U.S. DEPARTMENT OF ENERGY | Office of ENERGY EFFICIENCY & RENEWABLE ENERGY

Open Energy Data Initiative

Advancing Analytics and Research Innovation through Improved Data Access

The Open Energy Data Initiative (OEDI) aims to improve and automate access of high-value energy data sets across the U.S. Department of Energy's (DOE's) programs, offices, and national laboratories. Sponsored by DOE, this platform is being implemented by the National Renewable Energy Laboratory (NREL) to make data actionable and discoverable by researchers and industry to accelerate analysis and advance innovation.

Partners at other national laboratories will help utilize the platform for analysis, can be directly involved in determining requirements, and will be resources for providing new datasets that can be shared with the public to expand innovation.

What is OEDI?

OEDI is geared toward developing improved open data access to better enable analysis by DOE, national laboratories, industry, and the general public. A robust analysis platform will improve accessibility through a data lake, data catalog, and high-value open datasets hosted on multiple cloud computing platforms.

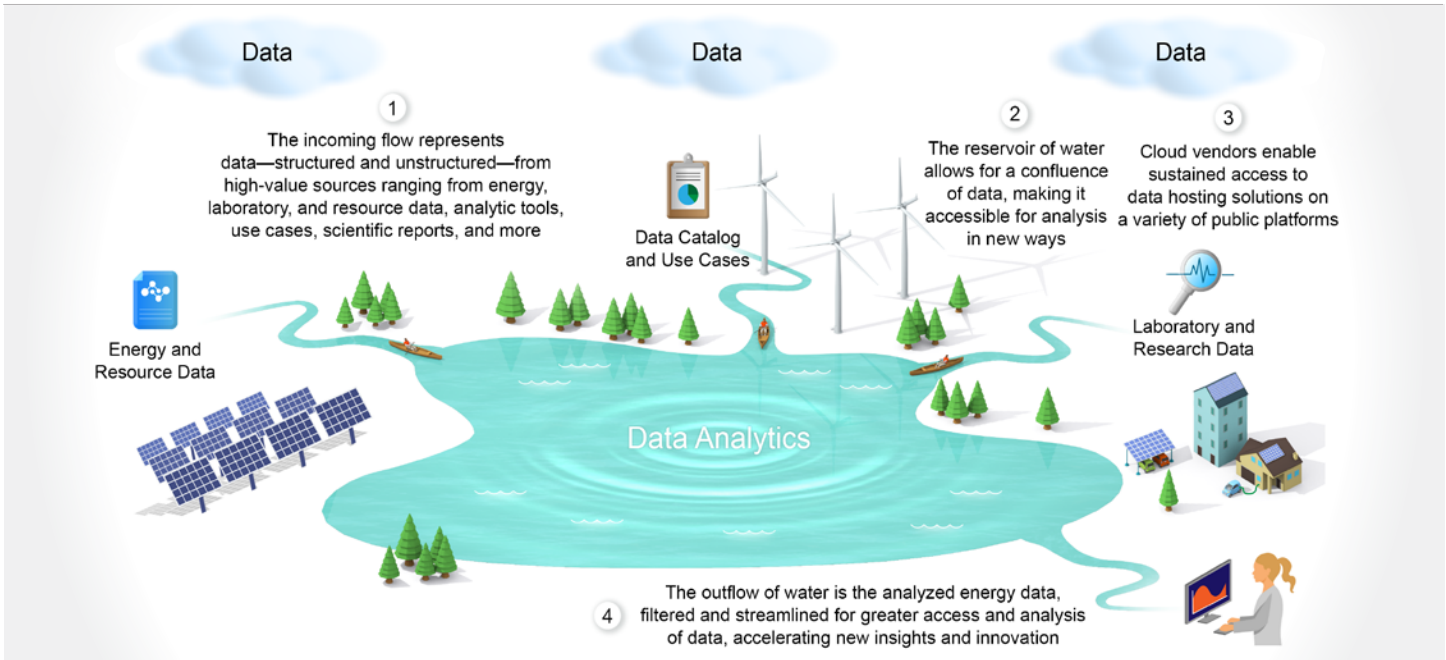
Cloud venders will have the opportunity to host high-value data sets that are in demand by data scientists and analysts, entrepreneurs, and industry experts. The data catalog can be utilized to find and access datasets through a variety of cloud hosting vendor sites. The data lake can be used to conduct interesting data mash-ups, deep analytics, and advanced computations for developing new and expanded datasets. The data lake will also enable key analytics tools to simplify processes while also advancing the potential for rapid, complex analysis.

The OEDI project will:

- Make relevant, useful connections between analysts and high-value open data
- Increase the accessibility and visibility of high-value data sets, distilled into relevant, actionable information
- Engage key stakeholders to collaborate and determine new data sets and use cases to assist in data analysis
- Allow construction of new data sets from numerous other varied data to expand the benefits of the data
- Provide a data lake to enable faster, easier, more advanced analysis and computation to accelerate novel data mash-ups, analysis, and innovation
- Develop data and computing relationships with cloud computing providers, national labs, and federal agencies
- Support next-generation research and analysis by allowing researchers to analyze big data without making expensive hardware investments (i.e. supercomputer).

Why OEDI?

Many researchers have difficulty accessing and utilizing big data and complex data because: 1) data can be difficult to find and use, 2) there is a lack of publicly available data, 3) it's often not in standard sizes or formats, and 4) large data sets can be costly to store and manage.



OEDI Data Lake to Source Information from DOE and 17 National Laboratories. *Illustration by Besiki Kazaishvili, NREL*

OEDI will help remove these barriers and improve accessibility for analysts and researchers. The platform will enable data scientists and analysts to explore, mash-up, and analyze data in a framework that speeds innovation, allowing for rapid computation while also utilizing portions of their manipulated data for other purposes. The platform will catalog the data sets and allow users to add public data, too.

Power of Partnerships

- **Cloud Vendors:** The OEDI project will work closely with various cloud hosting partners to ensure each data set can be publicly hosted and shared on a variety of platforms to eliminate the potential for vendor lock-in. In some cases, high-value data sets will be hosted for free; in other cases, the federal government will pay for the hosting; but in all cases, OEDI data will be open data, available to all who should want it. With these partners, DOE and NREL will increase the openness and usefulness of the data by expanding the endpoints for data access.
- **National Laboratories:** The data and information coming out of our national laboratories are valuable assets. This initiative gives laboratory staff a platform through which they can easily share datasets with the public. Access to big data analytics and computation in the cloud will enable innovation and data mash-ups like never before. Using key analysis tools only found in the cloud, researchers can analyze data faster and industry can leverage large-scale federal datasets in new ways.

How it Works?

- **The Data Lake:** OEDI will build a robust data lake on Amazon Web Services (AWS) that will enable analytical functionality and comprehensive data analysis. The data lake will leverage AWS analysis tools that will assist users with consolidating data in non-standardized formats, accelerate analytics, and allow users to download or move small portions of their analysis into their own AWS accounts. To assist scientists and analysts with understanding some of the data analytic possibilities, sample code snippets (Python Notebooks, SQL Statements) will be provided to demonstrate access and query capabilities.
- **The Data Catalog:** In order to easily find available data, OEDI is creating a data catalog that merges data from multiple cloud hosting locations. The catalog will include information about the data and data owner, metadata, licensing information, provenance, and links to all relevant data locations (if the data is housed on multiple cloud hosting vendor sites). The catalog will be designed to get users to the data as quickly as possible.
- **Partnership Agreements:** Based on a previous, successful federal partnership with a variety of cloud hosting vendors, OEDI strives to build similar agreements with all major cloud hosting vendors. These partnerships are important for building long-term sustained data hosting solutions that will enable big data to be openly shared with the public in a cost-effective manner. Our goal is to ensure high-value datasets are openly accessible via a variety of cloud hosting services, free of charge. Once these data are made available, industry can access and analyze the data on the platforms meet their needs.

What Kind of Data Will Be Included?

OEDI will start with two large datasets—examples of high-impact, high-demand resource data. The WIND Toolkit data includes meteorological conditions and turbine power for more than 126,000 sites in the continental United States for the years 2007–2013. The National Solar Radiation Database (NSRDB) is a serially complete collection of hourly and half-hourly values of the three most common measurements of solar radiation—global horizontal, direct normal, and diffuse horizontal irradiance—and meteorological data.

Future datasets will be dependent on laboratory researchers and other stakeholders. Criteria for OEDI datasets include usefulness to a broad audience and researchers who will support the inclusion of their data. Initial plans will focus on additional resource and technology data (such as solar, wind, geothermal, water, bioenergy, and transportation), utility rates, technology performance data, materials and market data, information from cities and states, and data related to grid systems.

How to Participate

- Are you a **Data Owner** who wants to **share** data? Contact us to let us know more about your data and who may be interested in your data.
- Are you a **Data User** who wants to **analyze and utilize** data? Contact us to learn more about the project, get access to pilot versions, and help guide functional requirements for the data platform.

Contacts

U.S. Department of Energy – Dave Rench McCauley
Dave.Rench-McCauley@ee.doe.gov

National Renewable Energy Laboratory – Debbie Brodt-Giles
Debbie.Brodt.Giles@nrel.gov

For more information, visit: https://openei.org/oedi

DOE/GO-102019-5171 · July 2019

U.S. DEPARTMENT OF ENERGY | Office of ENERGY EFFICIENCY & RENEWABLE ENERGY