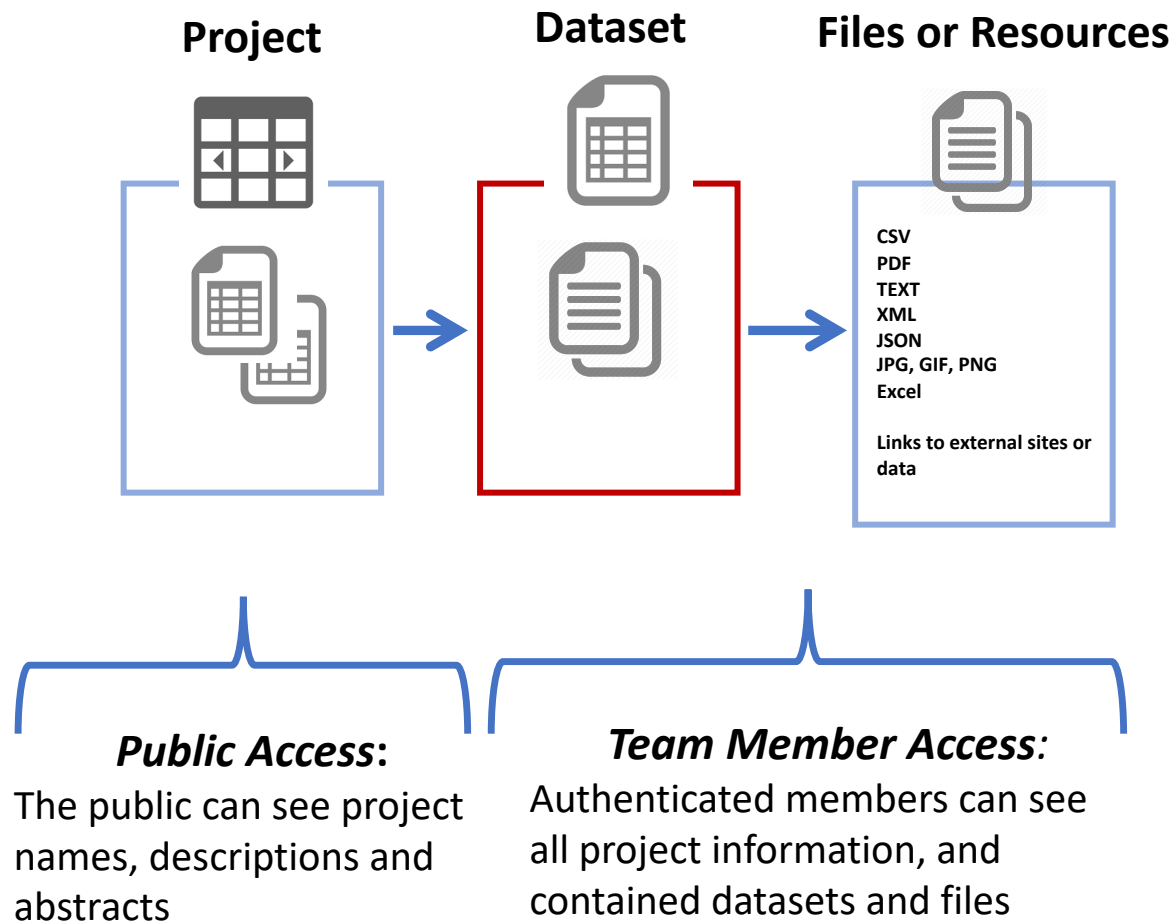
<https://github.com/EMN-Data>



Data Management Architecture



Data Security



- Researcher logins via Cognito
 - Login access works for all datahubs
- Data security applied at dataset level
- **All** data uploads as “private”
- Three Project Member permissions
 - Member – Can see all data in project
 - Editor – Can add new datasets and data
 - Admin – Can add new members to project
- Data uploads only from:
 - DuraMAT researchers
 - Vetted data sources (no general public data)
- Public data release under administrative control, requires authorization.

What's in the Data Hub

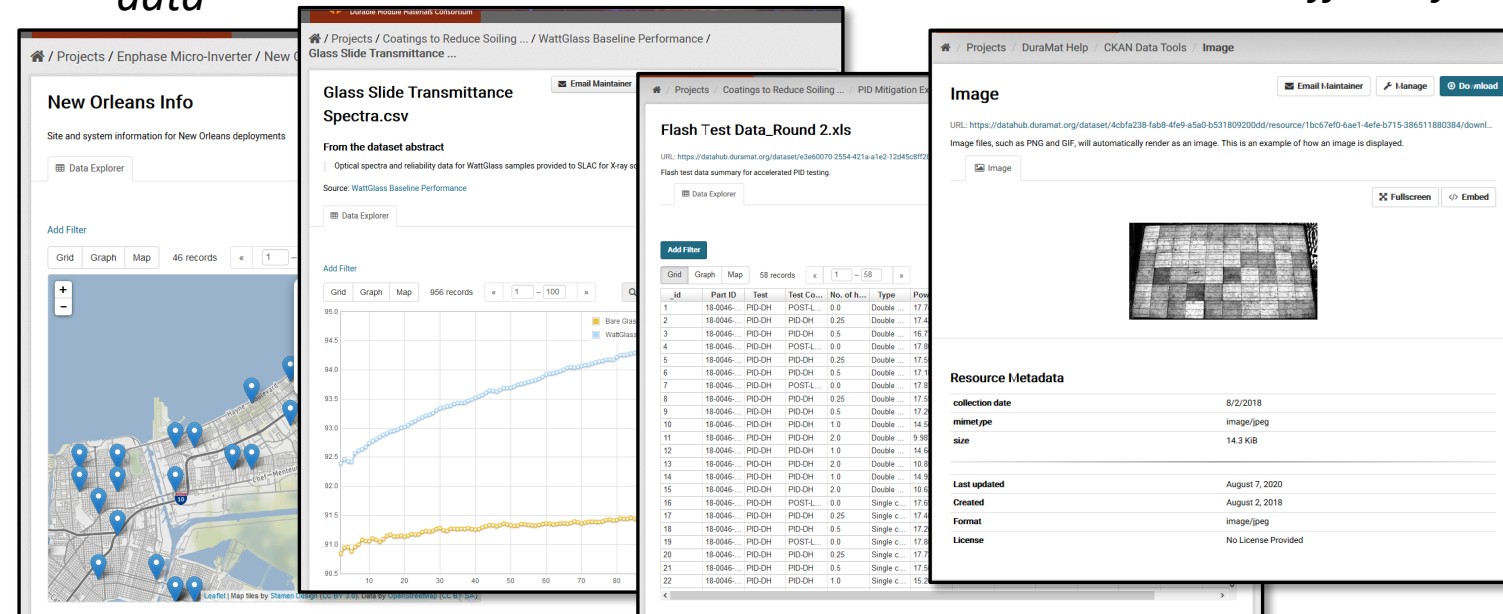
For the public:

- All project descriptions and abstracts
- Data tools
- Publicly available datasets
 1. *Albedo Data for Bifacial Systems*
 2. *Coatings to Reduce Soiling and PID Losses*
 3. *NREL Soiling map and supporting data*
 4. *Bifacial Experimental Single-Axis Tracking Field data*

For project members:



- Project datasets
- Data tools
- All publicly available datasets
 1. *Back side defect imaging in crystalline silicon photovoltaic modules*
 2. *Identifying Degradation Mechanisms in Fielded Modules using Luminescence and Thermal Imaging*
 3. *Combined Accelerated Stress Testing*
 4. *Effect of Cell Cracks on Module Power Loss and Degradation*

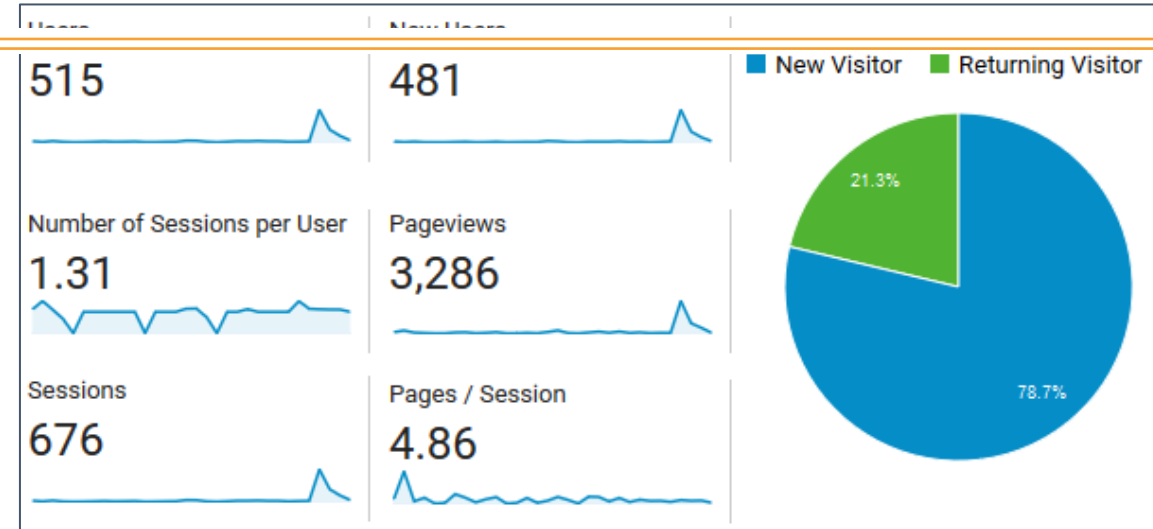
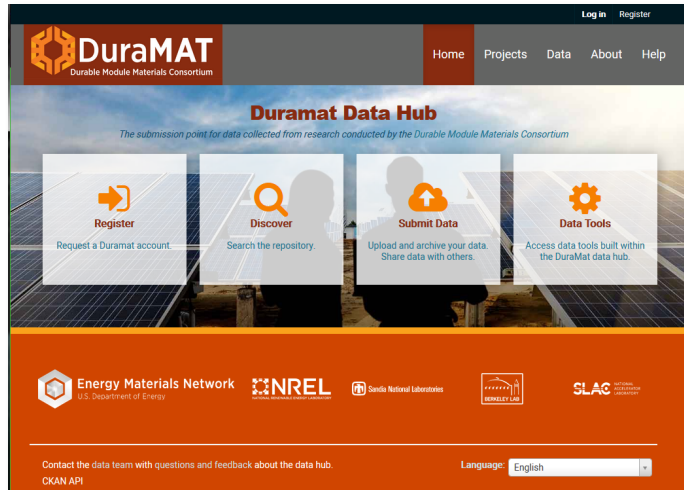


Data Tool Links:

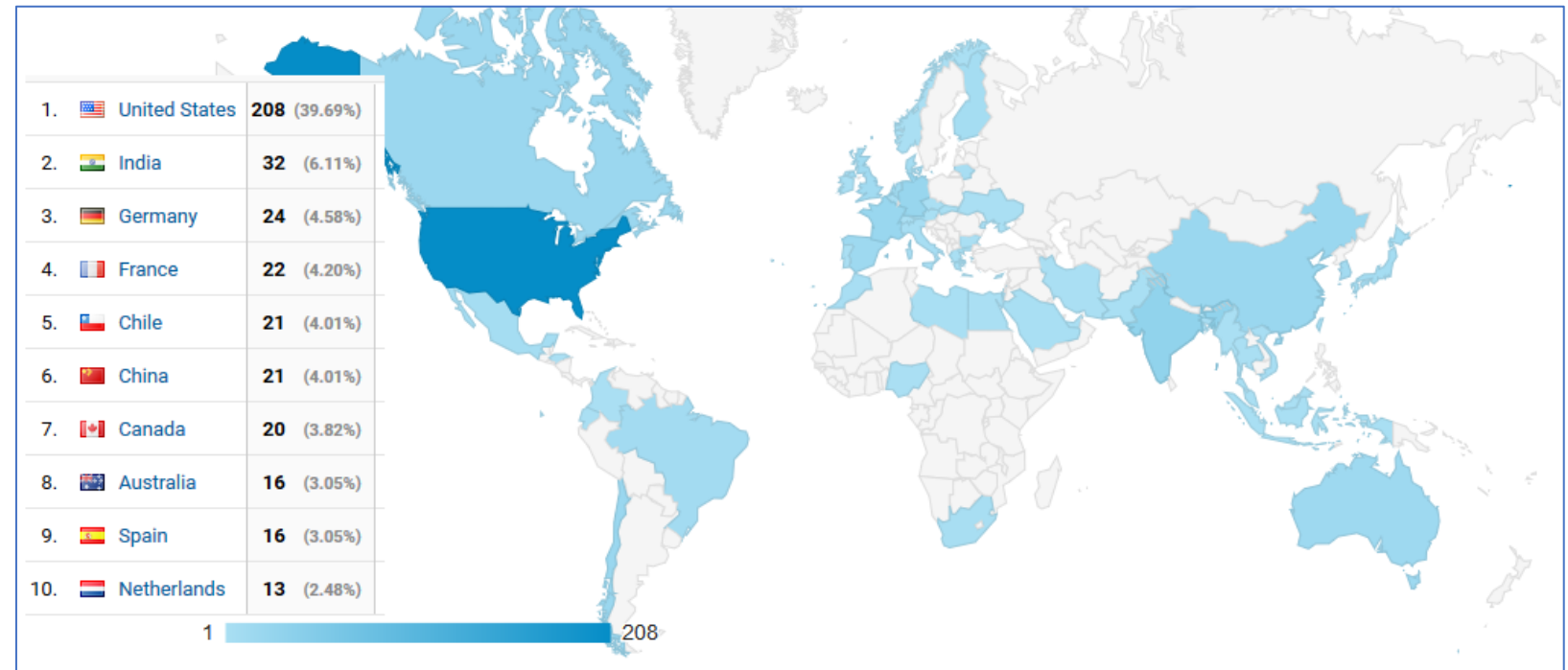
1. Rd Tools
 2. Photovoltaic String Calculator
 3. PV Climate Zone Stressors
 4. IV Curve Corrections
 5. Clear Sky Detection
-and more

Is Anybody Using It?

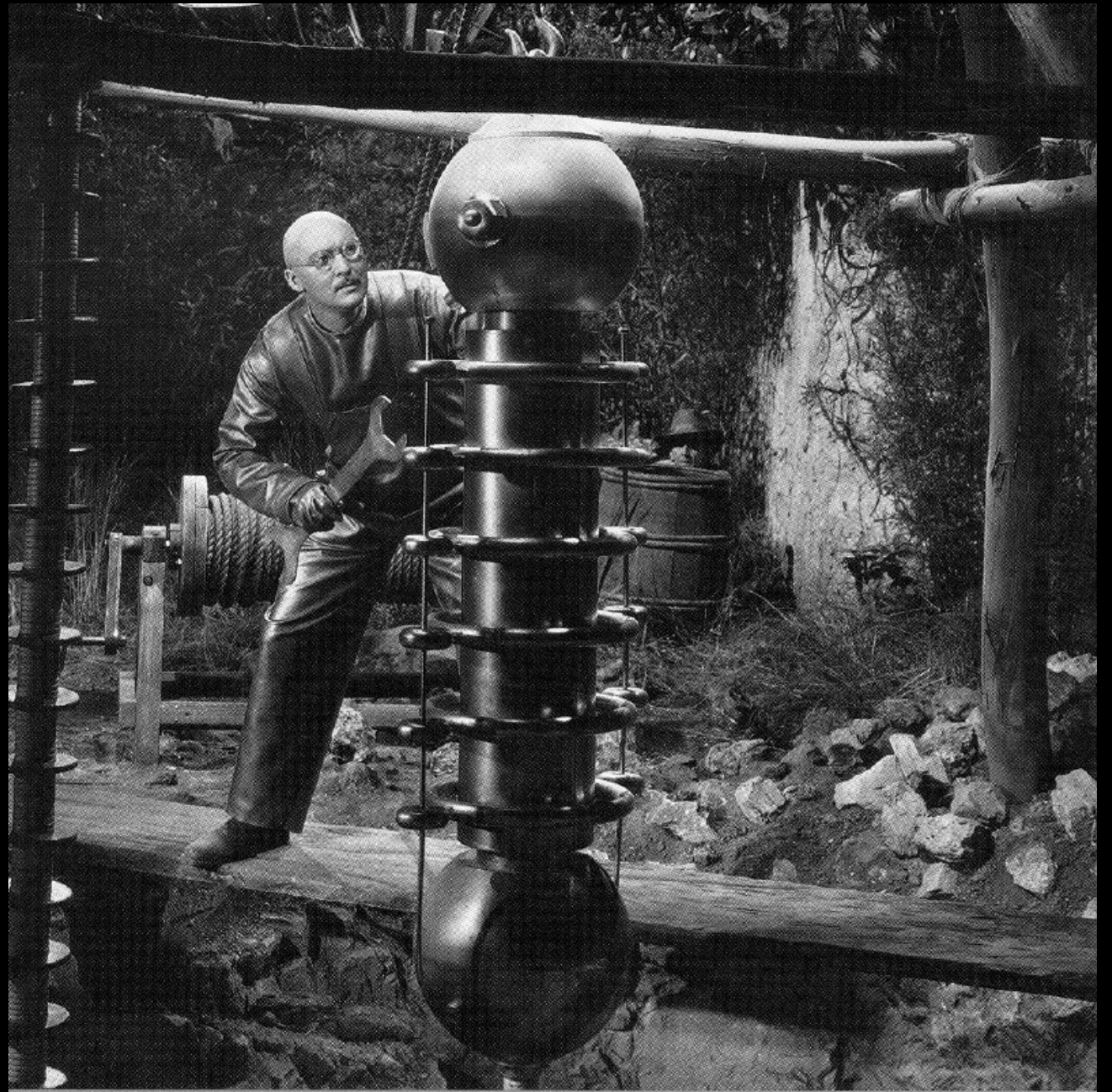
Google analytics July 1 –Aug 10, 2020:



Data Hub Stats
137 DuraMAT Users
63 Projects
128 Datasets
70 Public Datasets
2120 Files and Resources



Making it work

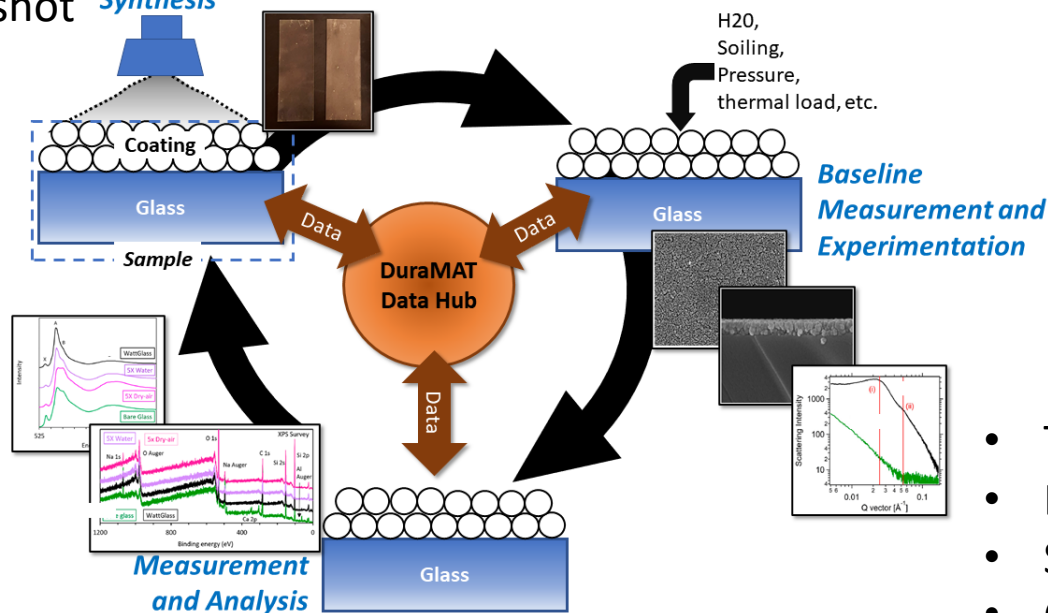


How do I set up my project?

A. Design the Project Archive

Workflow

1. Round robin experiments
2. Running data archive
3. One-shot *Synthesis*



Data

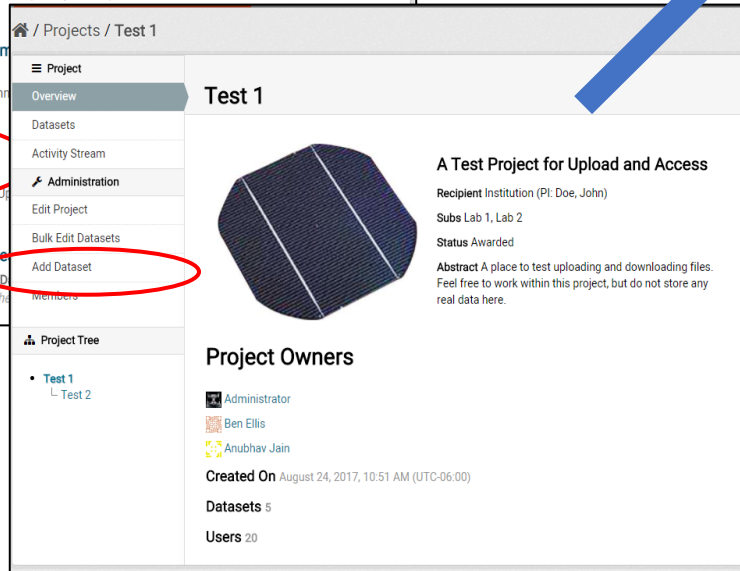
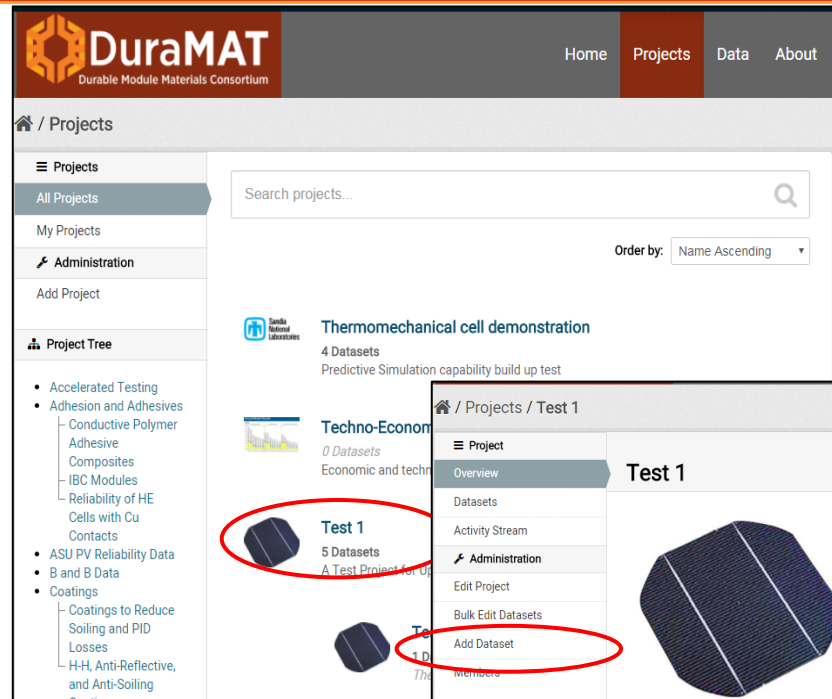
1. Instrumentation
2. **Metadata**
3. Validation / QA
4. Moving data
5. Other team access

B. Project Description Form

<https://datahub.duramat.org/dataset/new-project-request>

- Category of project
 - Accelerated testing
 - Adhesions and adhesives
 - Coatings
 - Encapsulants and barriers
 - Mechanics and cracking
 - Module forensics
 - Data analysis and quality assurance
 - Predictive simulation
 - Modeling
 - Field Data Studies
- Title
- Lead Organization and PI name
- Supporting organizations
- Current Status (Awarded, Complete, in progress)
- Abstract (Detailed description of the project and its goals)
- Access List (What DuraMAT members need to get to the data)
 - What kind of data access does each member need:
 - Member, Editor, or Admin

How do I put my project data in the Data Hub?

A screenshot of the 'Create Dataset' form in the DuraMAT Data Hub. The form is divided into two steps: '1 Create dataset' and '2 Add data'. The '1 Create dataset' step includes fields for 'Title', 'Project', 'Description', 'Tags', 'License', 'Dataset Metadata' (Author, Institution, Data Source Type), and 'Lab and Experimental' (Accelerated, Environmental Conditions, Measurement Types). A blue arrow points from the 'Add Dataset' button in the previous screenshot to the 'Add Data' button at the bottom right of this form.

1. Select the project
2. Click “Add dataset”
3. Fill out interactive form
4. Click button “Add Data”
5. Fill out interactive form

A screenshot of the 'Add New Resource' form in the DuraMAT Data Hub. The form is divided into two steps: '1 Create dataset' and '2 Add data'. The '2 Add data' step includes fields for 'Data' (Upload or Link), 'Name', 'Description', 'Format', and a 'What's a resource?' section. A blue arrow points from the 'Add Data' button in the previous screenshot to the 'Add New Resource' form.

Or you can use the API and perform the same steps programmatically.

A new multiple file (bulk) upload feature is in test and will be available next month.

How do I get my project data?

Using the Website

1. Log into the Data Hub (For private project data only)
2. Click on Projects and select the project from the list
3. On left hand menu, click “Datasets”
4. Click the dataset you want to access from the list
5. Select the data you wish to see (“View”) or download

Projects / NREL Soiling Map / Soiling Data

Dataset

Data and Resources

Metadata

Activity Stream

Administration

Edit Metadata

Edit Resources

Add New Resource

Soiling Data - Data and Resources

Project ID: 15e1b647-f595-4...6-a4d6-a98186...7ac

Dataset ID: 34897add-3765-4...5-9571-3de6f408d06f

Online Map Data_Oct_2017.csv

DA API

Site information

Data Site ID 3.csv

DA API

Field soiling data

Data Site ID 4.csv

View Download Edit

Can I change the data?

Click “Edit” next to the name of the file

Projects / NREL Soiling Map / Soiling Data / Online Map Data_Oct_2017.csv

Online Map

Data_Oct_2017.csv

URL: <https://datahub.duramat.org/dataset/34897add-3765-4205-9571-3de6f408d06f/resource/7a76577a-0f92-43ec-9654-228fa8c44745/download>

Site information

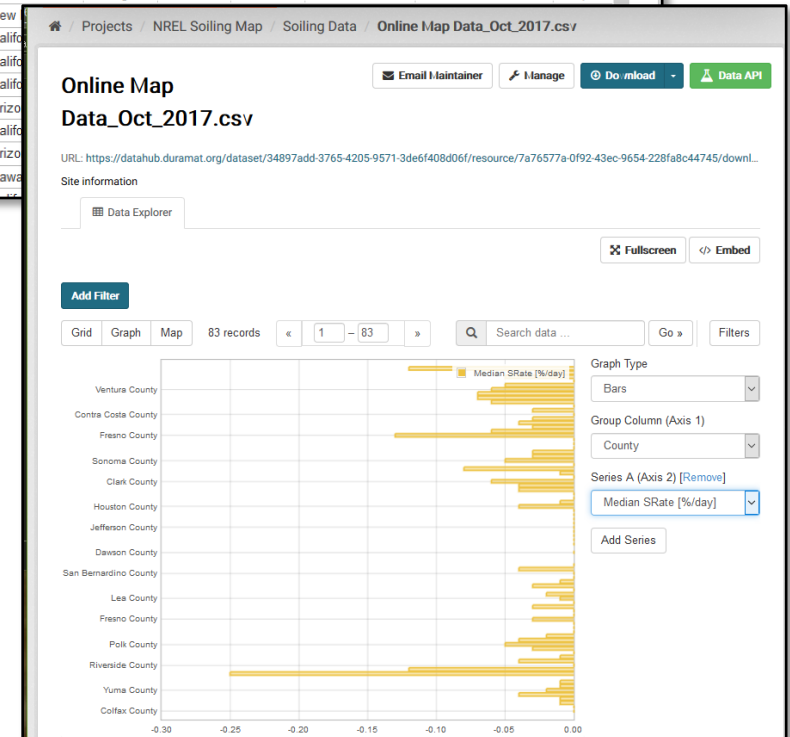
Data Explorer

Fullscreen Embed

Add Filter

Grid Graph Map 83 records

id	County	State	System...	Start Year	End Year	Mean SRatio	SRatio Uncertainty	Median SR
1	Colfax County	New Mexico	Soiling ...	2013	2016	0.997	0.001	0% per d
2	Luna County	New Mexico						
3	Imperial County	California						
4	Fresno County	California						
5	Los Angeles Co...	California						
6	Yuma County	Arizona						
7	San Luis Obisp...	California						
8	Pima County	Arizona						
9	Honolulu County	Hawaii						



Accessing data without the Web App

All items stored within the Data Hub is assigned a Universally Unique Identifier (UUID); projects, datasets, data files... *and even registered users*

user UUID
(a.k.a API Key) –
User Page

The screenshot shows the user profile for 'Mr. Cool Cat' (mccat). The profile includes a cat avatar, a bio 'Test subject in DuraMat', and statistics: 0 followers, 13 datasets, and 321 edits. The 'API Key' (a.k.a. API Key) is circled in red at the bottom left, showing the value: 55053878-2f25-3878-8cf3-96012f25505. The right side of the page lists several datasets, including 'Another Test Dataset', 'ds1', 'Compendium of photovoltaic degradation rates', 'Sample Instrument Synthesis and Log File', 'Sample File Uploads', and 'General Standards Documentation'.

project UUID – Project overview page
Project datasets page

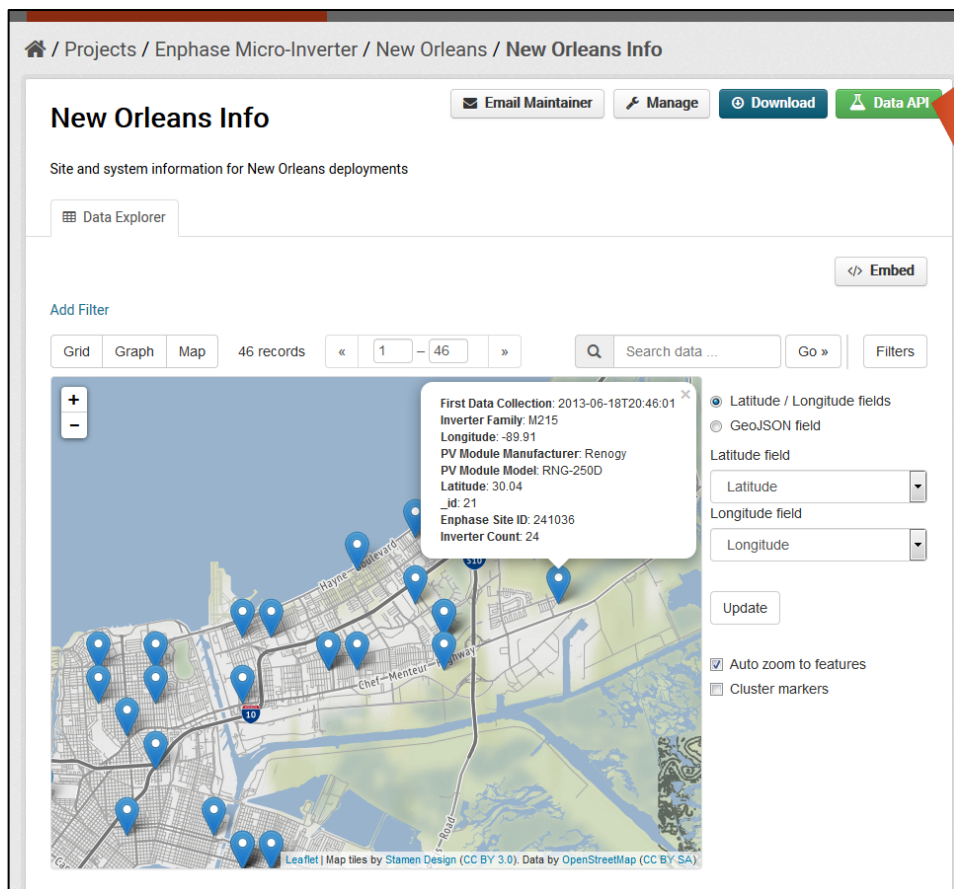
The screenshot shows the project overview page for 'Enphase Micro-Inverter'. The 'Project ID' is circled in red, showing the value: 6448b09a-f253-4e5c-9a97-e9e88b628df5. The page includes a sidebar with navigation links (Overview, Datasets, Activity Stream, Administration, Edit Project, Bulk Edit Datasets, Add Dataset, Members, Project Tree) and a main content area with project details and metadata.

dataset UUID – Dataset page

The screenshot shows the dataset page for 'EnPhase time-series data - New Orleans'. The 'Dataset ID' is circled in red, showing the value: 3d125a42-a098-4270-bb48-2fb046f3ca4a. The page includes a sidebar with navigation links (Dataset, Data and Resources, Metadata) and a main content area with dataset details and resources.

Accessing the data file through API

All files are accessible for download from the API, but CSV files are special, and the API will allow for querying and accessing records in the CSV file.



CKAN Data API

Access resource data via a web API with powerful query support. Further information in the main [CKAN Data API](#) and [DataStore](#) documentation.

Endpoints »

The Data API can be accessed via the following actions of the CKAN action API.

Create	https://datahub.duramat.org/api/3/action/datastore_create
Update / Insert	https://datahub.duramat.org/api/3/action/datastore_upsert
Query	https://datahub.duramat.org/api/3/action/datastore_search
Query (via SQL)	https://datahub.duramat.org/api/3/action/datastore_search_sql

Querying »

Query example (first 5 results)

https://datahub.duramat.org/api/3/action/datastore_search?resource_id=3796002f-017e-4420-95d3-89bfcf50bf91&limit=5

Query example (results containing 'jones')

https://datahub.duramat.org/api/3/action/datastore_search?q=jones&resource_id=3796002f-017e-4420-95d3-89bfcf50bf91

Query example (via SQL statement)

https://datahub.duramat.org/api/3/action/datastore_search_sql?sql=SELECT * from '3796002f-017e-4420-95d3-89bfcf50bf91' WHERE title LIKE 'jones'

Example: Javascript »

Example: Python »

```
import urllib
url = 'https://datahub.duramat.org/api/3/action/datastore_search'
fileobj = urllib.urlopen(url)
print fileobj.read()
```

API Demo

Need more help?

Go to the Data Hub <https://datahub.duramat.org>

Go to Projects > DuraMAT Help > Datasets > Help and Tutorial

- Data Hub User's Manual – Covers all of this and more
- DuraMAT Data Hub Api Walkthrough - Another copy of the API iPython notebook
- Video Tutorial

Acknowledgements:

(the developers, supporters and early adopters)

Kris Munch
Courtney Pailing
Chris Webber
David Ragar
David Whiteside
Teresa Barnes

Nick Wunder
Nalinrat Guba
Michael Bahl
Jainli Gu
Laura Schelhas
Stephanie Moffitt

Anubahv Jain
Todd Karin
Ben Ellis
Birk Jones
Cameron Stark
Margaret Gordon



DuraMAT
Durable Module Materials Consortium



Thank you



Energy Materials Network
U.S. Department of Energy

