Solar Big Data and Informatics

DATE: 03/16/2015

SUBJECT: Request for Information (RFI)

DESCRIPTION: In order to reduce solar transaction costs (soft costs) and increase transparency and efficient pricing, the U.S. Department of Energy's SunShot Initiative is exploring best strategies to support, expand, and streamline competitive and non-competitive efforts in the solar industry related to collecting, securing, managing, exchanging, and monetizing solar datasets, across the value chain in the U.S. from project origination to grid integration, operations, and decommissioning. Improved data availability reduces informational asymmetries, accelerates the maturity of solar related products and services, and increases solar deployment. Areas of key interest center on technical, economic, administrative, and legal barriers and opportunities to reduce costs of capital, lower operational risks, protect consumers, and increase efficient market activities.

BACKGROUND: 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 accessibility has been cited as a key opportunity to improve processes as wide and varied as consumer protection, financing, insurance, real estate, and grid integration. Existing data assets have already been instrumental in the growth of the solar industry. Ensuring their continued expansion will help expand access to solar and lower the overall cost of ownership. DOE has long taken an active role in compiling, hosting, disseminating, and enabling strategic use of datasets and tools (e.g. Green Button, Open EI, Open PV Project). Many of these datasets have been instrumental in expanding the growth of the U.S solar market. Funded by DOE and developed by National Renewable Energy Lab (NREL), irradiance data are widely used by most solar installers to estimate generation. Initiatives like the Utility Rate Database and Database of State Incentives for Renewables and Efficiency (DSIRE) have helped create transparency in the marketplace, and educate consumers and businesses about the economic benefits of solar energy in the U.S.

State governments have also made data accessibility and reporting a priority in their energy initiatives (e.g. New Jersey's Clean Energy Program: Solar) primarily to measure and quantify programmatic impacts. Since its inception, the California Solar Initiative (CSI), for example, has mandated the collection of data for all PV systems receiving state rebates for policy compliance reasons. System cost, size, characteristics and installer information were all compiled and made available to the public. The CSI database proved to be extremely valuable for numerous applications and businesses, serving as a de facto list of all PV systems operating in the state of California. Many market stakeholders (e.g. analysts, installers) have relied on CSI's data to gauge the state of the solar market in California and elsewhere. Like many other states, however, California has begun to dramatically reduce the number of PV systems that receive incentives, and the quality of data collected has been diminishing. As the solar market continues to grow rapidly, the need to treat solar datasets as strategic assets for a maturing industry has never been clearer. Reducing market uncertainty and facilitating efficient flow of data and information will protect consumers and lower costs. This in turn will increase market efficiency and boost consumer

confidence, thus contributing to the [SunShot goals](#) of making solar energy more accessible to all Americans.

PURPOSE: The purpose of this RFI is to solicit feedback from industry, academia, research laboratories, government agencies, and other stakeholders on issues related to technical, economic, administrative, and legal barriers limiting an efficient exchange of solar datasets among stakeholders. EERE is specifically interested in information on practical, industry-driven, financially enduring, and self-sustaining strategies that treat big data as strategic assets for the U.S. solar industry. This is solely a request for information and not a Funding Opportunity Announcement (FOA). EERE is not accepting applications.

DISCLAIMER AND IMPORTANT NOTES: This RFI is not a Funding Opportunity Announcement (FOA); therefore, EERE is not accepting applications at this time. EERE may issue a FOA in the future based on or related to the content and responses to this RFI; however, EERE may also elect not to issue a FOA. There is no guarantee that a FOA will be issued as a result of this RFI. Responding to this RFI does not provide any advantage or disadvantage to potential applicants if EERE chooses to issue a FOA regarding the subject matter. Final details, including the anticipated award size, quantity, and timing of EERE funded awards, will be subject to Congressional appropriations and direction.

Any information obtained as a result of this RFI is intended to be used by the Government on a non-attribution basis for planning and strategy development; this RFI does not constitute a formal solicitation for proposals or abstracts. Your response to this notice will be treated as information only. EERE will review and consider all responses in its formulation of program strategies for the identified materials of interest that are the subject of this request. EERE will not provide reimbursement for costs incurred in responding to this RFI. Respondents are advised that EERE is under no obligation to acknowledge receipt of the information received or provide feedback to respondents with respect to any information submitted under this RFI. Responses to this RFI do not bind EERE to any further actions related to this topic.

PROPRIETARY INFORMATION: Because information received in response to this RFI may be used to structure future programs and FOAs and/or otherwise be made available to the public, **respondents are strongly advised to NOT include any information in their responses that might be considered business sensitive, proprietary, or otherwise confidential.** If, however, a respondent chooses to submit business sensitive, proprietary, or otherwise confidential information, it must be clearly and conspicuously marked as such in the response.

Responses containing confidential, proprietary, or privileged information must be conspicuously marked as described below. Failure to comply with these marking requirements may result in the disclosure of the unmarked information under the Freedom of Information Act or otherwise. The U.S. Federal Government is not liable for the disclosure or use of unmarked information, and may use or disclose such information for any purpose.

If your response contains confidential, proprietary, or privileged information, you must include a cover sheet marked as follows identifying the specific pages containing confidential, proprietary, or privileged information:

This is a Request for Information (RFI) only. EERE will not pay for information provided under this RFI and no project will be supported as a result of this RFI. This RFI is not accepting applications for financial assistance or financial incentives. EERE may or may not issue a Funding Opportunity Announcement (FOA) based on consideration of the input received from this RFI.

Notice of Restriction on Disclosure and Use of Data:
Pages [list applicable pages] of this response may contain confidential, proprietary, or privileged information that is exempt from public disclosure. Such information shall be used or disclosed only for the purposes described in this RFI DE-FOA-0001318. The Government may use or disclose any information that is not appropriately marked or otherwise restricted, regardless of source.

In addition, (1) the header and footer of every page that contains confidential, proprietary, or privileged information must be marked as follows: "Contains Confidential, Proprietary, or Privileged Information Exempt from Public Disclosure" and (2) every line and paragraph containing proprietary, privileged, or trade secret information must be clearly marked with double brackets or highlighting.

EVALUATION AND ADMINISTRATION BY FEDERAL AND NON-FEDERAL PERSONNEL: Federal employees are subject to the non-disclosure requirements of a criminal statute, the Trade Secrets Act, 18 USC 1905. The Government may seek the advice of qualified non-Federal personnel. The Government may also use non-Federal personnel to conduct routine, nondiscretionary administrative activities. The respondents, by submitting their response, consent to EERE providing their response to non-Federal parties. Non-Federal parties given access to responses must be subject to an appropriate obligation of confidentiality prior to being given the access. Submissions may be reviewed by support contractors and private consultants.

REQUEST FOR INFORMATION CATEGORIES AND QUESTIONS:

CATEGORY 1: Solar Data Marketplace

1. Describe available datasets and resources for solar (residential, commercial, utility scale) or datasets that are relevant to the solar deployment and usage across the following data classes? Please provide details about data accessibility (e.g. public, private), availability (e.g. real-time, monthly) variety, volume, providers, users, uses, or any other key characteristic? Respondents may use Table 1 to answer this question.
   Suggested list of data classes:
   a. Customer Data (Demographics, financial information, experience with energy products)
   b. System Site Data (location, irradiance, utility data, zoning, electric rate structure, load profile, electric feeder characteristic, roof characteristics, shading, environmental characteristics)
   c. PV System- Level Data (array profile, inverter characteristics, communications capabilities, other system design characteristics, installer, system ownership/contract terms, incentives, critical dates (e.g. contract date), fees (e.g. permits))
   d. Fleet Operations & Maintenance (production, failure modes, equipment replacement)
   e. Other

2. What barriers or costs in the solar market could be addressed by increasing access to data-driven information?

3. What are the data gaps that exist today for solar (residential, commercial, utility)? What value can be derived by filling these gaps?

4. What datasets or information would increase solar deployments, reduce financial risks and boost consumer confidence? Which of these datasets would be the most impactful and what would its impact be?

5. What types of datasets/tools/capabilities are needed for faster, easier, better:
   a. Economic valuation of solar
   b. Consumer protection (e.g. contracts, certified installers, verified equipment, safety)
   c. Solar financing/investing
   d. Insurance (e.g. asset insurance, performance insurance, warranty/equipment insurance)
   e. Compliance with local regulations (e.g. permits, codes, safety)
   f. Real estate dynamics
   g. Grid integration, system operation/maintenance, and grid management/operations (forecasting, dispatchability, demand response).
   h. Transparency and self-regulation

6. How can solar datasets be improved to benefit multiple stakeholders? What are technical, legal, or economic barriers to improvement? How can the interests of these different stakeholders be incorporated into designing, and making these datasets available and accessible?

CATEGORY 2: Administrative/ Legal Considerations

7. What types of models for data collection and management exist, other than government-mandated or incentivized methods (e.g. California Solar Initiative, U.S. Energy Information Administration)? How can solar deployment datasets remain robust in a post-incentive world?

8. What roles do state, local & Federal governments play in data disclosure? What is the private sector's role?

9. How might private entities improve and make datasets available? How can this be done without disclosing proprietary business practices?

10. How can data disclosure increase benefits to consumers? What are the primary issues associated with privacy and solar data? What examples of privacy violations are relevant or commonplace in other industries?

11. For a given data type (deployment data, performance data) what rights might stakeholders (consumers, utilities, financial community, state and local program administers, etc.) have to confidentiality, IP, privacy, ownership, or other issues?

12. How might legal uncertainties about proper usage of datasets be reduced?

CATEGORY 3: Data Systems and Interoperability

13. What are the best practices for data management, quality, and security in other industries? What can the solar industry learn from these industries regarding data management?

14. What are the requirements (legal or otherwise) for sharing or exchanging data, including buying or selling? What are unique requirements for solar data? What obligations for data collection or data warehousing and sharing might exist?

15. How do you define aggregated or anonymized data? What are the costs and benefits associated with aggregated and/or anonymized data? What are the technical, legal and/or financial challenges to high resolution data and/or data quality?

16. What level of data granularity (spatial, temporal, data variables) best balances industry needs and costs?

17. How might a multi-use industry data standard be created? What might those standards look like?

18. What other technical challenges are preventing data exchange and access? As state level solar incentive programs are phased out, there has been a trend to migrate data collection and disclosure traditionally associated with those programs to the interconnection process (e.g. California's proposed model).What are the potential costs and benefits of this model as compared to other approaches?

CATEGORY 4: Competitive/Innovative Data Driven Business Models

19. How might non-solar entities help the solar industry collect/analyze/monetize solar data?

20. Who is buying and/or selling solar data? What successes can the solar industry build on? What business models can incentivize consumers to release their data?

21. What tools already exist in the marketplace that can be (or are being) applied to solar data?

22. The value of data is often abstract until there is a mature product. How do other industries monetize data, and how might the solar industry begin to monetize its data?

23. What value streams throughout the solar lifecycle might be monetized?

24. What "blue skies" exist for solar data? What might the future look like?

REQUEST FOR INFORMATION RESPONSE GUIDELINES: Responses to this RFI must be submitted electronically to solardata@ee.doe.gov no later than 5:00pm (ET) on April 10, 2015. Responses must be provided as attachments to an email. It is recommended that attachments with file sizes exceeding 25MB be compressed (i.e., zipped) to ensure message delivery. **Only electronic responses will be accepted.**

Please identify your answers by responding to a specific question or topic if possible. Respondents may answer as many or as few questions as they wish.

EERE will not respond to individual submissions or publish publicly a compendium of responses. A response to this RFI will not be viewed as a binding commitment to develop or pursue the project or ideas discussed.

Respondents are requested to provide the following information at the start of their response to this RFI:
- Company / institution name;
- Company / institution contact;
- Contact's address, phone number, and e-mail address.

Table 1 – Optional Format to Respond to Question 1

Name of Dataset	Green Button *(example)* **http://www.greenbuttondata.org/**	<name/link>	<name/link>	…
Provider/ Owner	NIST developed the data standard. Participating utilities own the data.	<details>	<details>	…
Category	Account-level electricity consumption data.	<details>/ Choose an item.	<details>/ Choose an item.	…
Accessibility	Open and free to rate payers accessing their data.	<details>/ Choose an item.	<details>/ Choose an item.	…
Availability	Monthly.	<details>/ Choose an item.	<details>/ Choose an item.	…
Data Sources	Utility reported.	<details>	<details>	…
Uses and Users	Virtual Energy Audits, Usage Reconciliation, Benchmarking Resource Usage, Validating Efficiency Goals, Efficiency Programs, Fact Based Service Deployment, Management Policy Validation, Quality of Service Analysis, Operational Management, Benchmarking.	<details>	<details>	…
Volume (Entries)	87 utilities committed. Data standard by NIST provides the list of all variables that can be used to implement the standard.	<details>	<details>	…
Variety/ Quality	Variable depending on which data variables the utilities choose to provide but data variable formats are consistent.	<details>	<details>	…
Other	There is a number of APIs to help developers write new applications (Green Button Connect).	<details>	<details>	…