



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

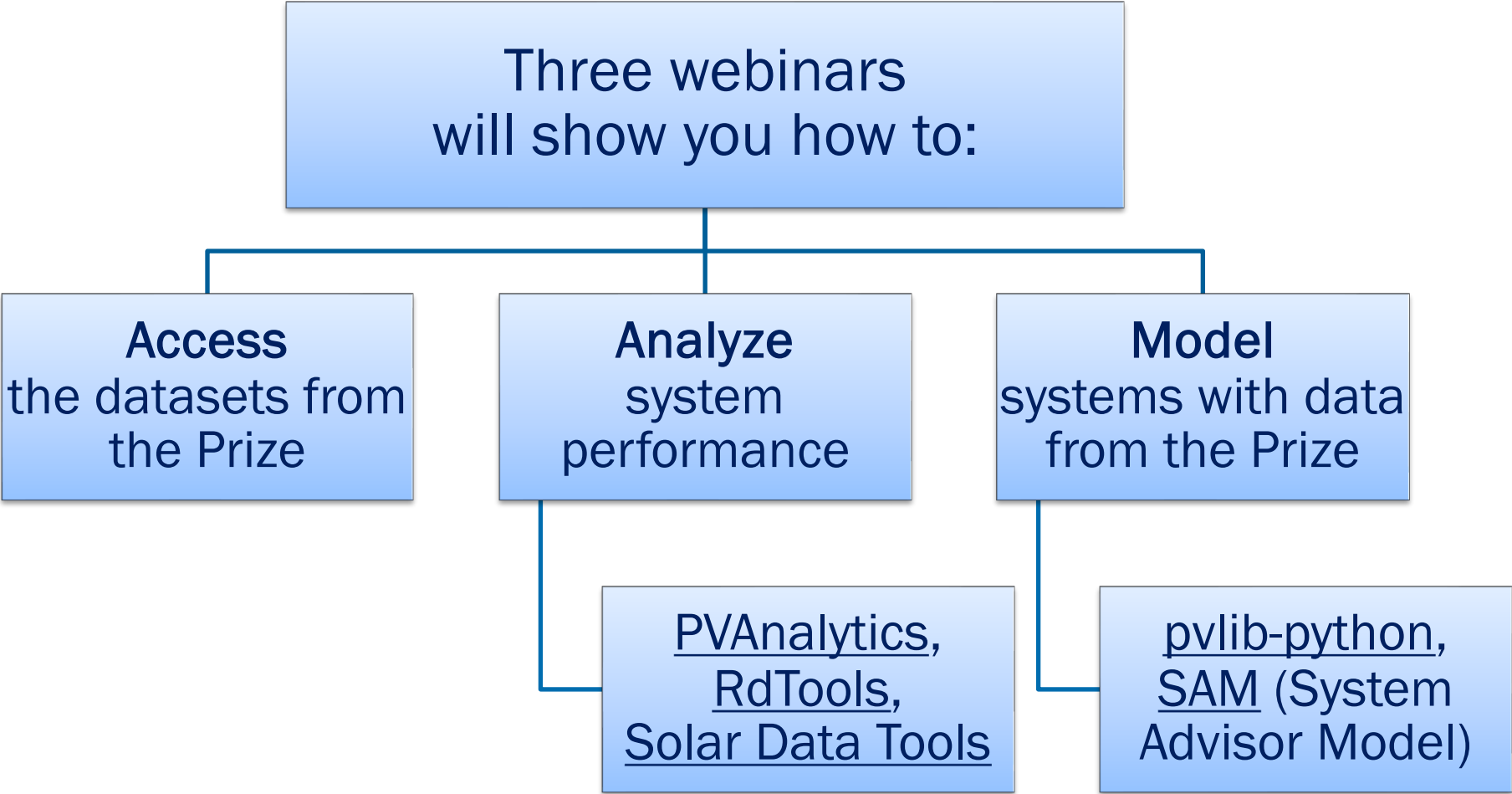
The winners' data sets are shared publicly via a dedicated platform:

[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

Congratulations to Our Prize Winners!

Track A: Medium-sized Systems



Track B: Large-sized Systems



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



PVDAQ Data Map

This dynamic map represents a census of PV installations located across the United States that have public data available. The map is constantly expanding as new sites are developed. If you are aware of PV sites that should be added to the map or have a correction, please click on the "Contribute to the PVData Map" button below.

Displayed Results: 5

[Contribute to the PVData Map](#)

Site Filters

Array Configuration

- Fixed Roof
- Fixed Ground
- Single-axis Tracking
- Dual-axis Tracking

Photovoltaic Technology

- Monocrystalline PV
- Bifacial PV

System Size kWdc

- < 1 kW
- 1-5 kW
- 5-10 kW
- >10 kW

Data Source

- PVDAQ
- PV Output
- DOE Data Prize

US Region

- Northeast
- West
- South
- Southeast
- Midwest



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





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

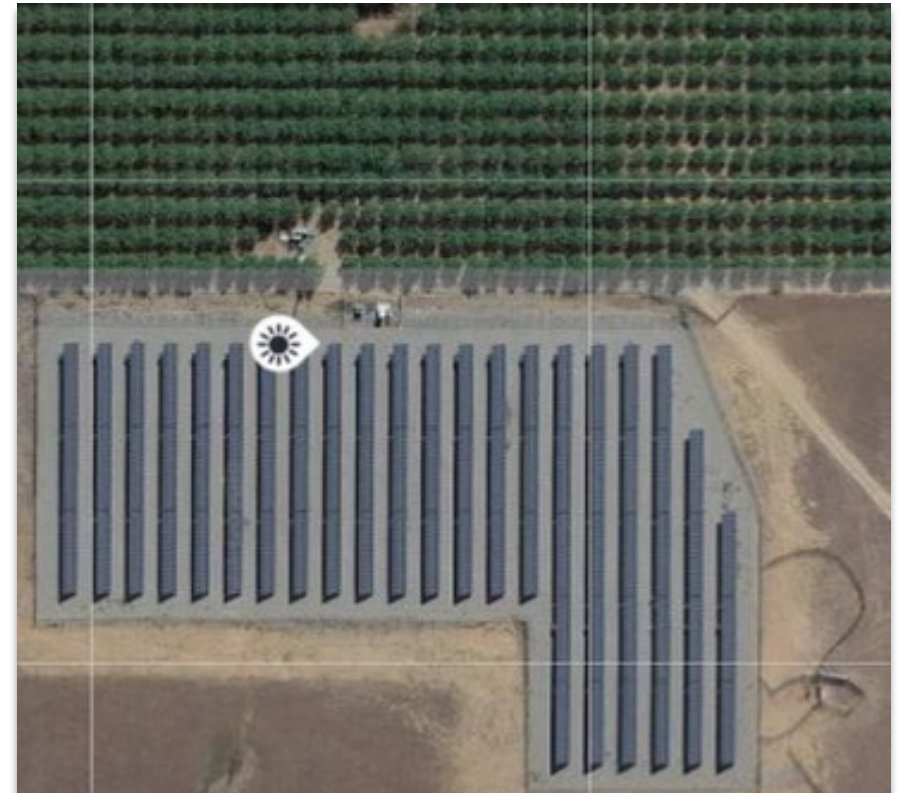
State	Colorado
Size (kW _{dc})	4,738
PV Technology	CdTe
Array Configuration	Single-axis tracking
Years of data	6.2
Minimum temporal resolution	5 min.
Channels	438
Dataset Size (GB)	1.53





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





Fixed Roof

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

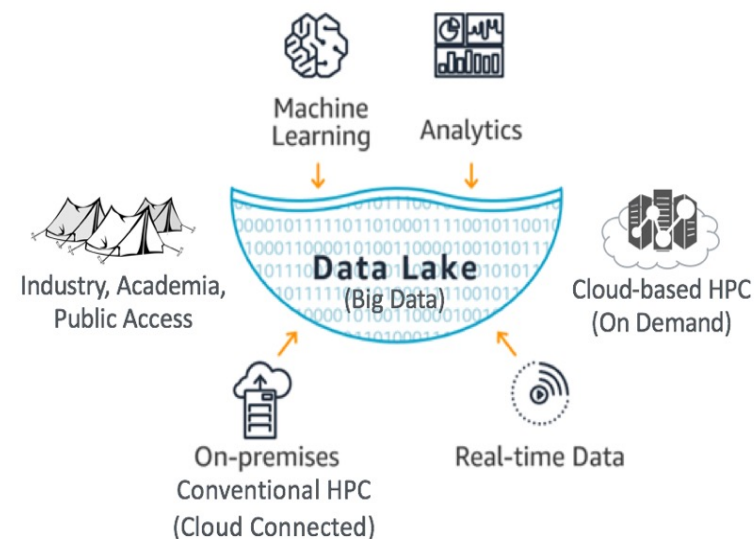
National Renewable Energy Laboratory

What is EDI?

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

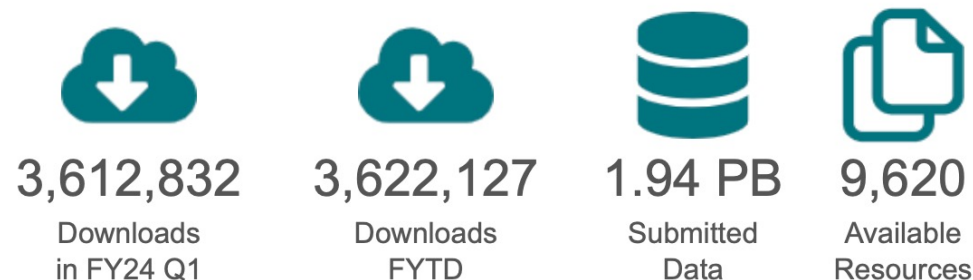


Value of Sharing Data

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



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>

The screenshot shows the Open Energy Data Initiative (OEDI) website. The main heading is "Photovoltaic Data Acquisition (PVDAQ) Public Datasets" with a DOI of 10.25984/1846021. The page is publicly accessible and has a license of CC-BY. The description states that the NREL PVDAQ is a large-scale time-series database containing system metadata and performance data from a variety of experimental PV sites and commercial public PV sites. The datasets are used to perform on-going performance and degradation analysis. Some of the sets can exhibit common elements that affect PV performance (e.g. soiling). The dataset consists of a series of files devoted to each of the systems and an associated set of metadata information that explains details about the system hardware and the site geo-location. Some system datasets also include environmental sensors that cover irradiance, temperatures, wind speeds, and precipitation at the site.

The "Related Datasets" section lists several datasets, with the "2023 DOE Solar Data Prize" dataset highlighted in a red box. This dataset is described as containing directories of data and metadata files for the winners in the DOE Solar Data Prize for 2023. The metadata files have detailed information about the system and more. The dataset is 417.74 GB and is available in the View Data Lake.

Other datasets listed include "Available Systems Information", "Data Documentation", "OEDI Data Registry on AWS", "PVDAQ Public Data Lake", "PVDAQ Public Data Lake - CSV", and "PVDAQ Public Data Lake - parquet".

The page also includes a "Citation Formats" section with options for RIS, MLA, APA, Chicago, BibTex, and DOI. The MLA format is selected, showing the citation: Deline, Chris, Perry, Kirsten, Deogill, 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.

Project Details: Project Name DOE PV Fleet Performance Data Initiative, Project Number 38258.

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

OEDI AWS S3 Explorer for the Open Energy Data Initiative oedi-data-lake / pvdaq / 2023-solar-data-prize

Show 50 entries

Object	Timestamp	Size
2105_OEDI/		
2107_OEDI/		
7333_OEDI/		
9068_OEDI/		
9069_OEDI/		

OEDI AWS S3 Explorer for the Open Energy Data Initiative oedi-data-lake / pvdaq / 2023-solar-data-prize / 9069_OEDI

Show 50 entries

Object	Timestamp	Size
data/		
metadata/		

OEDI AWS S3 Explorer for the Open Energy Data Initiative oedi-data-lake / pvdaq / 2023-solar-data-prize / 9069_OEDI / data

Show 50 entries

Object	Timestamp	Size
9069_Combiner DC Input_01.01.01_DC CURRENT STRING 01 (A).csv	2024-01-02 10:17:27	5 MB
9069_Combiner DC Input_01.01.02_DC CURRENT STRING 02 (A).csv	2024-01-02 10:17:27	5 MB
9069_Combiner DC Input_01.01.03_DC CURRENT STRING 03 (A).csv	2024-01-02 10:17:27	5 MB
9069_Combiner DC Input_01.01.04_DC CURRENT STRING 04 (A).csv	2024-01-02 10:17:27	5 MB
9069_Combiner DC Input_01.01.05_DC CURRENT STRING 05 (A).csv	2024-01-02 10:17:27	5 MB
9069_Combiner DC Input_01.01.06_DC CURRENT STRING 06 (A).csv	2024-01-02 10:17:27	5 MB
9069_Combiner DC Input_01.01.07_DC CURRENT STRING 07 (A).csv	2024-01-02 10:17:27	5 MB
9069_Combiner DC Input_01.01.08_DC CURRENT STRING 08 (A).csv	2024-01-02 10:17:27	5 MB
9069_Combiner DC Input_01.01.09_DC CURRENT STRING 09 (A).csv	2024-01-02 10:17:27	5 MB
9069_Combiner DC Input_01.01.10_DC CURRENT STRING 10 (A).csv	2024-01-02 10:17:27	5 MB
9069_Combiner DC Input_01.01.11_DC CURRENT STRING 11 (A).csv	2024-01-02 10:17:27	5 MB

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

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.


```
{
  "System": {
    "system_id": 2107,
    "public_name": "Farm Solar Array",
    "power(kW DC)": 893.0,
    "started_on": "2017-01-01 00:00:00",
    "comments": "Actual timezone is US/Pacific",
    "data_prize": "t",
    "timezone_code": "PST8PDT",
    "us_region": "West",
    "agripv": "f",
    "first_timestamp": "2018-12-15 0",
    "last_timestamp": "2023-11-15 1",
    "years_of_data": 4.9205,
    "data_time_resolution": "5-15 m",
    "number_data_channels": 125,
    "data_size": "445 MB"
  },
  "Site": {
    "site_id": 7211,
    "public_name": "FSA_1",
    "location": "Arbuckle, CA",
    "us_region": "West",
    "latitude": 38.996306,
    "longitude": -122.134111,
    "elevation(m)": 10.0,
    "climate_type": "Csa"
  },
  "Mount": {
    "Mount 0": {
      "mount_id": 17262
    }
  },
  "Inverters": {
    "Inverter 0": {
      "inverter_id": 29943,
      "name": "inv_fss_1",
      "manufacturer": "ABB",
      "model": "TRIO-27.6-TL-OUTD-S1B",
      "type": "string",
      "quantity": 1,
      "serial_num": "",
      "comments": "",
      "time_interval": "1L",
      "modules_per_string": "",
      "num_strings": ""
    },
    "Inverter 1": {
      "inv_01_ac_power_inv_149583": {
        "metric_id": 149583,
        "sensor_name": "inv_01_ac_power (kW)",
        "common_name": "AC power",
        "raw_units": "kw",
        "units": "kw",
        "calc_scale": 1.0,
        "calc_offset": 0.0,
        "calc_details": null,
        "aggregation_type": "avg",
        "source_type": "INVERTER",
        "source_id": null,
        "comments": "",
        "offline": false
      }
    }
  }
}
```

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.

```
"inv_01_ac_power_inv_149583": {  
  "metric_id": 149583,  
  "sensor_name": "inv_01_ac_power (kW)",  
  "common_name": "AC power",  
  "raw_units": "kw",  
  "units": "kw",  
  "calc_scale": 1.0,  
  "calc_offset": 0.0,  
  "calc_details": null,  
  "aggregation_type": "avg",  
  "source_type": "INVERTER",  
  "source_id": null,  
  "comments": "",  
  "offline": false  
},
```



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](https://doi.org/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-AC36-08G028308. 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

