DISCLAIMER

Open Energy Data Initiative (OEDI) FY22-24: Final Technical Report

Jon Weers

National Renewable Energy Laboratory

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated under Contract No. DE-AC36-08GO28308

This report is available at no cost from
NREL at www.nrel.gov/publications.

Technical Report
NREL/TP-6A20-94426
September 2025

Open Energy Data Initiative (OEDI) FY22-24: Final Technical Report

Jon Weers

National Renewable Energy Laboratory

Suggested Citation
Weers, Jon. 2025. *Open Energy Data Initiative (OEDI) FY22-24: Final Technical Report*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-6A20-94426. https://www.nrel.gov/docs/fy25osti/94426.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated under Contract No. DE-AC36-08GO28308

This report is available at no cost from NREL at www.nrel.gov/publications.

Technical Report
NREL/TP-6A20-94426
September 2025

15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored by NREL for the U.S. Department of Energy (DOE), operated under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

Cover photos (clockwise from left): Josh Bauer, NREL 61725; Visualization from the NREL Insight Center; Getty-181828180; Agata Bogucka, NREL 91683; Dennis Schroeder, NREL 51331; Werner Slocum, NREL 67842.

Final Technical Report (FTR)

Cover Page

a. Federal Agency	Department of Energy	
b. Award Number	DE-EE0038443	
c. Project Title	DOE Open Energy Data Initiative	
d. Recipient Organization	National Renewable Energy Laboratory	
e. Project Period	*Start:* 10/01/2021	*End:* 9/30/2024
f. Principal Investigator (PI)	Jon Weers Data Science and Technology Lead, NREL Jon.Weers@nrel.gov 303-275-3701	
g. Business Contact (BC)	Jeff Cook Acting Program Lead Jeff.Cook@nrel.gov 303-275-3676	
h. Certifying Official (if different from the PI or BC)		

Signature of Certifying Official **Date**

1. **Acknowledgement:** This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the **Program Award Number DE-EE0038443**.

2.

3. **Project Summary:**
   The DOE Open Energy Data Initiative (OEDI) is a partnership between the National Renewable Energy Laboratory (NREL), the U.S. Department of Energy (DOE), and major cloud providers; Amazon, Microsoft, and Google to provide universal access to big data on cloud platforms. The OEDI Data Repository is a centralized online hub of high-value energy research open datasets, which are aggregated from the U.S. Department of Energy's Program Offices, National Laboratories and other collaborators. OEDI aggregates smaller, domain-specific data repositories, allows direct data submissions, and includes support for big data through its energy data lakes hosted on the cloud. The OEDI team developed cloud-hosted data lakes making high-value data universally accessible. This new capability helps researchers, entrepreneurs and the public access and readily use petabytes of open energy datasets.

4. **Project Objectives and Outcomes:** The primary objective of this project is to enable open science through enhancements to the OEDI data catalog and the development of data pipelines and an advanced analytics engine to facilitate easier access to cloud-hosted data lakes, incorporate usage analytics, and provide users with helpful guides on how to use big data and perform analyses in the cloud.

   Outcomes:

   4.1 Increased Availability of Open Data

   Expanding the network of data sharing partners, increasing exposure and dissemination, while also increasing the volume of cloud-accessible data.

   Throughout the course of this project, OEDI's public data footprint has increased by 8,252%, up from 32 terabytes (TB) in 2018 to over 2.61 petabytes (PB) in 2024 (Figure 1). High-value data from DOE's Solar, Wind, Geothermal, Buildings, and

Marine energy programs are now available in OEDI data lakes. By the end of Fiscal Year (FY) 2024, OEDI datasets contained more than 10,867 distinct data resources and were downloaded more than 14 million times (Figure 1). Monthly and quarterly impact reports showing these stats were distributed to DOE and other OEDI stakeholders on the 1st day following the reporting period.

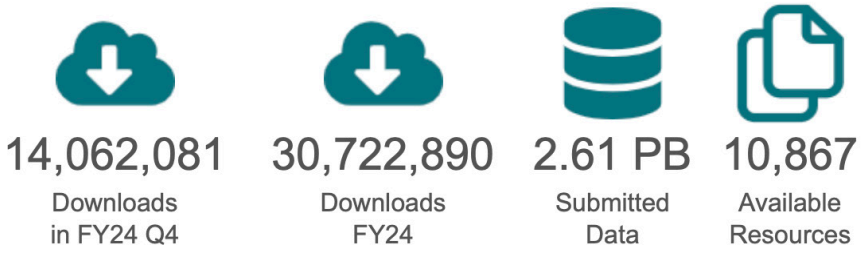

Figure 1 Image from the OEDI Impact Report for the fourth quarter of Fiscal Year 24 (Q4)

4.2 Improved Discoverability through the Expansion of OEDI Data Sharing Network
OEDI metadata is automatically shared across a network of other sites (Figure 2), allowing OEDI dataset to appear in the data catalogs of those sites. Only the metadata is shared, along with pointers to the original datasets on OEDI, to avoid duplication of effort and additional data storage costs. Users of any of these sites can seamlessly access OEDI data without being redirected to another site, allowing OEDI datasets to be discovered by users of dozens of sites across the web.

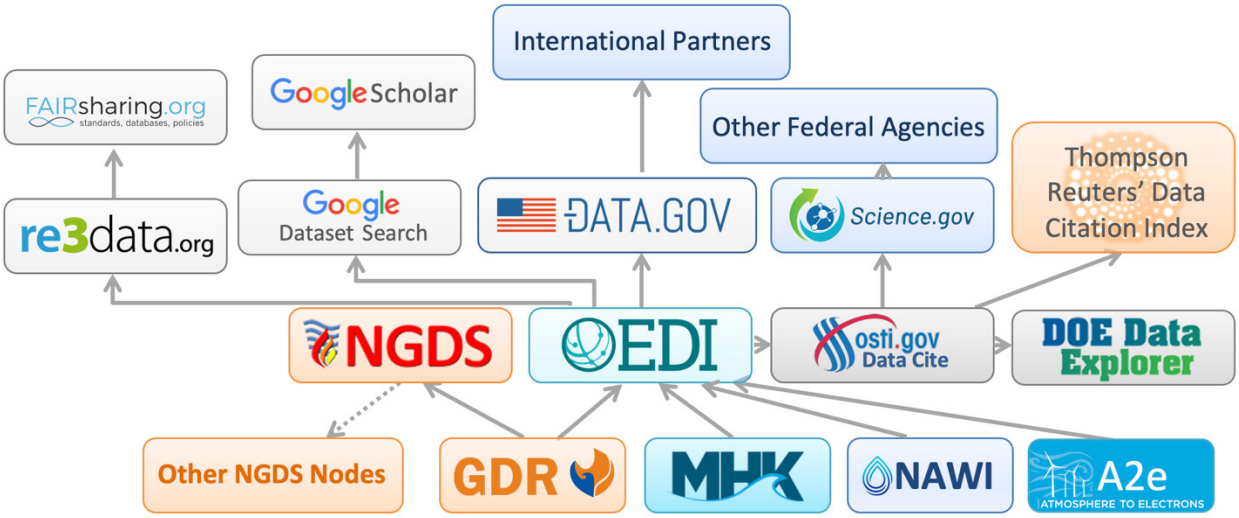

Figure 2 Network diagram of the OEDI Data Sharing Network.

Improvements made to expand this network of data sharing partners, including integration with Google Dataset Search and r3data.org helped OEDI achieve a nearly exponential increase in impact year-over-year (Figure 3).

By the end of FY24, data on OEDI had been downloaded more than 27 million times by a broad spectrum of users, including researchers, analysts, engineers, students, industry professionals, entrepreneurs and policy makers.

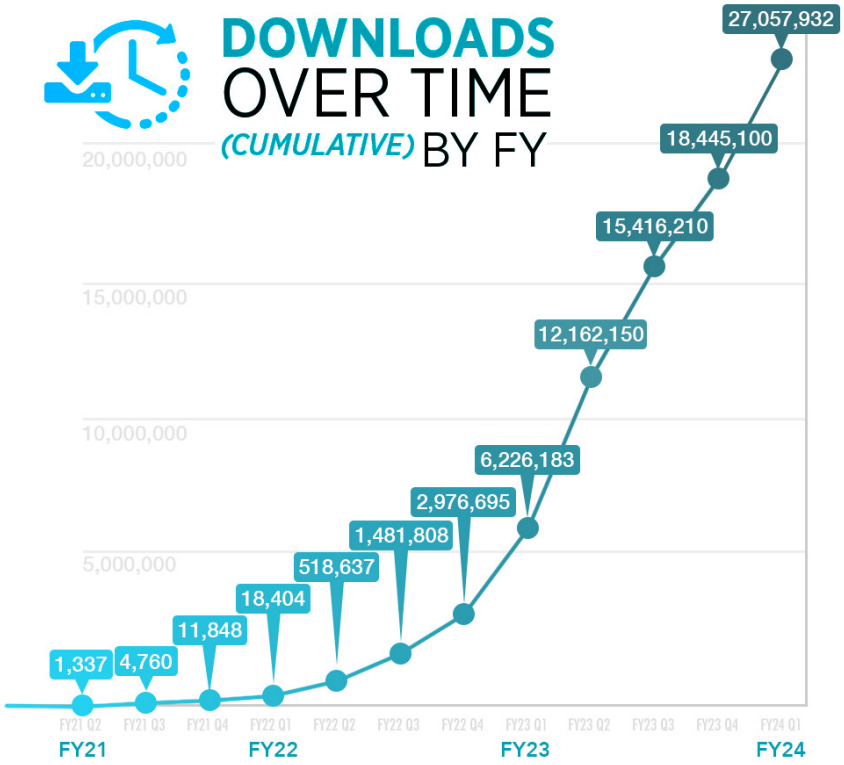

Figure 3 OEDI Cumulative data downloads over time, from FY21 to FY24.

4.3 Development of an Advanced Analytics Engine

Tracking these downloads was made possible through NREL's development of an advanced analytics engine, which tracks downloads through OEDI's network of data sharing partners. Typical web analytics platforms, such as Google Analytics, track events through web browsers and link the events to the website being visited. However, since OEDI data is available on multiple sites across its network of data sharing partners, another means of measuring impact had to be developed. Additionally, OEDI enables universal access to data through a variety of means that do not involve a web browser, including programmatic access, command line tools, programming languages, and machine learning platforms. OEDI's advanced analytics engine tracks data access requests at the server level from multiple end points to capture download requests federated through the network of data sharing partners as well as direct access requests through programming languages and native cloud tools. These capture points then feed analytics data back to OEDI's centralized analytics engine which processes them into meaningful impact metrics.

4.4 Expansion of OEDI Data Lakes to Multiple Cloud Providers

One of the goals of this project was for OEDI to expand access to its data lakes from AWS to other cloud platforms, such as Microsoft Azure and Google Cloud Platform

(GCP), to be cloud agnostic and to have a uniform presence on all three cloud providers. At its height, OEDI had more than 80% of its data on 2 or more cloud providers. However, changes in the strategic direction of Microsoft's public data programs and the operational pause of the Google public data program forced the removal of some OEDI data lakes and prevented new data lakes from being added to these cloud providers. As several more petabytes of data were added to OEDI data lakes on AWS, the amount of data hosted on 2 or more cloud providers diluted to 21% at the end of FY24. As a result, the OEDI team shifted project resources from prioritizing a cloud agnostic approach to one prioritizing data quality and interoperability with cloud-native tools for big data querying, machine learning, and artificial intelligence.

4.5 Curation of New Data

Throughout the course of the project, new data submissions were curated by the OEDI data curation team. The metadata associated with each data submission were reviewed for:

- **Relevance** to ensure the metadata is descriptive of the data submitted
- **Appropriateness** to make sure they are suitable for eventual public release and do not inadvertently contain personally identifiable information (PII) or other obviously inappropriate content
- **Completeness** to ensure no critical components are missing

More information can be found on OEDI's About page: https://data.openei.org/about.

4.6 Development of AskOEDI: OEDI's Artificially Intelligent (AI) Research Assistant

In FY24, NREL developed and launched AskOEDI, an Artificially Intelligent (AI) research assistant that leveraged work done previously to made OEDI metadata machine-readable and a Large Language Model (LLM) integrated with metadata and supporting documents from OEDI datasets. AskOEDI improves the discoverability of DOE data by helping guide users to datasets beyond simple keyword searches. It enables users to find answers to questions about the underlying assumption or methodologies used in a dataset, discover information contained within supporting documents, and explore data from projects related to their research objectives.

AskOEDI is a Retrieval-Augmented Generation (RAG) application. It enhances responses from ChatGPT's LLM by incorporating OEDI metadata and supporting documents. This allows AskOEDI to generate more relevant and accurate responses based on OEDI's curated datasets. Figure 4 diagrams the data flow from OEDI metadata through a vectorized database to the response generated by ChatGPT.

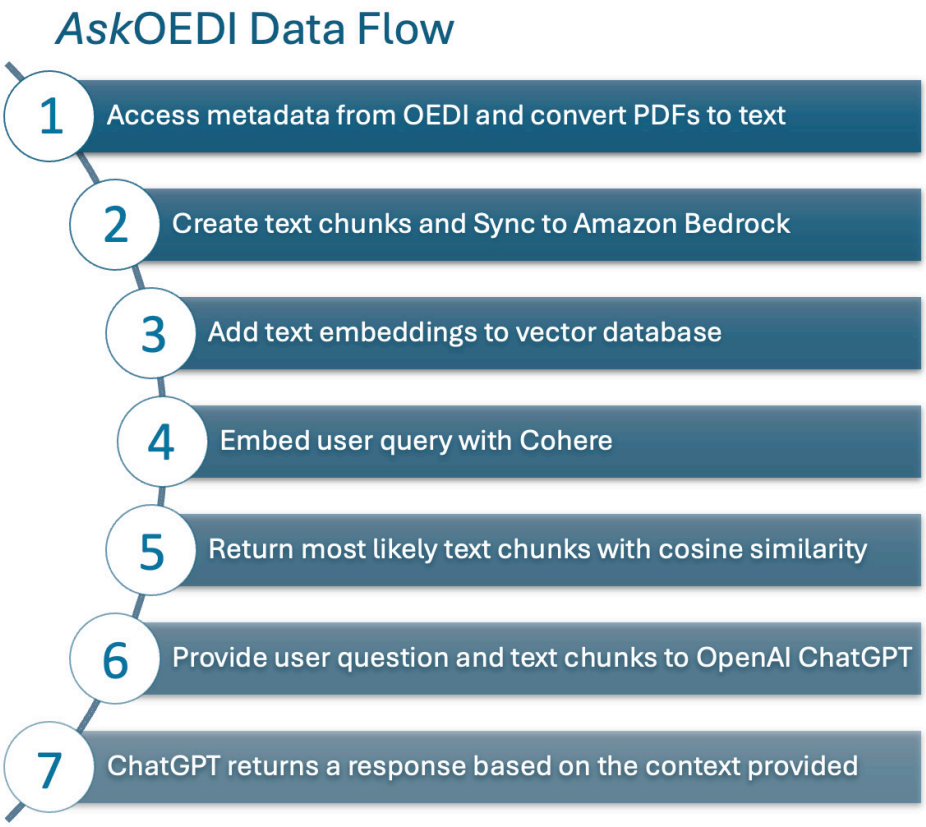

Figure 4 Flow of information from OEDI metadata to AskOEDI response

4.7 Key Milestones Completed During this Project:

- Develop universally accessible data lakes
- Uniform presence on each of the 3 cloud providers' public dataset programs
- Develop an Advanced Analytics Engine to track data lake usage and impact
- Publish cloud costs case studies
- Demonstrate integration of OEDI data with native cloud query and analysis tools (Athena, BigQuery)
- Integrate OEDI data with AI/ML (AskOEDI)
- SolarTRACE Technical Report

5. **Path Forward:** NREL will continue to maintain the core OEDI platform to ensure the preservation and continued availability of high-impact DOE datasets by:

a. Maintaining the OEDI catalog and data lakes cloud architecture, including the application of DOE-mandated cyber security patches.
b. Assisting DOE-funded projects with updates to existing OEDI datasets and data lakes (e.g. annual updates to the NSRDB or additional data channels for the WIND Toolkit), with a primary focus on updating AWS data lakes.
c. Maintaining consistency of the data stored in data lakes across the three cloud providers when possible and economical. If necessary, remove outdated or incomplete data from under-supported open data programs on Google or Azure.

d. Supporting DOE-funded projects to effectively use native cloud ecosystems to leverage OEDI data lakes to promote collaborative data access and analysis, including data access, management, and dissemination best practices.
e. Maintaining connections to data sharing partners, including Data.gov and DOE's Office of Science and Technical Information (OSTI) to ensure the discoverability of DOE data.
f. Providing technical assistance to OEDI users and assist with data collaboration needs.
g. Continuing to curate incoming data sets and resolve broken links in existing datasets to ensure the quality of OEDI data and metadata.

Additionally, NREL will monitor the use of AskOEDI and collect metrics on its impact, including usage, operating costs, questions asked, and the quality of provided answers. An analysis of these metrics will be used to modify AskOEDI to increase its impact and the quality of generated answers.

6. Inventions, Patents, Publications, and Other Results:

6.1 Publications produced by this project:

SolarTRACE Technical Report (Pending DOI):

Cruce, Jesse, Lara, Dalila, Frey, Noah, Dalecki, Emily and Cook, Jeff. *Residential and Small Commercial Solar Photovoltaic and Storage Permitting, Inspection, and Interconnection Timelines—A Retrospective Review (2017–2023).* Golden, CO: National Renewable Energy Laboratory. NREL/TP-6A20-92539.

OEDI was also featured in the following publications:

Weers, Jon, Anderson, Arlene, and Taverna, Nicole. "Connecting People to Data: Enabling Data Connected Communities through Enhancements to the Geothermal Data Repository." GRC Transactions, Vol. 47, 2023. (OSTI ID: 2041087)

Weers, Jon, Porse, Sean, Huggins, Jay, Rossol, Michael, and Taverna, Nicole. “Improving the Accessibility and Usability of Geothermal Information with Data Lakes and Data Pipelines on the Geothermal Data Repository.” GRC Transactions, Vol. 45, 2021. (OSTI ID: 1827897)

News stories:

Daugherty, Justin. " Dive Into a Lake of Data: Open Energy Data Initiative Increases Big Data Access for Everyone." *NREL News*, 23 Oct. 2024, https://www.nrel.gov/news/program/2024/dive-into-a-lake-of-data-open-energy-data-initiative-increases-big-data-access-for-everyone.html.

6.2 Other Results:

- **The Open Energy Data Initiative (OEDI) web site**: https://data.openei.org/
  Home to the OEDI Data Catalog and access portal to all 3,235 datasets on OEDI, including the over 2.6 petabytes of data in OEDI's 53 data lakes. The site includes a comprehensive keyword search, a frequently asked questions (FAQ) page, data access instructions, and AskOEDI, an AI Research Assistant capable of answering in depth questions about OEDI datasets.
- **The OEDI Public GitHub repository**: https://github.com/openEDI
  A public code repository containing public dataset documentation, a machine-readable version of OEDI's open data registry, and a library of open source data access tools for accessing and processing data in cloud-based data lakes.

7. Project Team:

Principal Investigator: Jon Weers
Chief Architect: Jon Weers
Chief Cloud Engineer: Jay Huggins
Data Curation Lead: Nicole Taverna
AI/ML Experts: Grant Buster and Slater Podgorny
UI/UX Designers: Sandy Lee and Tom King
Cloud Engineers: David Rager, Jianli Gu, Michael Bartlett, Swapnil Bhatkar, Damon Hayley, David Whiteside, Reid Olsen and Matt Heine
Data Hackathon Coordinators: Jon Weers and Danny Zimny-Schmitt
Data Curators: Scott Mello, Adrienne Lowney, Nathan Danigelis and RJ Scavo
Developers: Jon Weers, Sean Morris, Erika Curry-Elrod, Hanna Fields, Tom King and Jay Huggins
Project Coordinators: Eva Brod, Rachel Eilert, Hanna Watts, and Lisa Temple