DEEDS: A Next-Generation Scientific Data Sharing and Analyzing Platform for PV Applications

Reza Asadpour*
Dept. of Electrical and Computer Engineering
Purdue University
West Lafayette, USA
rasadpou@purdue.edu

Muhammed Tahir Patel*
Dept. of Electrical and Computer Engineering
Purdue University
West Lafayette, USA
patel707@purdue.edu

Ann Christine Catlin
ITaP Research Computing
Purdue University
West Lafayette, USA
acc@purdue.edu

Muhammad A. Alam
Dept. of Electrical and Computer Engineering
Purdue University
West Lafayette, USA
alam@purdue.edu

***Abstract*—An ever-expanding photovoltaics (PV) community has been publishing enormous amounts of data regarding all aspects of PV technology. The data generated via these studies and reports range from experiments, simulation, and standard tests to policies and economics of PV. These data require organization, systematic storage, and analyses. Previous works have focused on research-specific data environments and repositories. However, a comprehensive discipline-neutral platform for preserving, sharing, and analyzing the data has not been built. Digital Environment for Enabling Data-Driven Science (DEEDS) provides a unique solution to this problem. DEEDS enables a user to create datasets (projects), cases, and tools; and store data which can be structured, compared, and numerically analyzed, all on a single holistic online platform. In this paper, we demonstrate the capabilities of DEEDS using an example research study called the Solar PV Diagnosis. DEEDS platform has the potential to be used by the entire PV community to preserve various PV projects, interpret their performance and reliability, and to facilitate worldwide collaboration.**

Keywords*— *Data analysis, efficiency, degradation, PV diagnosis.

I. Introduction

We are witnessing an era of data-abundance. In every research and industrial study, an enormous amount of data is generated. The field of photovoltaics does not remain far from it. Ranging from the materials and process research to system-level solar farm deployment, from experiments to numerical simulations, and from standardized tests to economic policies, all aspects of PV technology generate the voluminous dataset. The structured storage and systematic analysis of these data are as essential as their generation. This methodology enables productive use of data to be applied in real-life applications and allows sharing the data for research and industrial collaborations. Moreover, the data and corresponding analysis should be easily discoverable and available to the PV community for future research.

Previous works on research data management have focused on research-specific customized environments. For example, sites like the Encode Genome Browser and the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) are tailor-made for the genomic research community. Thus, there is a need for a holistic discipline-neutral online platform to preserve, share, and analyze data. The Registry of Research Repositories (www.re3data.org) is a valuable resource that provides access to 2000 repositories for research data that encompasses both structured data and repository files. In the PV community, DuraMAT (https://datahub.duramat.org/) [1] provides a platform to save and organize data related to several PV projects in the USA. However, both the above-mentioned platforms lack data analysis tools and scientific workflows.

Digital Environment for Enabling Data-Driven Science (DEEDS, https://datacenterhub.org/) [2]–[4] is a web-based platform that provides storage, exploration, and sharing of datasets with additional unique features of preserving the complete scientific workflows – from input raw data to output results and reports – and analyzing tools all in one place. Moreover, the platform can be used for any scientific discipline including multi-disciplinary collaborations. Within the PV community itself, all the verticals of PV research namely materials, chemical processes and fabrication, device-level

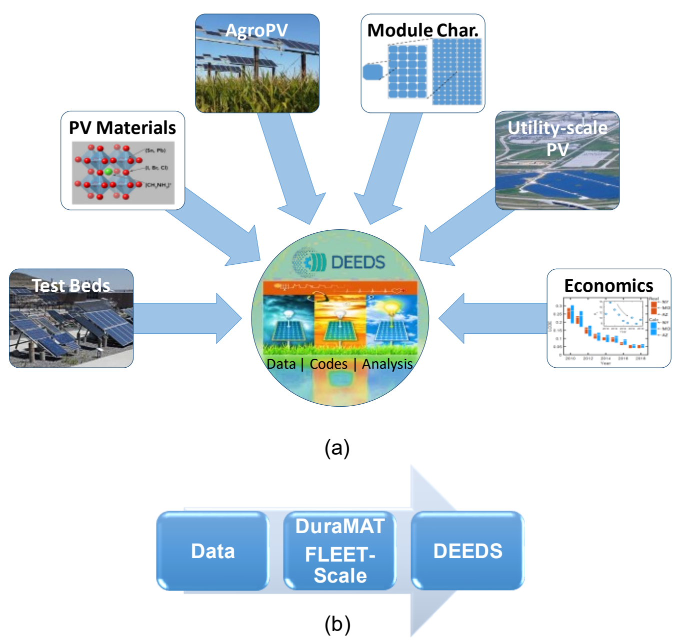
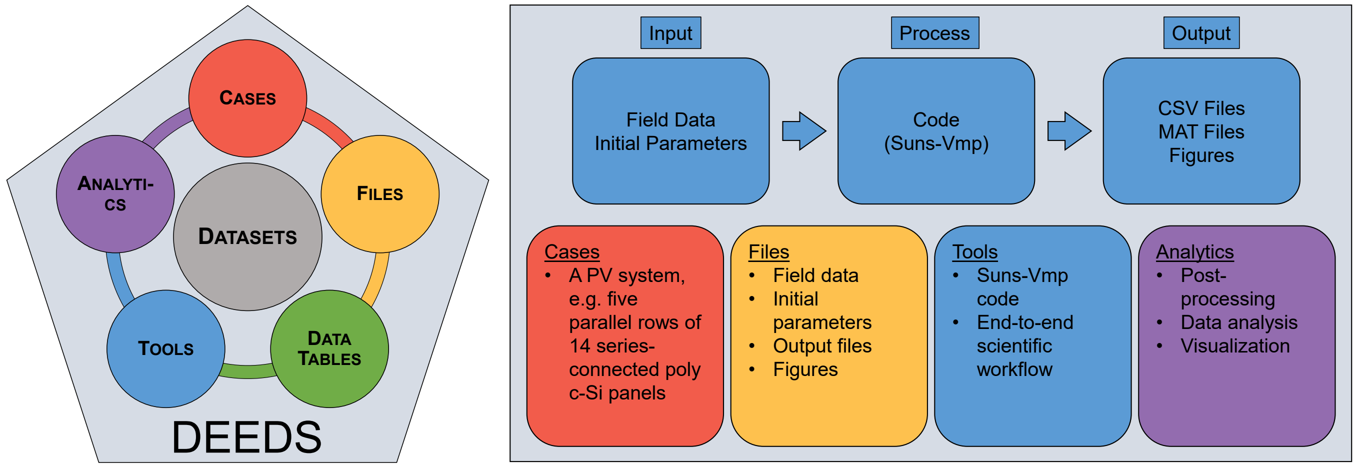


Fig. 1. (a) A schematic representation of DEEDS working as a common platform for all aspects of PV research. (b) DEEDS is capable of hosting analytical tools such as degradation modeling, a natural next step after storing the data in repositories, for e.g. DuraMAT, Fleet-Scale, etc.

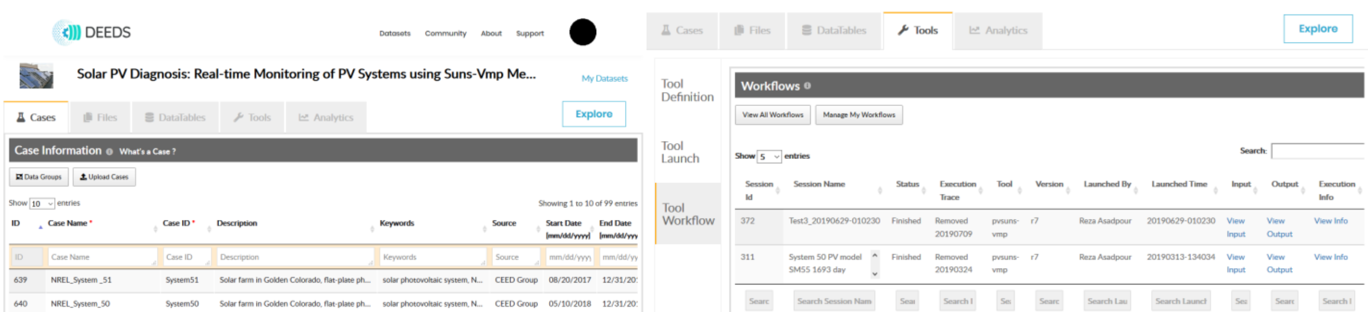
This work is supported by the National Science Foundation under Grant No. #1724728.

* The authors contributed equally to this work.



(a) (b)

Fig. 2 (a) Hierarchy of DEEDS organizing a dataset. (b) An example of hierarchy and process flow of DEEDS Solar PV application.



(a) (b)

Fig. 3 Screenshot of (a) A typical DEEDS dataset showing Cases, Files, DataTables, Tools, and Analytics. (b) Tools tab displaying the end-to-end scientific workflows.

simulation and characterization, farm-level installations, utility-scale analyses, and economic policymaking, can be simultaneously preserved including complete scientific processes, workflows, and analyses, as shown in Fig. 1.

In this paper, we showcase the capabilities of DEEDS using an example of a PV research study (dataset) called the Solar PV Diagnosis. The study requires raw IV data, weather data, and pristine module parameters as input. It then runs a processing code to find the degradation of PV module parameters as the output. In the next section, we discuss the framework of DEEDS and how it can be used for the Solar PV Diagnosis project.

II. DEEDS Framework

DEEDS as an environment for data management offers a solution to structuring mass data generated by PV farms or tests. The hierarchy of the DEEDS platform is summarized in Fig. 2a.

A. Dataset

The dataset includes all the information needed for the project. The information includes and is not limited to case studies, trials, specimens, experiments, measurements, and other related data. Fig. 3a shows a screenshot of a typical dataset with **Cases, Files, DataTables, Tools, and Analytics**.

The first section of the dataset is called "**Cases**" that has the information about the specimens, trials, location of the experiment or other means to categorize the projects or case studies. For example, a **case** could be a PV system at a certain location, an IEC test, or a specific PV material being explored.

Once the case is established the user can store the files related to the case. There can be different types of files that include input files, literature related to the project, manuals explaining the methods being used, output files, figures, codes, and other types depending on the project. Note that the user can create customized categories for their dataset.

The DataTables section is used to preserve, compare and explore quantitative and qualitative attributes of different cases. It manages data as different layers of MS Excel-like tabular data sheets that are linked to each other and support the representation of complex hierarchical and multi-dimensional data models.

In the Tools section, there are three sub-sections: **Tool definition, Tool Launch, and Tool Workflow**. In the definition section, one can upload a code (tool) that processes and analyzes the input data or import from a predefined set of tools in DEEDS collection. The code, for example, can be a MATLAB file that has multiple input arguments and the user can define the arguments. Once the tool is defined and verified the authorized users can launch the tool. Users must run the tool launch and select input files or input arguments (if necessary) and then choose a server and possible runtime that ensures enough time for the code to run completely. Note that the computation/

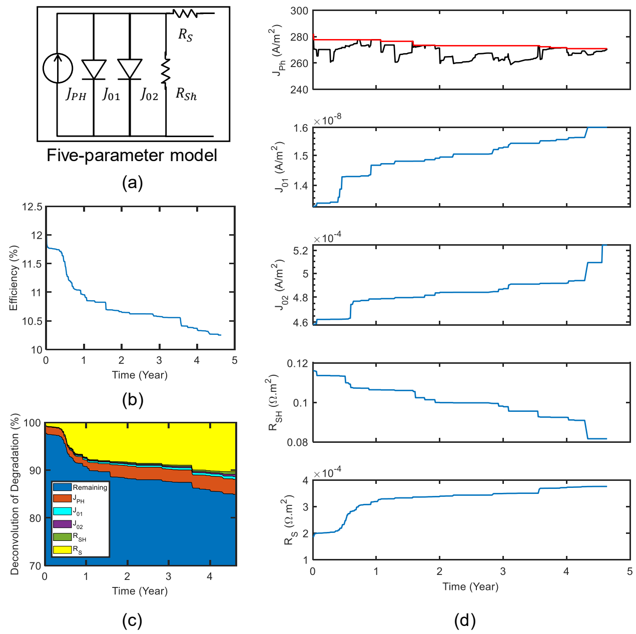


Fig. 4. (a) Five-parameter model equivalent of a module (b) the efficiency of the system as a function of time at nominal condition (c) the breakdown of role of each component of the five-parameter model in the degradation of the system (d) degradation of each of the five parameters as a function of time.

simulation tool can run on the DEEDS server or on DEEDS-linked HPC facilities to overcome the performance limits of locally-owned machines. The third section is tool workflow and is a unique capability provided by DEEDS (see Fig. 3b). It connects the input files to the code and then to the automatically uploaded output so that reproducibility and traceability are guaranteed. DEEDS workflow creates rich metadata that logs the input, output, user that launched the tool, the time of initiating a simulation, and input and output files related to and generated by the simulation. Such detailed logging helps with several aspects including tracking the users or groups involved, transition from a person that is leaving a project to a new member that is continuing the work. Tool workflow, in general, helps in preserving and sharing the end-to-end information related to an experiment/simulation and further facilitates multi-institutional collaborations.

Finally, in the R-based Analytics section, a user can visualize data, compare cases, perform regression analysis, and other customized numerical analyses. Users can also access the built-in Jupyter to use the Python programming language for analysis.

B. Explore: An overview of the dataset

The explore menu gives the user the ability to have access to the files belonging to each case. By clicking on one case the breakdown of the files based on the predefined categories will be shown. The user has access to the original tools (codes) by clicking on the Tools subsection here. Then the user can download the source files and view the arguments related to each source code.

III. An Example Dataset: Solar PV Diagnosis

In this section, we explain a solar PV diagnostic tool that was developed by Sun et al. [5]. The tool uses a method called Suns-Vmp to extract solar panel five-parameter out of maximum power points at different illumination during a day. The five parameters are: (a) J_{ph}, the photocurrent density generated due to illumination; (b) J_{01} is the diode recombination current with an ideality factor of one; (c) J_{02} is recombination current in the depletion region of the diode with an ideality factor of two; (d) R_{sh} is the shunt resistant; and (e) R_s, the series resistance. Since the illumination changes over a day, there is a unique set of five-parameter that can generate the same maximum power point as the system measures data.

With regards to DEEDS, the dataset is called **Solar PV Diagnosis** and the case is a system that consists of a string of five parallel rows of 14 series connected poly c-Si solar panels. Uploaded input files are (a) the initial five-parameter at the time of installation extracted from datasheet; (b) field data that includes measured maximum power points at half-hour intervals, and weather data that includes ambient temperature, illumination, and wind speed (see Fig. 2b). Users launch the tool which is the Suns-Vmp MATLAB code already developed and uploaded to DEEDS. The workflow logs input and output files and the status of the simulation during runtime i.e., whether it is still running, there is an error, or the simulation is completed.

Fig. 4 shows the auto-uploaded results of the four and a half years of field data as input. The method extracts the five-parameters as a function of time and then calculates the efficiency of the system at nominal conditions based on the extracted parameters (see Fig. 4b). Due to such ability, the method can deconvolve the percentage of the efficiency that is being lost because of the degradation of each of the parameters that are shown in Fig. 4c. The evolution of parameters as a function of time is shown in Fig. 4d, based on the deconvolution of parameters, the series resistance has the highest increase and causes most of the loss.

IV. Conclusions

DEEDS can be used for data preservation, sharing, and analysis for academic or industrial applications. Here we demonstrated a solar PV use case to show various features of the DEEDS online platform. However, the discipline-neutral nature of DEEDS makes it a powerful tool to be used for all the facets of PV research and assists collaborations.

Acknowledgment

This work is supported by the National Science Foundation under Grant No. #1724728.

References

[1] "DURAMAT, U.S. Department Of Energy." [Online]. Available: https://datahub.duramat.org/.

[2] A. C. Catlin, C. Hewanadungodage, and A. Bejarano, "Lifecycle Support for Scientific Investigations: Integrating Data, Computing, and Workflows," *Comput. Sci. Eng.*, vol. 21, no. 4, pp. 49–61, 2019.

[3] A. C. Catlin *et al.*, "A Cyberplatform for Sharing Scientific Research Data at DataCenterHub," *Comput. Sci. Eng.*, vol. 20, no. 3, pp. 49–70, 2018.

[4] C. HewaNadungodage *et al.*, "The DEEDS platform: Support for integrated data and computing across the research lifecycle," *Futur. Gener. Comput. Syst.*, Nov. 2019.

[5] X. Sun, R. V. K. Chavali, and M. A. Alam, "Real-time monitoring and diagnosis of photovoltaic system degradation only using maximum power point-the Suns-Vmp method," *Prog. Photovoltaics Res. Appl.*, vol. 27, no. 1, pp. 55–66, Jan. 2019.