Solar Big Data & Informatics Read Ahead

David Brower Center, Berkeley, CA

February 18-19, 2014

SunShot U.S. Department of Energy

Agenda

February 18, 2015

8:00	**Registration and Breakfast**
9:30	**Welcome and Introduction** ○ Elaine Ulrich, DOE Solar Energy Technology Office
10:00	**Panel: Freeing the Data: Regulatory and Legal Barriers to Data Proliferation (Overview of California Solar Initiative and Similar Datasets)** ○ Lewis Bichkoff, California Public Utilities Commission/California Solar Initiative ○ Justin Baca, Solar Energy Industries Association ○ Elaina Lucas, Utility API ○ Lead Panelist: Galen Barbose, Lawrence Berkeley National Laboratory
10:45	**Coffee Break**
11:00	**Breakout Discussion – Methods of Disclosure** ○ Is there an ideal method for data disclosure? ○ What data points can be reported? ○ What complications stem from these different methods?
12:00	**Lunch**
1:00	**Panel: Exploring Technical Issues and Interoperability with Solar Data** ○ Carolyn Moses, Energy Information Administration ○ Tefford Reed, Enphase Energy ○ Tom Tansey, SunSpec Alliance ○ Lead Panelists: Debbie Brodt-Giles, National Renewable Energy Laboratory & Bosco So, Presidential Innovation Fellow
1:45	**Breakout Discussion – Ideal Data Standards** ○ What are the technical issues for collecting and managing data? How can they be overcome? ○ What are best practices from other industries? ○ What value propositions require changes to data volume and velocity?
2:30	**Break**
2:45	**Discussion: Setting the Stage for Innovation** ○ What solutions were identified in breakouts? ○ How do technical and legal issues compound or alleviate each other? ○ What technical and legal prerequisites can enable innovation?
3:30	**Wrap Up and Overview of Thursday's Discussion**
4:00	**Adjourn**

February 19, 2015

8:00	**Registration and Breakfast**
9:15	**Welcome and Introduction**
9:30	**Tapping Value Streams – The Economic Potential of Solar Data and Informatics** ○ Vikram Aggarwal, EnergySage ○ Jason Kaminsky, KwH Analytics ○ Lead Panelist: Ben Hoen, Lawrence Berkeley National Laboratory
10:30	**Coffee Break**
10:45	**Breakout Discussion** ○ What gaps (data too limited, disaggregated, or nonexistent) are limiting the full economic potential of solar? ○ What players can be active in this space but aren't? ○ Are there value streams throughout the solar lifecycle are not being monetized?
11:45	**Lunch**
12:45	**Discussion: Key Themes and Next Steps** ○ What roles are appropriate for DOE to play? Where does the private sector fit in? ○ What industries have expertise or best practices that can be applied to solar? ○ What barriers can data remove? ○ How do these regulatory, technical, and economics issues all interact?
1:30	**Present Recommendations**
2:15	**Closing Remarks**
2:45	**Adjourn**

Introduction

As solar energy has transitioned from a nascent option to a growing and viable energy source providing over 2% of the nation's electricity capacity, access to high quality, robust and open datasets have emerged as a key need of the industry. Data access and quality have been cited as a key opportunity to improve processes as wide and varied as consumer protection, financing, insurance, real estate, and grid integration. However, challenges with legal and regulatory barriers to collection, and technical issues with collecting and managing data have prevented the industry from achieving its full economic potential.

Existing datasets have already been instrumental in the growth of the solar industry. For example, the California Solar Initiative required participants to disclose system data when they register for state incentives. While originally intended as a compliance mechanism, CSI served as a de facto database of roughly half of the PV systems in the US. This data has been instrumental in informing the industry and has helped facilitate tremendous growth in the market. Furthermore, the transparency created by the CSI disclosure process has helped push the solar industry towards established best practices, enabling them to self-regulate in a rapidly evolving market.

As state incentive programs begin to expire, new challenges and opportunities will arise with regards to solar data. This workshop will help to identify the current landscape of available data, as well as gaps in the market, and develop recommendations for improvement. Deep dives around the themes of regulatory and legal issues, technical capabilities and economic value streams will help to illuminate current offerings, successes from other industries and market need.

Notable Datasets

California Solar Initiative

The California State Incentive (CSI), provides rebates to customers who go solar in the in the territories of California's three largest Investor Owned Utilities: Pacific Gas and Electric (PG&E), Southern California Edison (SCE), and San Diego Gas and Electric (SDGE). In order to receive state incentives, users had to agree to release their data into the open and comprehensive CSI database. Historically, the CSI database has served as a de facto list of all PV systems operating in the state of California. System cost, size, characteristics and installer were all compiled and made available to the public. Many researchers and analysts have relied on CSI's data to gauge the state of the solar industry, and determine ways to respond to a rapidly evolving market.

With California dramatically reducing the number of PV systems that receive incentives, the number of systems contained in the CSI data has diminished. However, there was a successful push to make data disclosure a requirement for interconnection, which happens regardless of incentives. In November 2014, the California Public Utility Commission (CPUC) voted unanimously to include data requirements in the interconnection application. While California remains the largest solar market in the US by a large margin, similar efforts by other states could help to create an even larger and more useful resource, and therefore improve access to solar, and, potentially, lower costs.

California Eligible Equipment List

In addition to a disclosure requirement, CSI set technical requirements to ensure that state subsidized systems were of a sufficient quality. In order to be eligible to receive subsidies in CA a system had to be installed with modules that "(1) have safety certification (ANSI/UL 1703) from a Nationally Recognized Testing Laboratory and (2) submit electrical characterization data tested by a third party laboratory," and inverters that "(1) have safety certification (UL 1741) from a Nationally Recognized Testing Laboratory (NRTL) and (2) submit conversion efficiency data tested by a NRTL."[1] These modules and inverters were compiled into a public list that was maintained on the CSI website, and frequently updated. The California Eligible Equipment List offered consumers a quick and easy way to verify that the equipment they were purchasing was thoroughly vetted and of an adequate quality. Ensuring the this list or a similar database is current and comprehensive is an important resource for consumer protection purposes.

Other Datasets

Many other states also report system characteristic data, but the quality and reporting frequency can vary widely. States use different data standards, and report different aspects of the PV systems (e.g. some states require module and inverter models to be reported, while others only require a capacity rating). Therefore, integrating these datasets can be difficult. Additionally, data not traditionally required by incentive programs, such as production data or appraisal data could have a multitude of uses for the solar industry. Likewise industry can modify and expand public datasets into more useful formats. A sample of relevant datasets is included as an appendix to this document.

Data Collection in a Post-Incentive World

Several methods have been proposed to ensure that datasets can remain available after state incentives are discontinued. The first is transitioning information and disclosure requirements from the incentive application to the interconnection application, as was done in CA by the CPUC. Several other states have looked to follow California's example. This would ensure data on all net metered systems is recorded, and has been cited as an especially efficient process that will not create unnecessary paperwork or red tape.[2]

Suggestions for disclosure of characteristics of larger utility-scale systems, which might fall outside the interconnection process noted above, have included shifting reporting to the Energy Information Agency (EIA) or the Federal Energy Regulatory Commission (FERC). Both these solutions have been cited as potentially reducing paperwork burden of installing a solar system, making it faster, easier and cheaper to go solar.[3]

[1] California Senate Bill 1 (Murray, Chapter 132, Statutes of 2006)
[2] See CPUC ruling for arguments.
[3] Forms EIA-860 and FERC-556 are required for the interconnection of systems larger than 1 MW. These forms could be easily expanded to incorporate solar specific data disclosure

Potential Uses of Solar Data

Financing

Data disclosure has been cited as a key opportunity to lower the cost of capital for solar energy systems, and investor are typically unwilling to dedicate funds without datasets substantial enough to minimize their risk. Providing investors with high-quality performance data can help increase their confidence in the viability of investing in solar energy and ensure that solar is seen as an asset. KwH Analytics and NREL/SunSpec Alliance's oSPARC platform have been designed to provide investors with the data needed to make financial decisions, but there is a demonstrated market need for more comprehensive data.

Real Estate

The real estate community depends on granular data to help consumers easily make decisions and compare the relative costs and benefits of different properties. Confusion surrounding the value PV adds to home value has prevented customers from entering the solar market. Researchers have developed guides and tools for appraising homes with PV, based on extensive market data primarily gathered from California home sales.[4] However, the expansion of the quality and availability of these datasets would improve the ability of home appraisers to accurately value PV systems.

Safety & Standards

Recently, concerns have emerged surrounding the lack of a whole system shutoff for PV systems, and the possibility of a live and damaged PV system presents safety concerns for firefighters and other first responders. An up-to-date database of the locations of PV systems would be of important use to these responders, as onboard computers could be programed to notify firefights of the presence of rooftop PV systems when they are responding to structural fires.

Microgrids and Grid Integration

Solar presents clear opportunities for grid resiliency and optimization. Providing clear information on an area's generating capacity and load presents opportunities for detailed feeder modeling, and microgriding.

Market, Policy & Regulatory Analysis

Much of the market and policy analysis on solar has used data provided by these state incentive programs. Continued access to datasets will be required for analysts to continue and expand their work, which has been instrumental in informing industry, consumers, and policymakers.

Benefits of Expanded Solar Data

Benefits to Solar Industry

- Access to open and comprehensive data will spur growth of solar market.
- Data can ease customer acquisition, as installers can quickly identify underdeveloped areas.

[4] Hoen, Ben, Sandra Adomatis, Thomas Jackson, Joshua Graff-Zivin, Mark Thayer, Geoffrey T. Klise, and Ryan H. Wiser. *Selling Into the Sun: Price Premium Analysis of a Multi-State Dataset of Solar Homes*., 2015.; PVvalue.com

- Improved forecasting can lead to feeders more tolerant of high concentrations of PV.
- Filing processes can be streamlined, leading to decreased paperwork burden.
- Growth of available performance data will help improve investor confidence and lower the cost of capital.

Benefits to Utilities

- Improved understanding of renewable load will allow for a more targeted approach to deployment. A more detailed understanding of feeder saturation points will be clear.
- Consistent and standardized data will allow for easy development of other software tools to assist utilities in understanding and capitalizing off of the growing solar market

Benefits to ISOs/RTOs and Balancing Authorities

- Enhance forecasting and generation models will allow ISOs/RTOs to better predict the impact of variable generation.

Benefits to Regulators

- Continued timely, accurate and comprehensive analysis is essential to informing policymakers of potential issues and improvements they will face as the way we produce and consume electricity shifts in the coming years.

Benefits to Consumers

- Data can enable simpler and more targeted solar deployment, making it faster, easier and cheaper for a customer to choose solar.
- Improved data transparency will make customers more aware of the cost and commitment of going solar, and innovative tools can be more easily developed to assist and educate customers.
- For customers that go solar, performance data can help them understand if their systems are underperforming. This data could be licensed to software engineers to allow customers easy access to software tools to help them better track their performance and consumption.
- Lowered cost of capital will make solar more economically viable for potential customers.
- Reliable data can demonstrate the added value of solar on both residential and commercial properties, ensuring that systems are adequately valued and customers see a return on investment.

Privacy Considerations

Hosts

As datasets have increased in timeliness and become less aggregated, concerns with privacy have also emerged. While some homeowners and privacy advocates may be wary of data disclosure requirements, the information contained in many proposed disclosures is by-and-large already publicly available, albeit in a less consolidated format. Most solar systems are visible from the street and all eventually become visible via satellite images. Google Maps, for example could be used to find all of the PV systems in a local jurisdiction, regardless of whether they choose to disclose their system data. Furthermore, the process of installing solar is captured by a public paper trail in the form of permits and applications. Companies like CoreLogic and Reed Construction Data compile and consolidate these paper trails into useful datasets, which they make available to their customers. Furthermore, real estate

multiple listing services (MLS) are compiled by companies like Zillow and Trulia into publically available datasets. Characteristics of most houses that have been on the open market typically have at least some of their characteristics listed publicly, and efforts are underway to make solar one of their surveyed characteristics.

Many customers that receive RECs or SRECs register their personal information into portals such as PJM's Generational Attribute Tracking Systems (GATS), and larger systems will have detailed records with FERC, both of which could serve as additional sources of info.

Consolidating all of this information into a single place may not have a substantial impact on personal privacy, and having an explicit point of disclosure this data could make data rights clearer for potential customers, who may be unaware of the extensive uses of their personal data already underway. Additionally, language could be incorporated to ensure that solar customers retain access to their data and can apply it for useful purposes.

Utilities

Records of utility scale systems are already relatively easy to access. Systems over the size of one megawatt have to report system characteristics, and often production data to EIA and FERC. The sitting and permitting processes of utility scale solar development is typically well documented and compressive, leaving a very long public paper trail.

Installers

Some solar companies have expressed concern about disclosing data as it might provide insight into their pricing, hardware preferences, or market share. However, installers have largely complied with reporting system data as a requirement to receive state incentives, and public companies have to disclose more comprehensive financial data to the SEC and other regulatory bodies.

Resources & Sample Datasets

State Datasets and Resources

California Public Utility Commission Ruling
This ruling provides the background and context for CPUC's decision to move data disclosure as a requirement to qualify for CSI incentives to a requirement for interconnection. It provides a summary of stakeholder testimony, and the methods for collection and list proposed data fields.

California Solar Statistics
The California Solar Initiative publishes the actual program data, exported from the CSI online application tool. Users can view program data summaries for all of California's solar programs. The site provides key statistics in several figures and tables, and users can download the complete Working Data Set for their own analysis.

NY SUN Data
NYSERDA compiles all residential and small commercial projects into a database and interactive map. New York and California Report similar system characteristics (size, cost, module and inverter characteristics, etc.). While address level data is not available, system's relative location can be seen on a map.

New Jersey Solar Installation Update
New Jersey makes both installed system, and planed system data available to the public. Unlike California and New York, they do not provide specific details on modules and inverters.

Massachusetts Interconnection Data
Massachusetts reports project information as they pass through the interconnection pipeline. This dataset is of similar detail to New Jersey's, and includes other technologies.

Other PV Deployment Databases

Generation Attribute Tracking System
Renewable systems that are eligible for RECs in PJM's territory often register in the GATS system for reporting. Users can filter by technology, but the dataset is less robust than many of the state-run systems.

The Open PV Project
The Open PV Project is a collaborative effort between government, industry, and the public that is compiling a comprehensive database of photovoltaic (PV) installation data for the United States. Data for the project are voluntarily contributed from a variety of sources including utilities, installers, and the general public. The data collected is actively maintained by the contributors and are always changing to provide an evolving, up-to-date snapshot of the US solar power market.

Utility Rate Datasets

Green Button Data
The Green Button Initiative allows households and businesses can use Green Button to access their own energy usage data from their electric utility. This data can be licensed or exported to different platforms and a growing set of companies are offering products, services, and apps that use Green Button data.
Utility Rate Database
This tool emphasizes the power of crowd-sourcing for helping the database have relevant and accurate information. There are currently over 40,000 rates in the database with more being added each day. These datasets are a great resource for a developer to insert sample utility rate data into an app related to nationwide utility rates.

Data Catalogues

NREL's Renewable Resource Data Center
NREL provides many tools and datasets related to solar energy (and other renewable energy). This site can direct users to information about the solar resource, map and GIS tools, and generation estimators like the System Advisory Model (SAM) and PVWatts.
Data.gov's Energy Datasets
Data.gov contains hundreds of datasets related to the energy industry. These datasets are open source, and can be compiled by anyone.
Open Energy Information
The Open Energy Initiative is a free, open source knowledge-sharing platform created to facilitate access to data, models, tools, and information that accelerate the transition to clean energy systems through informed decisions. Open EI currently has 1,338 downloadable datasets, and nearly 200 solar specific datasets.

Other Datasets

Residential Consumption Survey
This wide-ranging survey on residential energy details consumption patterns including demographics and housing characteristics
Solar PEIS
This site, administered by Argonne National Lab details solar resource data, and GIS data detailing the Bureau of Land Management administered land designated for potential solar development.
oSPARC
oSPARC is a user driven actuarial database to compare and assess long-term relative performance, health and reliability of Residential, C&I and Utility solar systems, with the goal of lowering the cost of borrowing via standardization and securitization.

Workshop Attendees

- Molly Buchan, Clean Power Finance
- Lakshmanan Venkatesan, Clean Power Finance
- Scott Tewel, Clean Power Research
- Jake Saper, Emcap
- Carolyn Moses, Energy Information Administration
- Vikram Aggarwal, Energy Sage
- Jamie Johnson, Energy Sense
- Tefford Reed, Enphase
- Robbie Adler, Faraday
- Justin Lee, Genability
- Jason Kaminsky, KWH Analytics
- Galen Barbose, Lawrence Berkeley National Laboratory
- Ben Hoen, Lawrence Berkeley National Laboratory
- Will Klein, Mosaic
- Debbie Brodt-Giles, National Renewable Energy Laboratory
- Robert Margolis, National Renewable Energy Laboratory
- Michael La Marca, NRG
- Femi Omitaomu, Oak Ridge National Laboratory
- Scott Hinson, Pecan Street Energy
- Steve Hanawalt, Power Factors
- Steve Scakes, Power Factors
- Mary Rottman, Rottman Associates
- Geoff Klise, Sandia National Laboratory
- Sita Kuteira, Solar Census
- Aaron Woro, Solar Census
- Justin Baca, Solar Energy Industries Association
- Michael Palmquist, Solar Nexus
- Cedric Brehaut, Solichamba
- Conor Farese, Soligent
- David Herrmann, Sun Number
- Ryan Miller, Sun Number
- Gary Wayne, SunRun
- TJ Keating, SunSpec
- Tom Tansy, SunSpec
- Elena Lucas, UtilityAPI
- Qingyao Kong, Wells Fargo
- John Previtali, Wells Fargo
- Dixon Wright, Wells Fargo

Department of Energy Staff

- Elaine Ulrich
- Ammar Qusaibaty
- Craig Connelly
- Bosco So
- Monica Andrews
- Dan Boff