U.S. DEPARTMENT OF ENERGY

Virtual Tutorial Series

Open-Source Tools & Open-Access Solar Data

An overview of the datasets acquired from the Solar Data Bounty Prize

Tassos Golnas, *Solar Energy Technologies Office (SETO)*
Noreen Gentry, *Solar Energy Technologies Office (SETO)*
Robert White, *National Renewable Energy Laboratory (NREL)*

Agenda

1. Webinar Series Overview
2. PV Data Usage & Solar Data Bounty Prize Recap
3. Prize Winners & Datasets Intro
4. Datasets Content & PV Prize Site Details
5. Metadata File Content vs. Data File Content
6. Exploring & Downloading the Data
7. Citing the Datasets
8. Questions

Data: A Means to an End

Better photovoltaic (PV) models and system performance through high-quality data.

PV models are important in:

- Project development and valuation
- Power plant operation and maintenance.

Better system performance means lower cost of solar electricity.

Solar Data Bounty Prize Purpose

Support industry and academic research efforts to **develop**, **improve**, **evaluate**, and **validate** models of real-world PV system performance in diverse locations.

Solar Data Bounty Prize Recap

Goal: Incentivize system owners to share information-rich datasets from their assets

Phases and Prize Pools

- Two-stage, two-track program
- Up to $1,415,000 in cash prizes

Stage 1 Submission Materials

- System metadata
- One month or more of irradiance time series data

Stage 2 Submission Materials

- Complete time series data

Results

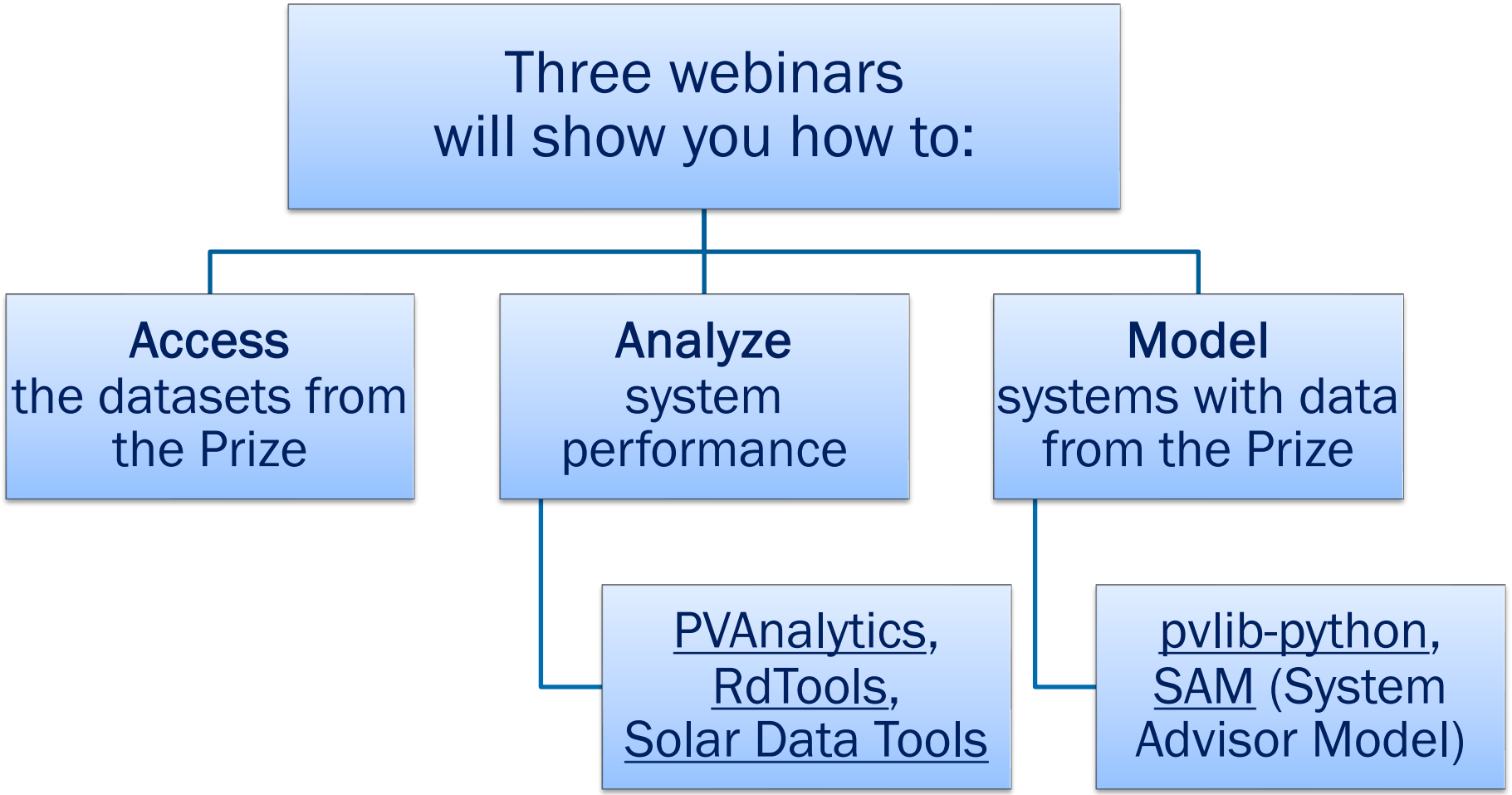
The winners' data sets are shared publicly via a dedicated platform:

PVDAQ/PVData Map | Open Energy Information (openei.org)

Open-Access Data & Open-Source Tools

Scan to Register

for Webinar 2 and 3!



Prize Winners & Datasets Intro

AMERICAN MADE
U.S. DEPARTMENT OF ENERGY

Congratulations to Our Prize Winners!

Track A: Medium-sized Systems

island way technology

SILICON RANCH

WEXUS

Track B: Large-sized Systems

SILICON RANCH

Southern Company

By the Numbers

NEW Publicly Available Data

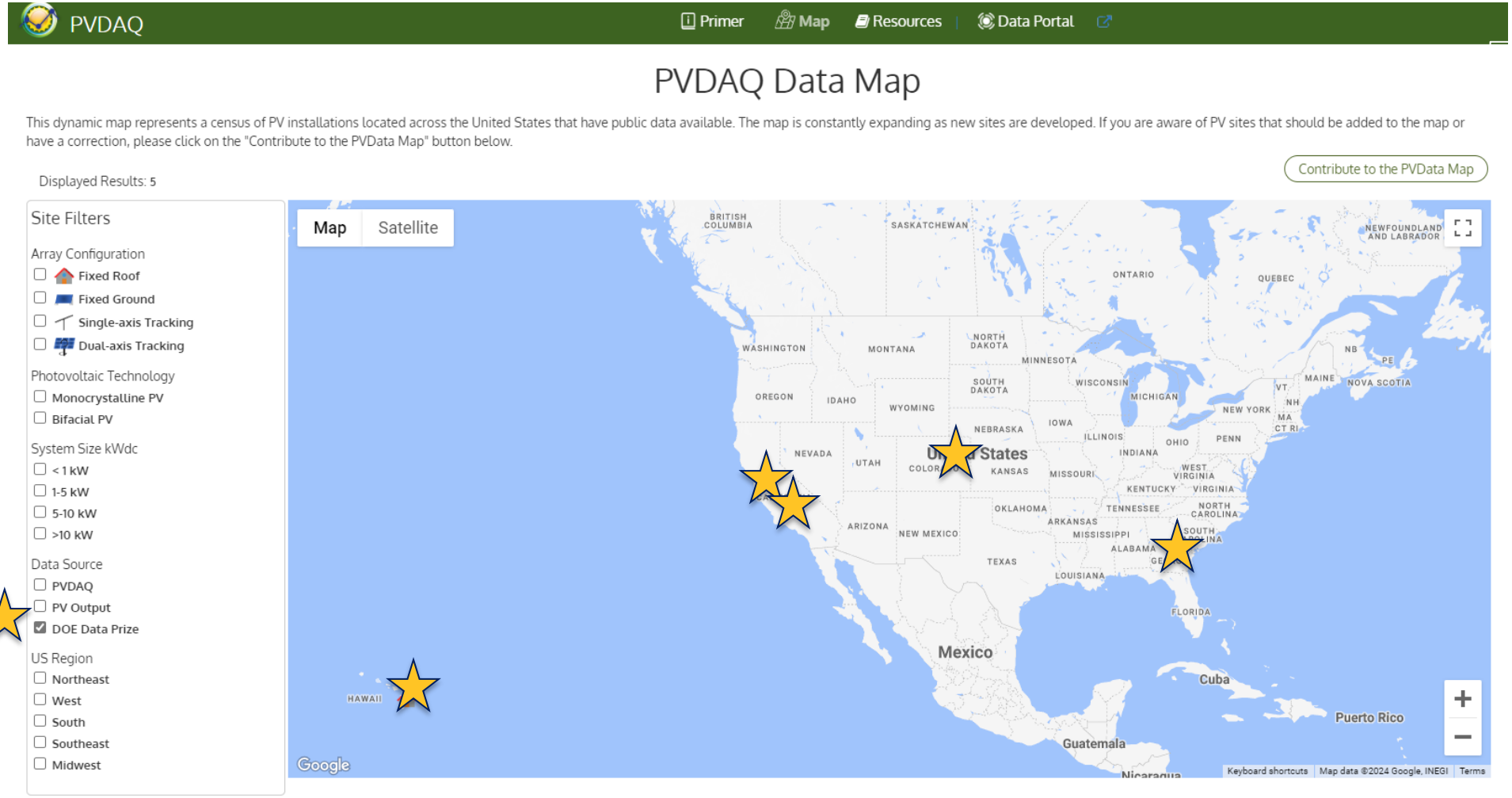
Through the Solar Data Bounty Prize there are now:

- 5 different systems across 4 US states
- 110 kW_{dc} to 257 MW_{dc} system sizes
- 417 GB of data
- > 4 billion data points
- > 9,500 sensor channels
- 6.6 years average
- 10 sec. to 15 minutes time resolution

Data available on OEDI and through PVDAQ.

See the Data!

Solar Data Bounty Prize Datasets Intro



See the Data!

https://openei.org/wiki/PVDAQ/PVData_Map

Shine On Solar Facility

State	California
Size (kW_{dc})	257,600
PV Technology	Multi-Si
Array Configuration	Single-axis tracking
Years of data	7.0
Minimum temporal resolution	10 second
Channels	4086
Dataset Size (GB)	392.3

Fixed Ground

Simon Solar Farm

State	Georgia
Size (kW_{dc})	38,687
PV Technology	Multi-Si
Array Configuration	Fixed Ground
Years of data	7.8
Minimum temporal resolution	5 min.
Channels	4798
Dataset Size (GB)	23.3

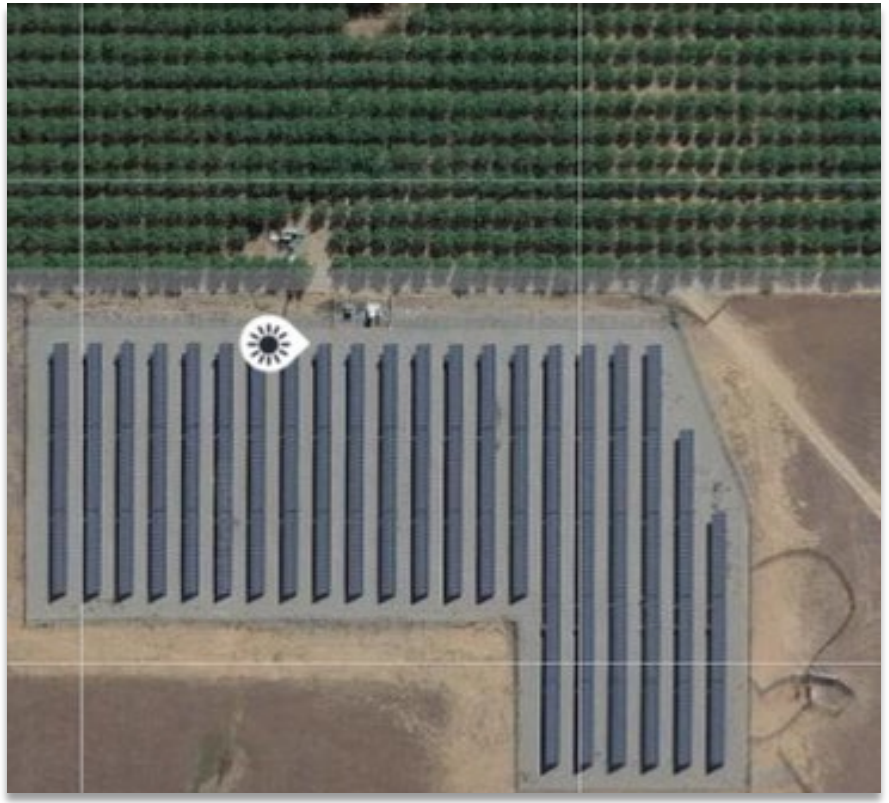
SR CO

State	Colorado
Size (kW_{dc})	4,738
PV Technology	CdTe
Array Configuration	Single-axis tracking
Years of data	6.2
Minimum temporal resolution	5 min.
Channels	438
Dataset Size (GB)	1.53

Fixed Ground

Farm Solar Array

State	California
Size (kW_{dc})	893
PV Technology	mono-Si
Array Configuration	Fixed Ground
Years of data	6.9
Minimum temporal resolution	5-15 min.
Channels	125
Dataset Size (MB)	445



Maui Ocean Center

State	Hawaii
Size (kW_{dc})	110
PV Technology	mono-Si
Array Configuration	Fixed Roof
Years of data	4.9
Minimum temporal resolution	5-15 min.
Channels	57
Dataset Size (MB)	199

Datasets Content & PV Prize Site Details

Robert White
National Renewable Energy Laboratory

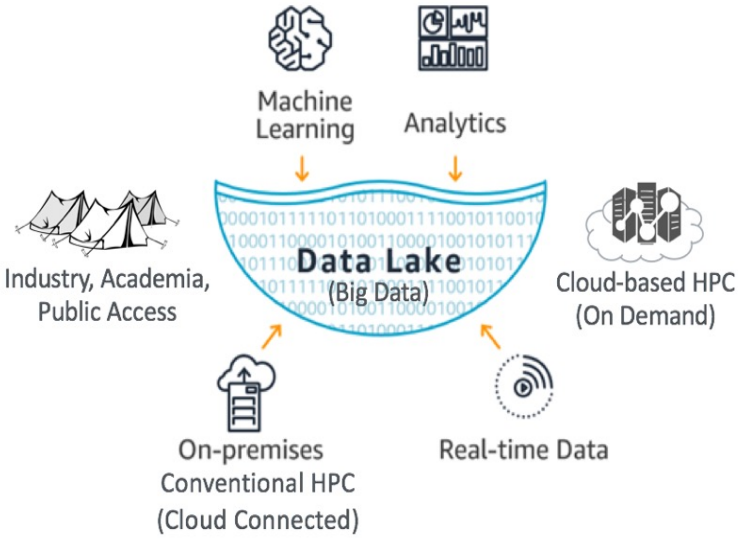
AMERICAN MADE
U.S. DEPARTMENT OF ENERGY

What is OEDI?

Open Energy Data Initiative

A partnership between NREL, DOE, Amazon, Microsoft, and Google to provide universal access to big data in the cloud

- Centralized Data System (Data Lake)
  - Cloud-based, single access point
  - Reduce duplication of effort, storage
  - Less transfer time, more research time



Value of Sharing Data

Success should be measured not when a project is completed or an experiment concluded, but when scientific and technical information is disseminated.

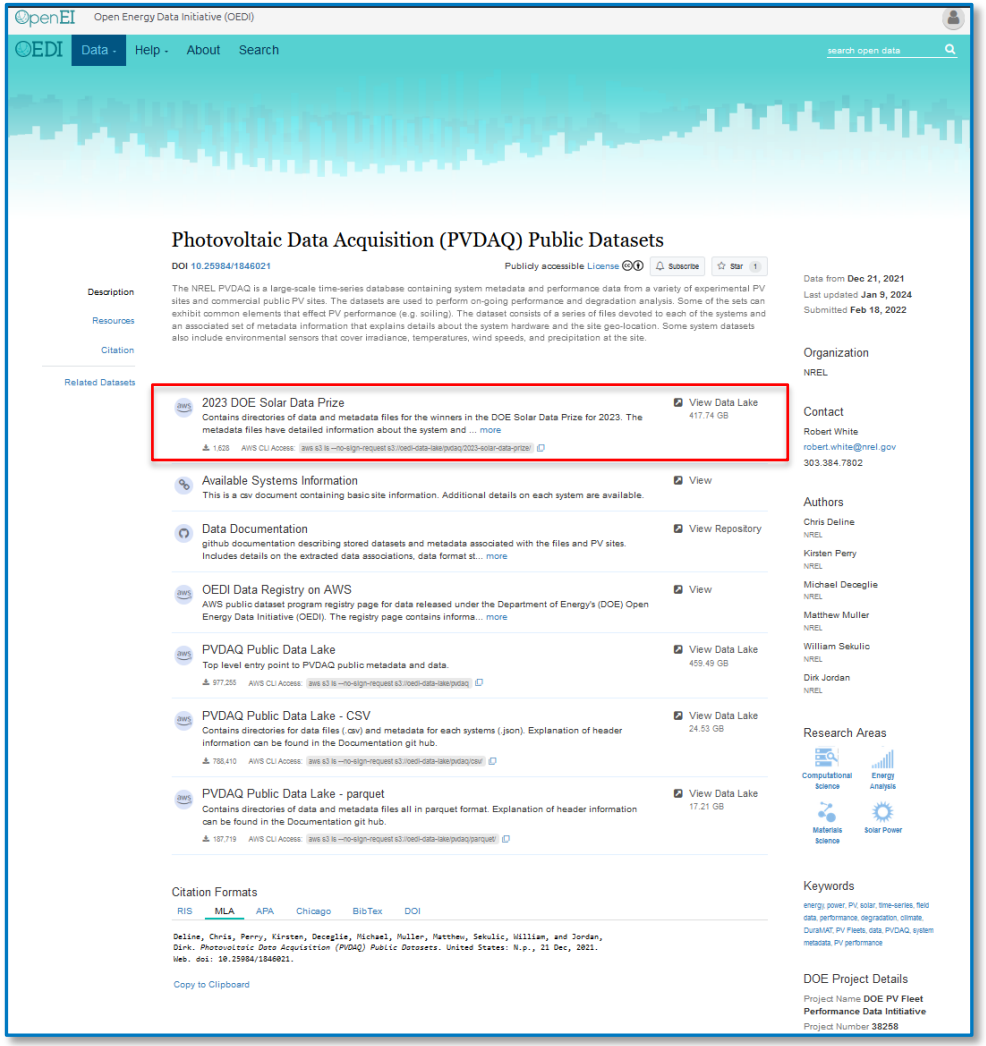
- DOE Strategic Plan, May 2011, p. 43-44

FY24 Q1 Impact Report

3,612,832	3,622,127	1.94 PB	9,620
Downloads in FY24 Q1	Downloads FYTD	Submitted Data	Available Resources

Archiving the Solar Bounty Prize Data

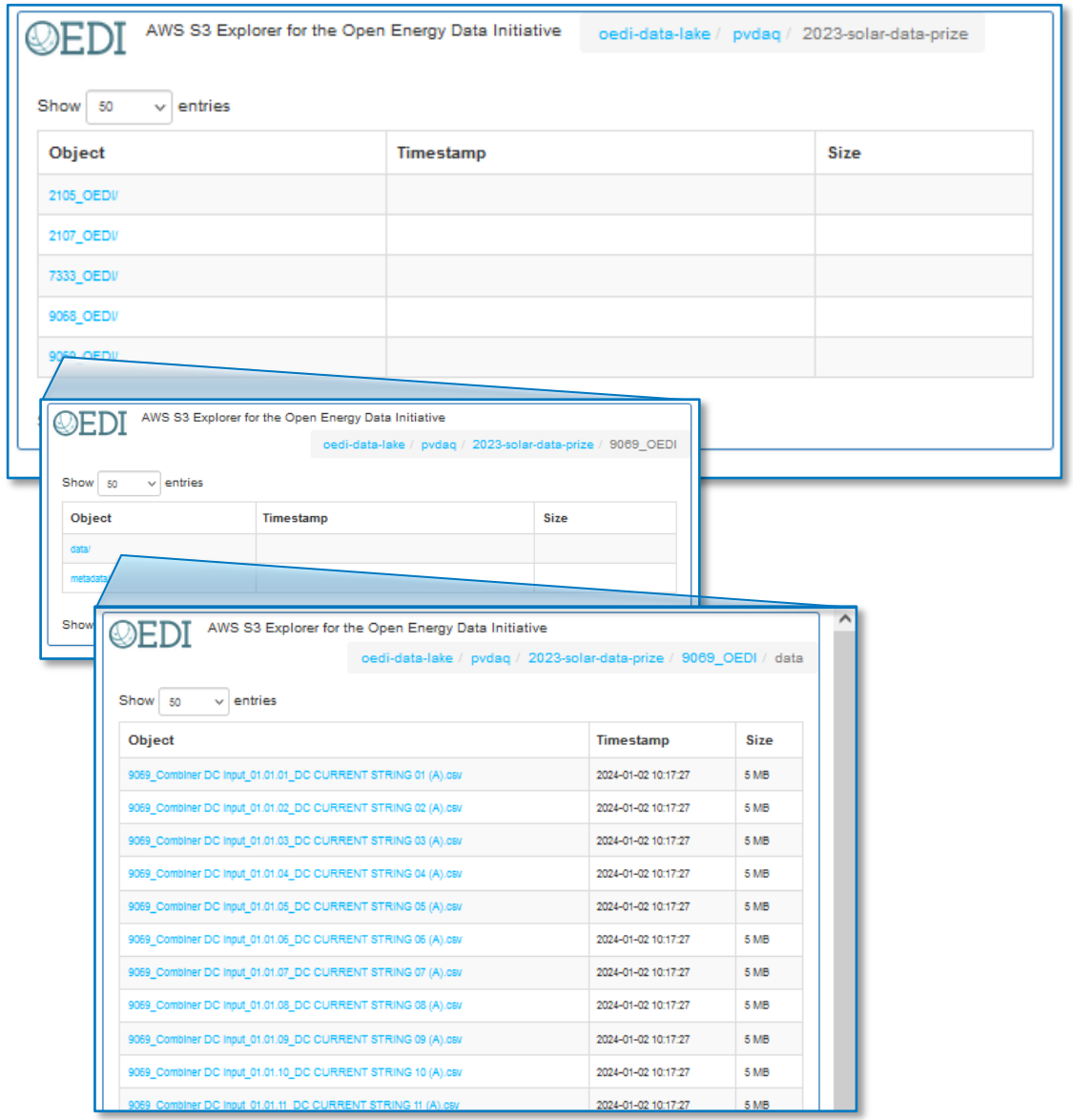
- Archived in the Open Energy Data Initiative Data Lake – PV Data Acquisition(PVDAQ) portfolio.
- The files are within an AWS Simple Storage Solution(S3) repository.
- Other public PV site data is also available on this site.



https://data.openei.org/submissions/4568

What's in the Solar Bounty Prize Data?

- Each site is listed by a unique number
- Within each Site folder is:
  - A data folder
    - Multiple data files
  - A metadata folder
    - A single descriptive JSON file describing the
      1. Site
      2. Hardware
      3. Sensor or Metric Channels



What Are the Details of the PV Sites?

Track A (< 5MW)

ID	Name	Power (DC)	Mounting	Resolution	Data Size
2105	Maui Ocean Center	110 kW	Fixed	5 & 15 minute	445 MB
2107	Farm Solar Array	893 kW	Fixed	5 & 15 minute	199 MB
9068*	SR_CO	4.8 MW	Tracking	5 minute	1.53 GB

Track B (> 5MW)

ID	Name	Power (DC)	Mounting	Resolution	Data Size
7333*	Shine-On Solar Facility	257.6 MW	Tracking	10 second	392.3 GB
9069	Simon Solar Farm	38.69 MW	Fixed Tilt	5 minute	23.3 GB

**Category Winner*

Metadata File Content vs. Data File Content

& Intro to Downloading and Citing

AMERICAN MADE
U.S. DEPARTMENT OF ENERGY

What's in the Metadata Files?

Each JSON Metadata file is divided into a series of sections describing a site.

System – Base metadata, timezone, power, etc.
Site – Geo location of PV Array

Hardware

- Mounts
- Inverters
- Modules
- Trackers
- Combiner boxes
- Weather Stations, Pyranometers, etc.

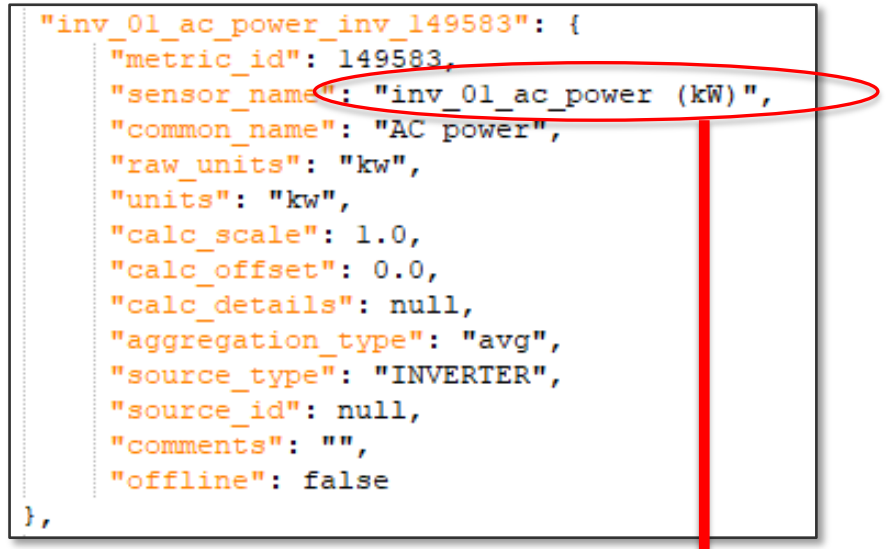
Metrics

- Details on each sensor channel from the hardware names, units, aggregation type, etc.

What's in the Data Files?

(Linking them to the Metadata)

- Datafiles are all CSV format
- All columns have descriptive headers
  - Each column header will match a name in the metadata JSON file under "metrics":
- The first column in each file is a timestamp in ISO 8601 FORMAT -> YYYY-mm-dd HH:MM:SS
- The dataset may be divided into files by hardware, timestamp, or a combination.



iso timestamp	inv_01_dc_current (A)	inv_01_dc_voltage (V)	inv_01_ac_power (kW)
2017-11-01 09:00:00	11.919	739.342	10.406
2017-11-01 09:05:00	12.932	737.205	11.084
2017-11-01 09:10:00	13.906	737.972	11.866
2017-11-01 09:15:00	14.769	735.418	12.599

How to Review the Sites and Detailed Information

Go to the Interactive Website at:

https://openei.org/wiki/PVDAQ/PVData_Map

DEMO

How to Download the Data

Method A

1. Go to the OEDI Website: https://data.openei.org/submissions/4568
2. Navigate to a Data Prize Winner's folder
3. Click on the filename.

DEMO

Method B **Use the Python AWS BOTO3 application to access the archived data**

1. Go to our repository: https://github.com/NREL/pvdaq_access
2. Clone the repository
3. Run the code.

DEMO

Method C **Using python, and the s3fs and pandas modules, open files directly in code:**

```
import s3fs
data = pd.read_csv("s3://oedi-data-lake/pvdaq/2023-solar-data-prize/2107_OEDI/data/2107_meter_15m_data.csv")
```

Citing the Solar Bounty Prize Data

The Data is part of the PVDAQ Open Energy Data Initiative datasets and is citable using the following:

Deline, Chris, Perry, Kirsten, Deceglie, Michael, Muller, Matthew, Sekulic, William, and Jordan, Dirk. *Photovoltaic Data Acquisition (PVDAQ) Public Datasets*. United States: N.p., 21 Dec, 2021. Web. doi: 10.25984/1846021.

This and other citing formats can be found at the bottom of the OEDI-PVDAQ landing page.

Questions?

www.nrel.gov

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 Solar Energy 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.

Photo from iStock-627281636

NREL
Transforming ENERGY