

White Paper: A Data Repository and Search Engine for Big Data Research in Power Systems

Big Data Access Working Group

IEEE PES Subcommittee on Big Data and Analytics for Power Systems

Prepared by

Hamed Mohsenian-Rad, University of California, Riverside, USA

Ning Zhou, Binghamton University, USA

Payman Dehghanian, George Washington University, USA

August 2019

Contents

1. Introduction	3
2. Website	3
3. Data Sets	4
4. Search Engine.....	6
5. Contribute Data.....	10
6. Volunteer Activities.....	11

1. Introduction

The Big Data Access (BDA) working group was established in 2017 as part of the IEEE PES Subcommittee on Big Data and Analytics for Power Systems. The objective of the BDA working group is to facilitate public access to power systems data to promote big data research and development. The electric power industry, interacting with one of the largest customer-serving critical networks and going through drastic rapid changes in both business and technical paradigms, is presenting limitless opportunities for big data studies. However, research and development on big data are not possible unless one has access to big data. Accordingly, the BDA working group was formed in order to facilitate such access in an organized and sustainable fashion by identifying and addressing the related practical obstacles.

The BDA working group has recently launched its website at the following link:

<https://bigdata.seas.gwu.edu/>

As of August 2019, over 60 data sets are indexed at the above link, serving as the data repository for the BDA working group, with the ability to search across different types of power system data sets that can help quickly pinpoint adequate data sets to be used for data-driven research and education.

The purpose of this White Paper is to provide an overview of the features of the BDA working group's data repository and lay out the potential for expansion and extended efforts by working group's members.

2. Website

The front page of the BDA website is shown in Fig. 1.



Fig. 1. The front page of the website for the BDA Working Group.

Various information, including the news and activities of the working group, are available on the website. In this White Paper, our focus is to overview three aspects of the website:

- Data Sets
- Search Data
- Contribute Data

3. Data Sets

As of August 2019, over 60 data sets are made available on the website, covering various data sets, such as:

- Solar Radiation Data
- Wind Power Generation Data
- Distributed Energy Resources Data
- PMU Data
- Micro-PMU Data
- Smart Meter Data
- Power Quality Data
- Fault Detection Data
- Electric Vehicle Load Data
- Electric Vehicle Mobility Data
- Load Data - Appliances
- Load Data - Residential Building
- Load Data - Private Homes
- Load Data - Commercial Building
- Load Data - Office Building
- Energy Storage Data
- Independent System Operator Data
- Weather Data
- Electricity Market Data
- Electricity Pricing Data.

The data sets come from diverse geographical locations, including United States (California, New York, Texas, etc.), UK, Canada, Australia, Brazil, Belgium, Ireland, Switzerland, Germany, and Portugal.

The current list of data sets is shown in Fig. 2. Once you go to the tab “Data Sets” and then click on any of these data sets, you would see something similar to the page shown in Fig. 3. For the example in this figure, the data set is for distributed energy resources (DERs), covering 2 years of data at 1-hour intervals. The link to the actual data set is provided under “Data URL”. It should be noted that the BDA working group’s data repository is not intended to host the data, rather the data should be stored in an external location.

DATA SETS	
Data Set 1	Smart Meter Energy Demand Research Project – UK
Data Set 2	Solar Power Data for Integration Studies – USA
Data Set 3	NYSERDA Distributed Energy Resource (DER) Data Set – USA
	□ □ □
Data Set 60	Power Quality Data Set
Data Set 61	Summer PV Generation Data Set – Brazil
Data Set 62	Winter PV Generation Data Set – Brazil

Fig. 2. The list of current data sets as of August 2019.


NYSERDA DISTRIBUTED ENERGY RESOURCES (DER) DATA	
NYSERDA Distributed Energy Resources (DER) Data – USA	
Description	NYSERDA Distributed Energy Resources (DER) Data
Data Types	Distributed Energy Resource
Categories	DER Data
Format	CSV
Sampling Intervals	1-Hour
Starting Time (Year)	2016
Time Duration	2 Years
Total Size	N/A
Geographic Location	New York State, USA
Geographic Resolution	State-Wide (with Street Address)
Access Tools	Website: New York State Energy Research and Development Authority (NYSERDA)
Data URL	Data 
User Manual	N/A
Keywords	<ul style="list-style-type: none"> • Distributed Energy Resources • Anaerobic Digester • Combined Heat and Power • Energy Storage • Fuel Cell • Solar Photovoltaic • New York • NYSEDA

Fig. 3. A sample data set.

4. Search Engine

The most important feature of the website is the ability to search within the data sets. It helps students and researchers to quickly pinpoint data sets to be used for data-driven research and education.


The easiest way to search within the data sets is to type your intended keyword in the box for “Search Site”, as shown in Fig. 4 by the red arrow in this figure. Next, we go through several examples.

DATA SETS

Data Set 1

Smart Meter Energy Demand Research Project – UK

SEARCH SITE



QUICK LINKS

- Data Sets
- IEEE PES Subcommittee on Big Data & Analytics for Power Systems

Fig. 4. Search within the data sets by typing a keyword in the search box.

As an example, if you type “Smart Meter”, four data sets will be listed, as shown in Fig. 5.

The screenshot displays a search interface with the title "Search Results for: Smart Meter". On the left, four data sets are listed: "Smart Meter Data Set - Ireland", "Smart Meter Electricity Trial Data Set", "Texas Residential Smart Meter Data Set", and "Smart Meter Energy Demand Research Project - UK". On the right, there is a "SEARCH SITE" section with a search bar containing "Smart Meter", and a "QUICK LINKS" section with two links: "Data Sets" and "IEEE PES Subcommittee on Big Data & Analytics for Power Systems".

Fig. 5. The data sets that are listed when the search keyword is “Smart Meter”.

As another example, if you type “PMU”, four other data sets will be listed, as shown in Fig. 6.

The screenshot displays a search interface with the title "Search Results for: PMU". On the left, four data sets are listed: "LBNL Micro-PMU Data Set - USA", "EPFL Campus PMU Data Set", "Synthetic PMU Data Set", and "Distribution PMU Data Set". On the right, there is a "SEARCH SITE" section with a search bar containing "PMU", and a "QUICK LINKS" section with two links: "Data Sets" and "IEEE PES Subcommittee on Big Data & Analytics for Power Systems".

Fig. 6. The data sets that are listed when the search keyword is “PMU”.

Another example is to search for “Wind”, which returns five data sets, see Fig. 7.

The screenshot displays a search interface with the title "Search Results for: Wind". On the left, five data sets are listed: "Wind Generation Data Set", "MesoWest Weather Data Set for Wind and Solar Integration", "Wind Speed Data Set - USA", "Wind Power Generation Data Set - Belgium", and "Smart Meter Data Set - Ireland". On the right, there is a "SEARCH SITE" section with a search bar containing "Wind", and a "QUICK LINKS" section with two links: "Data Sets" and "IEEE PES Subcommittee on Big Data & Analytics for Power Systems".

Fig. 7. The data sets that are listed when the search keyword is “Wind”.

Similarly, you may search for “Solar”, which returns 10 different data sets, see Fig. 8.

The screenshot displays a search interface with the title "Search Results for: Solar". On the right, a "SEARCH SITE" box contains the text "Solar". Below it, a "QUICK LINKS" box lists two items: "Data Sets" and "IEEE PES Subcommittee on Big Data & Analytics for Power Systems". The main content area on the left lists ten data sets, each on a new line separated by a horizontal line:

- MesoWest Weather Data Set for Wind and Solar Integration
- Solar Generation Data Set – Worldwide
- Solar Radiation Data Set – USA
- Rooftop Solar Data Set – Australia
- DKASC Solar Power Data Set – Australia
- Solar Radiation Data Set
- Belgium Solar PV Power Generation Data Set – Belgium
- NREL Solar Radiation Data Set
- High-Resolution Solar Radiation Data Set
- Solar Power Data for Integration Studies – USA

Fig. 8. The data sets that are listed when the search keyword is “Solar”.

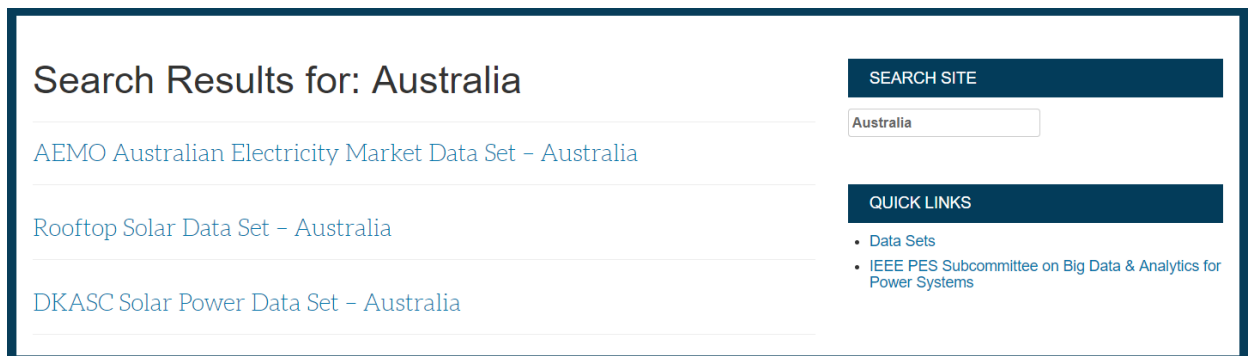
You can narrow down your search by adding “USA” to indicate the geographical location for the data. This will reduce the returned data sets to four, as listed in Fig. 9.

The screenshot displays a search interface with the title "Search Results for: Solar and USA". On the right, a "SEARCH SITE" box contains the text "Solar and USA". Below it, a "QUICK LINKS" box lists two items: "Data Sets" and "IEEE PES Subcommittee on Big Data & Analytics for Power Systems". The main content area on the left lists four data sets, each on a new line separated by a horizontal line:

- MesoWest Weather Data Set for Wind and Solar Integration
- Solar Radiation Data Set – USA
- Solar Radiation Data Set
- NYSERDA Distributed Energy Resources (DER) Data

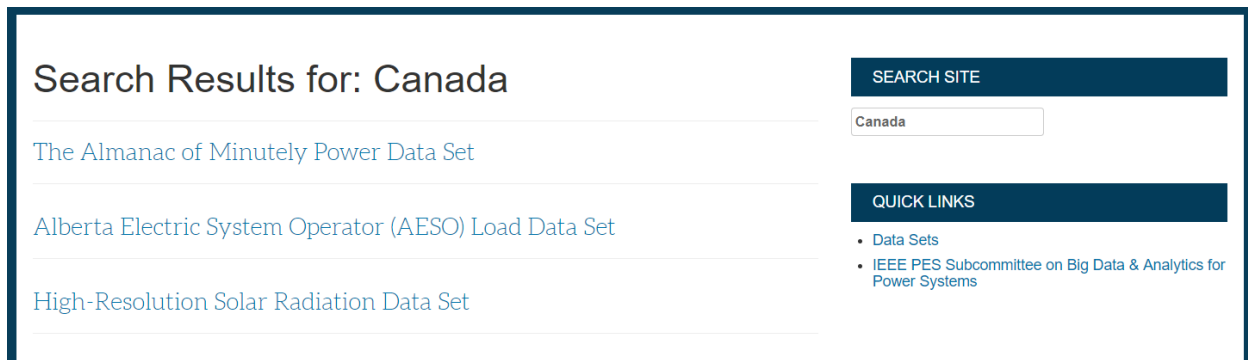
Fig. 9. The data sets that are listed when the search keyword is “Solar” and “USA”.

If the goal is to find data sets of different kinds for a particular country or geographical location, then the keyword to search can be set accordingly. Two examples are shown in Figs. 10 and 11.



The screenshot shows a search interface with the title "Search Results for: Australia". On the left, there is a list of three data sets: "AEMO Australian Electricity Market Data Set - Australia", "Rooftop Solar Data Set - Australia", and "DKASC Solar Power Data Set - Australia". On the right, there is a "SEARCH SITE" section with a text input field containing "Australia". Below that is a "QUICK LINKS" section with two bullet points: "Data Sets" and "IEEE PES Subcommittee on Big Data & Analytics for Power Systems".

Fig. 10. The data sets that are listed when the search keyword is “Australia”.

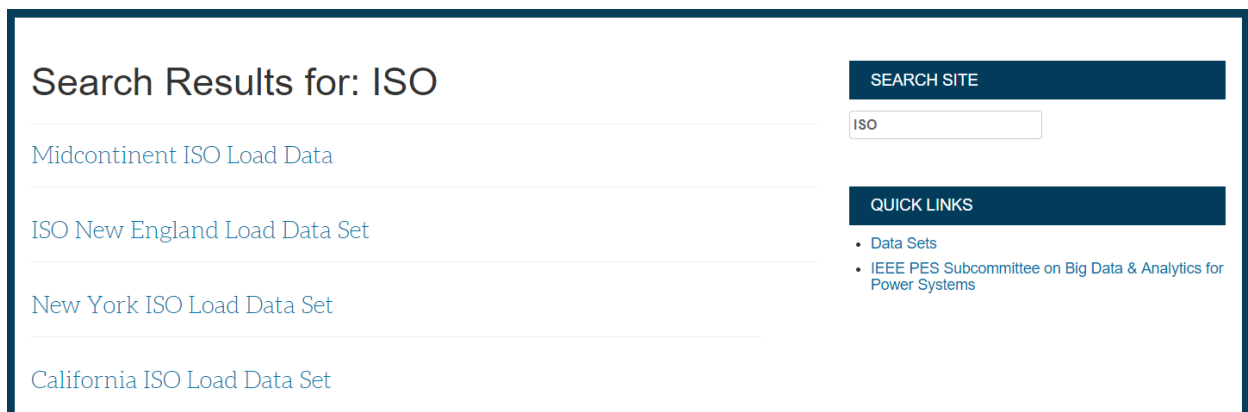


The screenshot shows a search interface with the title "Search Results for: Canada". On the left, there is a list of three data sets: "The Almanac of Minutely Power Data Set", "Alberta Electric System Operator (AESO) Load Data Set", and "High-Resolution Solar Radiation Data Set". On the right, there is a "SEARCH SITE" section with a text input field containing "Canada". Below that is a "QUICK LINKS" section with two bullet points: "Data Sets" and "IEEE PES Subcommittee on Big Data & Analytics for Power Systems".

Fig. 11. The data sets that are listed when the search keyword is “Canada”.

The search keyword could be in the title, as in the case for Australia, or somewhere else in the description of the dataset, as in the case for Canada.

If you search for ISOs, four data sets will be returned from four different ISOs in the USA, see Fig. 12.



The screenshot shows a search interface with the title "Search Results for: ISO". On the left, there is a list of four data sets: "Midcontinent ISO Load Data", "ISO New England Load Data Set", "New York ISO Load Data Set", and "California ISO Load Data Set". On the right, there is a "SEARCH SITE" section with a text input field containing "ISO". Below that is a "QUICK LINKS" section with two bullet points: "Data Sets" and "IEEE PES Subcommittee on Big Data & Analytics for Power Systems".

Fig. 12. The data sets that are listed when the search keyword is “ISO”.

5. Contribute Data

The form to complete in order to contribute a data set to the data repository is shown in Fig. 13.

CONTRIBUTE DATA

ABOUT THE CONTRIBUTOR

First Name (required)

Last Name (required)

Affiliation (required)

Email (required) *

Address (required)

City State/Province Zip/Postal
Country

Phone Number (required)

ABOUT THE DATA

Data Title

Data Types

Data Categories

Data Format

Data Sampling Intervals

Data Starting Time (Year)

Data Time Duration

Data Total Size (GB)

Geographic Location

Geographic Resolution

Data Access Tools

Data Owner

Data URL

Data User Manual

Data Keywords

Data References

Data Discussion Forum

Fig. 13. The form for contributing a data set.

The research community and public are welcome to contribute data sets. You do not need to own the data to share it. In fact, while this working group is interested in working with utilities and other entities

to help them share their data, the primary focus in this working group is to share the data that is *already available to the public*. However, the goal here is to bring various data in one place and make it easy for the research community to find the data that they need through simple search features, as listed in Section 4.

6. Volunteer Activities

The BDA working group encourages membership and volunteer activities. To be a member, please send an email to one of the working group's leadership team as listed on the first page of this White Paper.

Volunteer activities may include but not limited to the following:

- **To Contribute Data Set:** The research community and public are welcome to contribute data sets, in particular, sharing data that is already available to the public, see Section 5 for details.
- **To Maintain Data Sets:** As more data sets are added to the working group's data repository, it becomes increasingly important to actively maintain the data sets. For example, since data sets are contributed by different active volunteers, it is likely that over time some data sets are repeated; therefore, there is a need to clean up the data sets and reconcile the repeated data sets. Also, the external links for some data sets may become obsolete over time. Such data sets will have to be removed or the external links must be updated.
- **To Add References and Contribute to Discussion Forum:** Equally important to contributing and maintaining data sets is to gradually provide references to research papers and other publications that have previously used such data. The research committee members are also encouraged to contribute to the discussion forum to share their experience working with a data set.
- **To Expand Website:** The working group also encourages members who can actively help with maintaining and expanding the website and its features, review submitted data sets, etc.