# User Guide for the Distributed Solar Public Data

### Version Date: September 7, 2022.

## Background

Lawrence Berkeley National Laboratory (Berkeley Lab) collects project-level data on residential and non-residential photovoltaic (PV) systems. The data are sourced primarily from state agencies and utilities that administer PV incentive programs, solar renewable energy credit registration systems, or interconnection processes. In order to leverage this dataset for broader use, Berkeley Lab has issued a public data file, which can be downloaded at <http://trackingthesun.lbl.gov>.[[1]](#footnote-1) The public project-level dataset is updated once annually with data from the previous calendar year, and may also be updated on an interim basis as improvements to the data cleaning methodology and supplementary data fields are developed.

## What is Included in the Public Data File?

The data file includes only grid-connected residential and non-residential PV systems, defined to consist of rooftop systems, regardless of size, and ground-mounted systems up to 5 MWAC. Ground-mounted projects larger than 5 MWAC are considered utility-scale and are not included in this dataset.

The current version of the public data file includes more than 2 million PV systems installed through year-end 2021. The file includes 81 data fields describing key attributes of each system, which are listed and described in the table below. Note, though, that most fields are incomplete for most systems. The public data file is a single zip file, which contains a CSV file as well as this User Guide. Note that this file is too large to be loaded into Excel.

## What Data Cleaning Operations Are Performed?

The data collected for *Tracking the Sun* undergo extensive cleaning and quality control. Some elements of those operations are described in the table below. For additional information, please refer to Section 2 of the latest *Tracking the Sun* report (“Data Sources, Methods, and Sample Description”) and to Appendix A of the 2019 edition of the Tracking the Sun report, available for download [here](https://emp.lbl.gov/publications/tracking-sun-pricing-and-design).

**One important convention should be noted: Missing data are coded in the database as -1.**  Any operations performed on the data should therefore treat such values accordingly.

## Who to Contact with Questions?

Questions or comments specifically about the *Tracking the Sun* public data file may be directed to either Naïm Darghouth ([ndarghouth@lbl.gov](mailto:ndarghouth@lbl.gov)) or Galen Barbose ([glbarbose@lbl.gov](mailto:glbarbose@lbl.gov)).

**Data Fields in the Public Data File**

| Data Field Name | Description | Units | Details and Potential Data Quality Issues |
| --- | --- | --- | --- |
| data\_provider\_1 | Data Provider #1 | n/a | The entity that supplied at least a portion of the data, generally a utility or PV incentive program administrator. |
| data\_provider\_2 | Data Provider #2 (if applicable) | n/a | Another entity that supplied at least a portion of the data, generally a utility or PV incentive program administrator, if applicable. |
| system\_ID\_1 | System ID (from first Data Provider) | n/a | This is the system or application ID within the raw data file from the first data provider, if applicable. |
| system\_ID\_2 | System ID (from second Data Provider, if applicable) | n/a | This is the system or application ID within the raw data file from the second data provider, if applicable. A second system ID indicates that PV system data has been merged from two data providers. |
| installation\_date | Installation Date | date | For some data providers, the installation date may be based on the best available proxy, such as the date that an incentive claim was submitted or when the inspection was performed. |
| system\_size\_DC | System Size | kw (DC) | The total rated direct-current (DC) output of the module arrays at standard test conditions. These data are generally reported directly by the data provider, but in some cases must be estimated, for example, based on the module model and quantity or based on reported alternating-current (AC) capacity. |
| total\_installed\_price | Total Installed Price | dollars (nominal) | The total installed price for the system, prior to receipt of any incentives, as reported by the installer, host customer, or other incentive applicant. For third-party owned systems, the data may represent one of two things. If the third-party owner procured the system from an independent installation contractor, then the reported installed price likely refers to the intermediate sale price between the installation contractor and the third-party owner. If the third-party owner instead installed the system itself, then the reported installed price likely represents an appraised value. The installed price data may be subject to any number of other reporting inconsistencies, which may or may not be readily identifiable. In addition, the data may suffer simply from self-reporting errors, and the level of verification vary across data providers. |
| rebate\_or\_grant | Rebate or Grant | dollars (nominal) | The pre-tax value of any up-front rebate or grant provided by the entity supplying the data |
| customer\_segment | Customer Segment | n/a | Data on customer segment is mapped to one of six general types: RES, COM, SCHOOL, GOV, NON-PROFIT, and NON-RES, the last one being used only if more-specific information on non-residential customer type is unavailable. |
| expansion\_system | Expansion of an installed PV system | n/a | Indicates if the system is an expansion of a previously installed PV system. |
| multiple\_phase\_system | Part of a multiple phase PV system | n/a | Indicates if this is one part of a multiphase system. |
| new\_construction | New Construction | n/a | Indicates if the system was installed at the time of building construction. Data generally available for only those states or utilities that have separate programs or incentive rates for new construction vs. retrofits. |
| tracking | Tracking | n/a | Indicates if the system includes tracking equipment |
| ground\_mounted | Ground Mounted | n/a | Indicates if the system is ground-mounted (which may include pole-mounted systems). PV systems consisting of a combination of rooftop and ground-mounted arrays are coded as ground-mounted. |
| zip\_code | Zip Code | n/a | Host customer zip code (+4, when applicable) |
| city | City | n/a | Host customer city. Spellings have not been corrected or standardized. |
| state | State | n/a | Host customer state |
| utility\_service\_territory | Utility Service Territory | n/a | The electric utility service territory, when reported directly by the data provider; this data has not been cleaned or standardized. |
| third\_party\_owned | Third-Party Owned | n/a | Indicates if the system is third-party owned; that is, owned by an entity other than the site host and either leased or sold under a power purchase agreement to the site host. |
| installer\_name | Installer Name | n/a | Reported data, prior to being cleaned, is particularly "messy" given the complex spellings of models. These data have been cleaned and the spellings standardized to the extent feasible. |
| self\_installed | Self-Installed | n/a | Indicates if the system was installed by the site-host. |
| azimuth\_1 | Azimuth #1 | degrees | The horizontal direction of the array, where 180 degrees defines South facing PV orientation. Azimuth data reported by data providers was, in some cases, modified to adhere to this convention. Data fields are provided for up to three array orientations, though some systems may consist of a larger set of distinct orientations. |
| azimuth\_2 | Azimuth #2 | degrees |
| azimuth\_3 | Azimuth #3 | degrees |
| tilt\_1 | Tilt #1 | degrees | The vertical tilt of the array, where zero degrees corresponds to a flat array. As with the azimuth data, fields are provided for up to three array orientations, though some systems may consist of a larger set of distinct orientations. |
| tilt\_2 | Tilt #2 | degrees |
| tilt\_3 | Tilt #3 | degrees |
| module\_manufacturer\_1 | Module Manufacturer #1 | n/a | Reported data, prior to being cleaned, is particularly "messy" given the complex spellings of models. These data have been cleaned and the spellings standardized to the extent feasible. |
| module\_model\_1 | Module Model #1 | n/a |
| module\_quantity\_1 | Module Quantity #1 |  | Reported number of modules in system with manufacturer and model #1. |
| module\_manufacturer\_2 | Module Manufacturer #2 | n/a | These data have been cleaned and the spellings standardized to the extent feasible. |
| module\_model\_2 | Module Model #2 | n/a |
| module\_quantity\_2 | Module Quantity #2 |  | Reported number of modules in system with manufacturer and model #2. |
| module\_manufacturer\_3 | Module Manufacturer #3 | n/a | These data have been cleaned and the spellings standardized to the extent feasible. |
| module\_model\_3 | Module Model #3 | n/a |
| module\_quantity\_3 | Module Quantity #3 |  | Reported number of modules in system with manufacturer and model #3. |
| additional\_modules | Additional module model | n/a | Indicates whether there are more than three module models. |
| technology\_module\_1 | Technology, Module #1 | n/a | Identifies the module technology type. This is determined by cross-referencing module manufacturer and model names against equipment specification data available through solarhub.com and the California Energy Commission's list of eligible equipment. |
| technology\_module\_2 | Technology, Module #2 | n/a |
| technology\_module\_3 | Technology, Module #3 | n/a |
| BIPV\_module\_1 | BIPV Module #1 | n/a | Indicates if the modules are building integrated photovoltaics (BIPV), bifacial, and their nameplate capacity. These are determined by cross-referencing module manufacturer and model names against equipment specification data available through solarhub.com and the California Energy Commission's list of eligible equipment. |
| BIPV\_module\_2 | BIPV Module #2 | n/a |
| BIPV\_module\_3 | BIPV Module #3 | n/a |
| bifacial\_module\_1 | Bifacial Module #1 | n/a |
| bifacial\_module\_2 | Bifacial Module #2 | n/a |
| bifacial\_module\_3 | Bifacial Module #3 | n/a |
| nameplate\_capacity\_module\_1 | Nameplate Capacity, Module #1 | n/a |
| nameplate\_capacity\_module\_2 | Nameplace Capacity, Module #2 | n/a |
| nameplate\_capacity\_module\_3 | Nameplace Capacity, Module #3 | n/a |
| efficiency\_module\_1 | Efficiency, Module #1 | percent | Identifies the energy conversion efficiency of the modules. This is determined by cross-referencing module manufacturer and model names against equipment specification data available through solarhub.com and the California Energy Commission's list of eligible equipment. |
| efficiency\_module\_2 | Efficiency, Module #2 | percent |
| efficiency\_module\_3 | Efficiency, Module #3 | percent |
| inverter\_manufacturer\_1 | Inverter Manufacturer #1 | n/a | Reported data, prior to being cleaned, is particularly "messy" given the complex spellings of models. These data have been cleaned and the spellings standardized to the extent feasible. |
| inverter\_model\_1 | Inverter Model #1 | n/a |
| inverter\_quantity\_1 | Inverter Quantity #1 |  | Reported number of inverters in system with manufacturer and model #1. |
| inverter\_manufacturer\_2 | Inverter Manufacturer #2 | n/a | These data have been cleaned and the spellings standardized to the extent feasible. |
| inverter\_model\_2 | Inverter Model #2 | n/a |
| inverter\_quantity\_2 | Inverter Quantity #2 |  | Reported number of inverters in system with manufacturer and model #2. |
| inverter\_manufacturer\_3 | Inverter Manufacturer #3 | n/a | These data have been cleaned and the spellings standardized to the extent feasible. |
| inverter\_model\_3 | Inverter Model #3 | n/a |
| inverter\_quantity\_3 | Inverter Quantity #3 |  | Reported number of inverters in system with manufacturer and model #3. |
| additional\_inverters | Additional inverter models | n/a | Indicates whether there are more than three inverter models. |
| micro\_inverter\_1 | Microinverter, Inverter #1 | n/a | Indicates if the inverters identified are micro-inverters, solar/storage hybrid inverters, include built-in meters, and their output capacity. This is determined by cross-referencing inverter manufacturer and model names against equipment specification data available through solarhub.com and the California Energy Commission's list of eligible equipment. |
| micro\_inverter\_2 | Microinverter, Inverter #2 | n/a |
| micro\_inverter\_3 | Microinverter, Inverter #3 | n/a |
| solar\_storage\_hybrid\_inverter\_1 | Solar/Storage Hybrid Inverter #1 |  |
| solar\_storage\_hybrid\_inverter\_2 | Solar/Storage Hybrid Inverter #2 |  |
| solar\_storage\_hybrid\_inverter\_3 | Solar/Storage Hybrid Inverter #3 |  |
| built\_in\_meter\_inverter\_1 | Built-in Meter, Inverter #1 |  |
| built\_in\_meter\_inverter\_2 | Built-in Meter, Inverter #2 |  |
| built\_in\_meter\_inverter\_3 | Built-in Meter, Inverter #3 |  |
| output\_capacity\_inverter\_1 | Output Capacity, Inverter #1 |  |
| output\_capacity\_inverter\_2 | Output Capacity, Inverter #2 |  |
| output\_capacity\_inverter\_3 | Output Capacity, Inverter #3 |  |
| DC\_optimizer | DC Optimizer | n/a | Indicates if the system uses any DC Optimizers, based on the inverter manufacturer names. All systems using SolarEdge inverters are assumed to also include a DC optimizer. Systems using DC optimizers manufactured by other companies (e.g., Tigo) cannot be identified based on the inverter manufacturer; as such, the DC Optimizer field is coded as unknown for all systems with string inverters manufactured by companies others than SolarEdge. |
| inverter\_loading\_ratio | Inverter Loading Ratio | n/a | The Inverter Loading Ratio is the ratio of DC module capacity to AC inverter capacity. This is calculated from the reported or estimated values for System Size and System Inverter Capacity, described above. |
| dateOfBatteryInstall | Date of Battery Installation | date | For systems including storage only. Similar to the PTO date for the PV system, for some data providers, the battery installation date may be based on the best available proxy, such as the date that an incentive claim was submitted or when the inspection was performed. |
| battery\_rated\_capacity\_kWh | Battery Rated Capacity | kWh |
| battery\_model | Battery Model | n/a | As reported by the data provider. |
| battery\_rated\_capacity\_kW | Battery Rated Capacity | kW | If a storage system is installed at the same time as the solar system, these fields include the kW and kWh capacity of the storage system. |
| battery\_rated\_capacity\_kWh | Battery Rated Capacity | kWh |  |

## Disclaimer

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, or The Regents of the University of California. Ernest Orlando Lawrence Berkeley National Laboratory is an equal opportunity employer.

For more information on the Electricity Markets & Policy Group, visit us at [www.emp.lbl.gov](http://www.emp.lbl.gov)  
For all of our downloadable publications, visit <http://emp.lbl.gov/reports>



1. The public data file excludes any data provided under confidentiality agreements as well as other sensitive information that data providers requested to be withheld. [↑](#footnote-ref-1)