

## PV Fleets Data and Metadata Requirements

To facilitate our ability to provide the best and most thorough PV performance and degradation analysis, we have constructed a set of guidelines to make the transfer and processing of the data as easy as possible.

### Where Is Our Data Housed?

Your data will be stored in our secure NREL Fleet Data Warehouse, which consists of file archives and a relational database within the government Amazon Web Services (AWS) cloud. The file archives are placed within AWS Simple Storage Solution (S3) drives, and the time-series database has been built in the AWS RedShift platform. Access is restricted to only the researchers working on your data.

### Harvesting Your Data

We can ingest data and metadata through a variety of automated and manual means. We currently have developed automatic API connections to several third-party monitoring services. To date, we can currently acquire data from these repositories:

- AlsoEnergy
- Power Factors.

If your data is housed on any of these platforms, we can leverage existing methods to bring in your data automatically (This will require account authorization to obtain data access). If your data is warehoused with a different provider, we are willing to investigate creating additional API access methods if possible.

The default method for us to acquire data if the above is not possible is through manual files, transferred via Box, Google Drive, FTP sites, or other similar means. In most cases, emailing large data files will not be possible, except perhaps in the case of smaller files containing system metadata information.

### Metadata

The following is a list of information that will help us develop the best analysis for your fleet. These metadata fields are for each site we will be analyzing. If these details are not all available, analysis may still be possible on a limited basis, depending on the parameter in question.

#### **Location:**

- Latitude – specify format: decimal degree, deg/min/sec
- Longitude – specify E/W sign convention
- Elevation – specify unit (ft, m)
- Time zone or UTC offset for data from site

#### **System:**

- Commercial operation date
- DC capacity

- AC capacity
- Mounting type: fixed, tracked
- Number of subarrays
- Subarray information (list or array of)
  - Azimuth (South = 180) or axis azimuth for tracker (N-S = 180)
  - Tilt
  - DC capacity
  - AC capacity
  - Ground Coverage Ratio
  - Shading factors (optional SunEye values)
- Module identifiers (manufacturer, model, power, and so on)
- Inverter identifiers (manufacturer, model, power, and so on)
- Tracker (optional)
  - Tracker manufacturer
  - Rotation limit, Backtracking algorithm (optional)
- Other Site Documentation (optional)
  - Construction or electrical 1-lines
  - O&M information (if available) such as O&M contract \$/W, % Availability per year, Truck rolls / year

## Time Series Data

For each of the measurements in your time series, we need to know the following metadata: This information is typically provided as an additional document, but in some of the third-party monitoring providers, this is extracted as part of our harvest of your data.

- **Unique name** – *Examples:* POA, GHI, DHI, DNI, ambient\_temperature\_1, module\_temperature, power, energy
  - Specificity on type (in some cases): AC or DC
- **Device type** – *Examples:* thermopile\_pyronometer, reference\_cell, photodiode, inverter, power\_meter, current\_transducer, voltmeter, thermocouple, rtd, rain\_gauge, anemometer, tracker\_position, status\_flags
- **Units** – *Examples:* kW, kWh, W/m<sup>2</sup>, Wh/m<sup>2</sup> degC, m/s, A, V
- **Data value type** – Mean, median, max, integral
- **Interval** – Left, right, center, instantaneous. *For instance,* right-labeled data applies the timestamp to the end of a data averaging interval (“hour-ending”) as is typical with dataloggers or TMY climate data.

## Metadata to Data Connection

Wherever possible we need to understand the associations between the hardware devices. Examples would be which irradiance sensors are tied to which inverter’s subarray or are the sensors site wide and which inverters are tied to a meter.

It is important for us to be able to connect which hardware instrument is producing each sensor measurement.

## Data

The time series value is a list of (timestamp, value) pairs, or vectors (timestamp, value1, value2, ... etc). Depending on the dataset, all instruments may be in a single timeseries with multiple columns or they may be delivered in distinct single-column time series.

In cases where the data is provided as a series of CSVs, we would like the data to be structured in the following format, if at all possible.

*Example:*

Timestamp	ac_power_1	ac_power_2	module_temp	poa_irradiance	ambient_temp
2018-07-01 12:34:56	75.69433	78.36291	27.51348	528.0596	30.64497
2018-07-01 12:49:56	75.69857	100.7741	27.19358	533.2748	30.63336

## Required Data

For every site or system we analyze, we need the following time-series measurements to be provided:

- **Irradiance** – Be sure to indicate the type (POA< GHI< DNI, etc.)
- **AC Power** (Revenue grade meter)
- **AC Power** (from Inverter) – *Strongly Preferred*
- **DC Power** (From Inverter) - *Optional*
- **Other Meteorological** (Ambient temperature, wind speed, module temperature etc.)

## Granularity

We would like the data to be in 15-minute intervals if possible.

## Concerning Time Stamps

We would like for all timestamps to be provided in ISO-8601 format: yyyy-mm-dd HH:MM:SS . We can do translations of existing format if the data is in another common format.

## Additional Important Information

We would like to know about any curtailment or other availability data (periods of maintenance outage, times during which system is under offtaker control, known periods of inverter clipping, etc.)

### System Topology Map (optional)

Individual power data will be identified in a nested format with inverters assigned to system meters and subarrays assigned to inverters. This will allow us to roll up sub-field production to validate overall site performance.

Meter (e.g., Meter 1) > Inverter (e.g. Inverter 2) > Subarray 3 (e.g. Strings 8–12)

Site electrical layout will allow us to match strings to inverters, and inverters to production meters.