Power Sector Transmission & Distribution Data and Information Webinar Series and Summit Report

November – December 2023

Prepared by:
U.S. DEPARTMENT OF ENERGY,
OFFICE OF ELECTRICITY

Contents

I. Executive Summary

The power sector is adapting to accommodate a shifting generation mix, the electrification of other sectors, and the changing role of customers on the grid. Data are critical to understanding and adapting to these new system conditions and maintaining reliability, resilience, security, and affordability. The Department of Energy (DOE) Office of Electricity (OE) conducts research, development, and demonstration of technologies that advance power sector data collection, processing, and management. Transformative technologies for power sector data are creating new methods of operating the grid, enabling new markets, and facilitating the integration of renewable resources.

OE sponsored a four-part webinar series, culminating in an in-person summit, *Power Sector Transmission & Distribution Data and Information*, in October and November 2023. These meetings provided space for experts to share research and development on ideas and technologies that could modernize the U.S. power sector.

Each of the four webinars focused on a topic related to power sector data and information:

- **Topic 1: Transmission & Distribution (T&D) Information Sharing**: This webinar covered current data-sharing practices between transmission, distribution, and microgrid operators and identified the need for standardization to increase reliability in the future. Contributors recognized the need for multilateral data sharing among all stakeholders in the sector, allowing entities to share and collect necessary data while respecting differing privacy and security standards. Contributors also discussed the current data standards for utilities and distributed energy resources (DERs), distribution forecasts, and load shedding. They identified the need to improve interoperability through standardized data exchanges (including the T&D interface) and shared ontologies while using regulation to manage who has access to such data and when.
- **Topic 2: Cross-Sector and Open Data Sharing and Risks**: As grid operations evolve and become more complex, there is a growing need for data sharing between utilities and other stakeholders. This webinar identified challenges that are currently limiting data sharing and discussed the need to reduce operational and economic risks to utilities while creating opportunities for innovative and creative ways to access, manage, and share grid data. The webinar also highlighted several DOE-funded projects, which have supported these objectives, as examples of potential paths forward.
- **Topic 3: Sensor Systems and Platforms**: This webinar surveyed state-of-the-art methods for analyzing already available grid data and data that could be central to future grid operations. For example, the webinar showcased sub-synchronous oscillations, which are difficult to simulate but will be more frequent with the transition to inverter-based generation and load. Initiating the development of analytics to model such conditions is critical but challenging, given the large degree of uncertainty.
- **Topic 4: Sensor Data and Device Research**: This webinar explored data needs for a distributed grid with a focus on distribution and grid-edge systems. DERs and electric vehicles (EVs) are complicating demand-side operations and driving the need for higher-fidelity data. Additionally, time-synchronized measurements are increasingly becoming critical for coupling interconnected systems. A new breed of distributed sensor platform combined with standards-based interoperability can drive new insights from the fusion of electrical and non-electrical based data streams. To unlock the potential of these advanced sensors and measurements, while preserving the integrity and privacy of power sector data, it is critical to develop scalable innovations.

During the in-person summit, contributors discussed existing data practices, in both domestic and international power markets, and pathways for developing a data strategy to support the evolving U.S. power sector. Discussions included opportunities and risks related to technology, policy, markets, standards, and privacy and key considerations included data needs, interoperability, open and closed data sharing, and privacy and security.

Key takeaways from the discussion included:

- **Accurate, accessible, and timely data are critical for making decisions that maintain the reliability, resilience, affordability, and security of the grid.** The power sector requires data to inform short- and long-term decisions for operations and planning. Conventional data management methods must be updated and reimagined to meet the needs of a rapidly changing power sector.
- **Entire industries are emerging to extract value from data and power sector data has enormous potential**. Data are cheaper and easier to transport than ever before. Power sector data could improve decision-making, connect stakeholders, and support the creation of new businesses, services, and value streams. Because power sector data can be sensitive, an effective data strategy will be needed to balance value and risk.
- **Further research is needed to identify, develop, and characterize strategies for sharing and protecting power sector data**. Derivative data products that aggregate, anonymize, and otherwise obfuscate data are critical to facilitate shared data insights while respecting data producer privacy. Additionally, these methods need to be flexible to meet the needs of a diverse set of users and scenarios.
- **It will be critical to have standards and governance to ensure that power sector data are valid, transparent, and high quality.** To maximize stakeholder participation, entities must trust that data are being responsibly sourced and adequately vetted. Key ingredients for building this trust will include accuracy standards, testing or audits for compliance, and metadata in data streams to indicate validity.
- **Recent power sector data are needed to better understand changing power sector conditions.** Electrification, the retirement of conventional resources, and renewable integration have contributed to new reliability events that models are unable to accurately represent. For instance, Scotland's 8 Hz voltage oscillation event in 2021 could not be duplicated in model simulations. The historical data sets that are currently available are mostly insufficient for developing credible scenarios and models to represent emerging changes in power systems. To characterize and understand the risks and opportunities in the evolving grid, researchers and utilities need to build sustained partnerships to facilitate the study and sharing of recent, real-world data sets.
- **Improved synthetic data could support power system tool development**. Real power sector data can be challenging for researchers and industry developers to obtain from utilities and are costly to format for research. Because of this challenge, the initial development of power system tools often relies on data created from models of the power system (also called simulated data). Research and industry should partner to improve models and data sets used for mimicking power system scenarios and technologies. Investments in these improvements must consider changing conditions that are due to electrification and decarbonization. Additionally, it may become possible to harness artificial intelligence (AI) for this end, if the problem of training machine learning (ML) models using historical data can be solved.

- **A common strategy and roadmap are needed to accelerate industry innovation**. A shared framework would empower utilities and vendors to invest in and create new data products, adopt common data standards, and invest in data-sharing technologies.

II. Introduction

This report summarizes the discussions held throughout the *Power Sector Transmission & Distribution Data and Information* webinar series and summit.

DOE OE partnered with the Pacific Northwest National Laboratory (PNNL), Los Alamos National Laboratory (LANL), and Oak Ridge National Laboratory (ORNL) to organize the webinar series and summit. Each of these labs was critical in organizing the key points of discussion throughout the proceedings and developing this report.

Section III of this report covers key discussions from the webinar series and Section IV covers key discussions from the summit. The final section, Section V, reflects on and summarizes these findings.

III. Webinar Series

OE sponsored a four-part webinar series to facilitate discussions about power grid data. The webinars were held on October 11, 18, 25, and November 1, 2023. The series allowed experts to share research on power sector data and sensors and to identify gaps and opportunities for further technological advances. This section summarizes the topics discussed during the series.

The presentation material used in the webinar series can be found here: Power Sector Transmission & Distribution Data and Information Webinar Series | Webinar | PNNL

A. Topic 1: T&D Information Sharing

Host: Supriya Chinthavali (Group Leader, ORNL)

Presentations covered current data-sharing practices between transmission, distribution, and microgrids and discussed standardized data exchanges needed for future grid reliability. Experts presented on (1) transmission and distribution data sharing, (2) microgrid data landscape, and (3) regional data sharing for wide area situational awareness.

Transmission and Distribution Data Sharing

Presentations included VELCO – MAPLE LEAF/MAPLE BRANCH, presented by Alex Anderson (PNNL) and Dan Kopin (VELCO), GRIDS Overview: DER/Emergency/real-time telemetry data communication, presented by Supriya Chinthavali (ORNL), and Data Standards Landscape, presented by Scott Coe (GridOptimize).

Presentations highlighted several DOE initiatives and programs targeted at data exchange at the transmission level, distribution level and across the T&D interface. Key points discussed during the presentations included:

- The North American Electric Reliability Corporation (NERC) currently has no policies that apply directly to cloud-based systems, which is a limitation for utilities that want to move to the cloud.
- The MAPLE LEAF and MAPLE BRANCH cloud-based systems support research and development use cases only. No cyber assets that could have a 15-minute impact on the reliable operation of the bulk electric system (BES) are involved in this research.
- The MAPLE BRANCH, and other DOE-funded research projects, will evaluate whether and to what extent reactive power (Q) capability can be leveraged to support transmission-interface load power factor requirements.

- GRIDS, an OE-funded project, partnered with T&D utilities and developed network model management reports. The primary goal of GRIDS is to improve interoperability through standardized data exchange, specifically at the T&D interface, which can improve decision-making by increasing data availability and access to metadata.
- Participants discussed current data standards for utilities and DERs, distribution forecasts, and load shedding. Specifically, participants mentioned Common Information Model (CIM) and Multispeak, with CIM having applications in both T&D.

Microgrid Data Landscape
Presentation titled COMMANDER-Network of Microgrids, presented by Cameron Brooks (Think Microgrid).

Think Microgrid, in collaboration with ORNL, engaged with industry leaders to understand the data ecosystem surrounding microgrids. Several of the parties involved presented initial research results, which reveal that the data ecosystem surrounding microgrids includes the taxonomy of microgrid configurations, and data use cases and representative data flows of existing microgrids.

Regional Data Sharing for Wide Area Situational Awareness
Presentation titled Discovery Through Situational Awareness, presented by James Follum (PNNL).

PNNL presented an interconnection-scale view of the grid and discussed the reliability benefits that Phasor Measurement Unit (PMU) data flows can provide to system operators. PMU data can be shared over existing networks to support various applications in the transmission sector.

Q&A Discussion Highlight
Q: Given that Eastern Interconnection Situational Awareness Monitoring System (ESAMS) was demonstrated with the Eastern Interconnect Data Sharing Network (EIDSN), why would you now move to the cloud? Doesn't that raise security risks?

A: As detailed in a recent North American SynchroPhasor Initiative (NASPI) webinar (https://www.naspi.org/node/978), there are pathways to full Critical Infrastructure Protection (CIP) compliance while using the cloud. However, system operators have different attitudes toward moving to the cloud. During the initial demonstration, data was shared over EIDSN, but ESAMS was hosted by a single system operator. For long-term deployment, it does not make sense for one operator to shoulder the burden. EIDSN considered hosting ESAMS, but not every system operator in the Eastern Interconnection was interested. Hosting ESAMS in the cloud makes it easier to share the burden among interested system operators without having to get every system operator on board.

Q: ESAMS focused on the Eastern Interconnection where EIDSN allowed the data to be shared. Is data sharing significantly harder in other interconnections?

A: No, it is not necessarily harder in other interconnections. For example, the U.S. Western Interconnection has a data-sharing agreement. This arrangement permits data exchange between, such as that facilitated by the Western Interconnection SynchroPhasor Program-Wide Area Network (WISP WAN) - https://naspi.org/sites/default/files/2021-10/D2S6_02_brancaccio_quanta_20211006.pdf. Entities may be exchanging data bilaterally now.

B. Topic 2: Cross-Sector and Open Data Sharing and Risks

Hosts: Eric Andersen (PNNL) and Jim Follum (PNNL)

Presentations focused on sharing power sector data with trusted third parties and the public. Experts presented on (1) utility data-sharing risks and curation, (2) publicly available measurement repositories, and (3) DOE energy data initiatives.

Utility Data-Sharing Risks and Curation

Presentations included Utility Information Sharing: Challenges and Opportunities, presented by Jim Ball (WAPA), and Utility Data-Sharing Risks and Curating a National PMU Data Set, presented by Eric Andersen (PNNL).

Regulation is a significant hurdle for utility data sharing, especially NERC CIP standards at the transmission level and standards to protect Critical Energy Infrastructure information (CEII). Sharing data can be a liability, so most utilities rely on legal indemnities and nondisclosure agreements (NDAs) to share data with non-utility third parties. WAPA's Jim Ball suggested that utilities need a NERC "get out jail free" card[1] to minimize risk for good faith data sharing. PNNL is working with WAPA and Bonneville Power Administration (BPA) to better understand and mitigate data-sharing risks.

There was a brief discussion on the challenges of developing the National PMU Dataset. Challenges included accessing, cleaning, and protecting the data. To curate the data set, developers had to address inconsistencies in the data such as non-standardized significant digits, redundancies, and missing data. Additionally, they used anonymization to remove the names of substations, geo-tags, and any other indicators that could reveal the location or owner of the data source. NDAs were signed with both data providers and data set users, with explicit instructions for sharing.

Someone asked whether utilities had an interest in standardizing data formats. The format of data exports is largely determined by the archiving system used. While utilities are not likely to request specific data formats from their vendors, there are some common approaches, such as COMTRADE and JSIS-CSV.

Publicly Available Measurement Repositories

Presentations included The Grid Event Signature Library, presented by Aaron Wilson (ORNL), The Transmission Signature Library, presented by Jim Follum (PNNL), and GridSweep Instrument: Sharing the Data, presented by Alex McEachern (LBNL).

The Grid Event Signature Library (GESL) was developed as a platform to host and catalog electric grid events. The library was initially populated with distribution signatures and was expanded to include transmission signatures. Someone asked whether the signatures, which are several years old, are withstanding the changes to the bulk electric system (more inverter-based resources [IBRs], etc.). Presenters explained that they have been working with the transmission signature library (TSL) data to publish historical signatures, but they also have access to newer data sets. They have not yet seen changes that hinder the topics that they are working on but expect to continue adding to the TSL and GESL so they can effectively analyze emerging types of events. In particular, IBR event signatures are needed, and modern telemetry may not be optimal to detect, record, and provide those signatures. Additionally, there has been an interest in connecting external data sources to align with grid data (e.g., weather).

Alex McEchearn posed a question during his presentation which deserves further consideration. He

[1] This is a Monopoly board game reference. A "Get Out of Jail Free Card" protects a player from punishment if they land in jail.

explained that transmission utilities act as if they own the data, at least up to the step-down transformer. But in many instances, these data, such as frequency, rate of change of frequency, and some signatures, can all be detected and retrieved by consumers at the outlet. Therefore, McEchearn questioned why transmission utilities protect those data at the transmission level. Why aren't there more data in the open domain, especially those that are readily measured further down the line?

DOE Energy Data Initiatives
Presentations included ODIN – Outage Data Initiative Nationwide (ODIN), presented by Supriya Chinthavali (ORNL) and Open Energy Data Initiative (OEDI) for Solar Data and Analytics, presented by Kemal Celik (SETO).

DOE continues to make investments to make data more accessible and useful. ODIN and OEDI are two such examples.

Outage data are valuable to customers, neighboring utilities, and regional emergency management partners, but they are often unavailable, fragmented, and/or unstandardized. ODIN established a network of electric service providers who are committed to exchanging comprehensive and interoperable power outage data with designated stakeholders. This exchange can support restoration, reliability, risk mitigation, emergency response, and more.

While ODIN enables data exchange between stakeholders on the grid, OEDI supports modeling and model validation. OEDI enables access to data and algorithms for solar integration simulations. Capabilities include the adaptation of power systems analytics for distribution networks with high distributed solar generation participation, robust physics/network model-based algorithms, new ML algorithms based on large data sets, steady-state and transients' analysis, and data interfaces (CIM, OpenDSS, Gridlab-D).

C. Topic 3: Sensor Systems and Platforms

Hosts: Russell Bent (Scientist, LANL) and Andrey Lokhov (Scientist, LANL)

Experts presented on (1) power system operations with inverter-based resources, (2) localization of forced oscillations, and (3) modeling and analytics using sensor data.

Power System Operations with Inverter-Based Resources
Presenters: Nayak (EPG) and Yang (NREL)

We have limited models for understanding the dynamics and stability of modern power systems with a significant penetration of IBRs and their interactions with the bulk grid. As more IBRs come onto the grid, shared data platforms with analytical capabilities are needed that can leverage multiple data sources and provide real-time, as opposed to offline post-event intelligence. A few examples of platforms and use cases of grid operations with IBRs are mentioned below:

- Real-time platforms for power system operations with IBRs
    - Existing and future use cases include IBR monitoring, oscillations, asset health monitoring for equipment failure prevention, unified data platform, situational awareness, integrating grid dynamics analytics in real-time, grid resiliency and real-time assessment
    - Data sources include PMUs, DFRs, Point-on-Wave, Wide Area Monitoring System (WAMS), and Energy Management System (EMS)/Supervisory Control and Data Acquisition (SCADA)

- Distributed model identification for IBRs
    - Methods for identification of the nonlinear dynamical interactions between the bulk system and IBRs using sensors operating at a faster timescale compared to the dynamics of synchronous machines

Localization of Forced Oscillations
Presenters: Maslennikov (ISO NE), Lokhov (LANL)

Forced oscillations (FOs) are caused by abnormal operating conditions and may excite interarea modes and create large, undesirable transfers of energy. The only effective way to mitigate FOs is to locate and remove the forcing signal. Existing oscillation management tools, such as ISO NE's Dissipating Energy Flow, perform well for today's operational needs. However, increasing deployment of IBRs will likely create sub- and super-synchronous oscillations under resonance conditions, which existing management methods cannot handle.

For future research and development, traditional PMUs must be complemented with high-sampling rate synchronized measurements. High-speed measurements, of at least 200 samples per second for some applications, may require distributed implementation to avoid continuously sending high-resolution data to the server. Additionally, robust methods are needed for locating the source of FOs even with missing system parameters or samples, bad data, outliers, and noise. OE is investing in the development of maximum likelihood-based approaches, which could simultaneously identify the model and source of FOs under resonance conditions. The presentation included illustrations on real PMU and Frequency monitoring Network (FNET) data.

Modeling and Analytics Using Sensor Data
Presenters: Bernstein (NREL) and Kundu (PNNL)

Data-driven modeling, monitoring, and control should make use of all available data types for the best performance, including both traditional data, such as load flow and voltage measurements, and harmonic data. Algorithmic and modeling frameworks should take advantage of data fusion from both low-speed traditional sensors, such as SCADA devices and smart meters, and more advanced high-speed sensors, such as µPMUs.

Capabilities developed under OE funding at NREL include co-optimization of T&D, model-free real-time optimal power flow, optimal power flow with state estimation in the loop, real-time state estimation using stochastic gradient descent, online data-driven predictive control, fault and incipient failure detection, and dynamic flexibility estimation.

In continuation of the research presented, there is a need to study the feeder-level impact of harmonics, including (1) eddy (harmonic) current losses overheating transformers, (2) inaccurate load models leading to sub-optimal dispatch of DERs, and (3) the extension of EV charging load impact studies at distribution networks.

D. Topic 4: Sensor Data and Device Research

Host: Jim Ogle (Electrical Engineer, PNNL)

This webinar featured presentations that examined the data needs of an increasingly distributed grid, the role and value of higher resolution measurements and scalable sensor platforms, and analytics to assess the integrity of measurement data. Experts presented on (1) grid data analytics for a distributed grid, (2) measurement systems and sensor platforms, and (3) distributed grid operations.

Grid Data Analytics for a Distributed Grid
Presenters: Ogle (PNNL) and Scaglione (Cornell)

New data-driven methods are needed to operate DERs, and intelligent and responsive grid edge technologies. Data needs are becoming more complex and demanding, as the number of devices and entities sharing and using grid data increases. New distributed controls and coordination techniques are being applied to manage this scale. Additionally, the increasingly complex data landscape at the grid-edge changes data needs for system operations, to include non-utility resources for example, and introduces new privacy considerations.

Data producers and consumers are increasingly distributed, and the magnitude of distribution system data is increasing with more generation, responsive loads, EVs, sensors, and intelligent devices on the distribution system. With these changes, new data-centric analysis methods are needed to quantify data needs and co-optimize controls, communications, and measurement. This presentation highlighted a tool to optimize sensor placement to maximize observability of the distribution system.

As resources come onto the distribution system, a large amount of new data is available on consumer demand, behavior, and technology adoption. Data governance and transparency are critical for effective data sharing. Several countries, and some states such as California, have policies to protect grid data. Some potential solutions, for ensuring privacy, include secure multiparty computation, homomorphic encryption, and differential privacy.

Measurement Systems and Sensor Platforms

Presenters: Fahmy (Zaphiro), Kuruganti (ORNL) and Chen (Gridware)

As the grid becomes more distributed, more visibility of the distribution system and grid edge is needed to ensure reliable and responsive system operations. To meet this need, measurement systems and sensor platforms are adapting with higher sample rates, time-synchronized measurements, and distributed intelligence computing and sensing platforms.

Time-synchronized measurements at the distribution level have demonstrated use cases, including dynamic protection, DER monitoring and control, and forecasting and planning. For widespread deployment, there needs to be better standards for distribution PMU, a larger ecosystem of vendors to decrease costs, and a research community to identify high-impact applications for business cases. New, low-cost wireless sensor technologies have been developed to facilitate measurements at the edge and behind the meter. Using these types of sensors for timely information in operations requires distributed measurement and control architecture with real-time edge-utility data processing and analytics for computational and communication efficiency.

Many electrical system failures and wear mechanisms are a function of physical and environmental factors that are not measured directly through grid sensors. New breeds of grid sensors that collect non-electrical, mechanical, and environmental measurements have revealed incipient failures, the location and type of faults, and asset health.

Distributed Grid Operations

Presenters: Riepnieks (PNNL), Mahapatra (PNNL), Moutis (CCNY)

As we increasingly rely on data to plan and operate the distribution system, data integrity is critical. Research is exploring methods to validate measurement data at the source and detect anomalies in sensor data streams. Having trusted, accurate measurements is critical for reliable and resilient operations.

Power system measurements and standards largely assume that the electrical signal is sinusoidal. There is currently no standard definition for how a power measurement algorithm manages uncertainty in data in real operations. DOE projects have focused on developing techniques to identify when measurement data are suspect at the source and to inform the data consumer whether the data should be trusted. Updated IEEE standards, for the measurements of electrical power quantities under different signal conditions, are expected to be published in 2024.

Model-informed, advanced ML techniques have been employed to detect anomalies in sensor data streams, such as time-synchronized measurements from PMUs in real time. These techniques are agnostic to the cause of the data anomaly (instrument malfunction, misconfiguration, aging, etc.). Methods have been developed to generate synthetic data sets to train and validate data anomaly and integrity tools because there is limited real data. Open-source tools and data, from DOE projects, will be made available to the research community.

Aging equipment, limited monitoring and automation, unplanned DER and EV integration, and lack of digital models and data for forecasting and analysis make it difficult for distribution operators to ensure reliability. High accuracy and high granularity monitoring can enhance distribution system reliability with the detection of faults, power quality issues, reverse power flows, transformer monitoring, and conductor sensing for fire detection.

IV. Summit in Washington DC

DOE OE, with PNNL as the host, held a summit with the participants from the webinars as well as other industry experts, researchers from academia and national laboratories, and government stakeholders to discuss power sector data needs. This two-day summit was held on November 14 and 15 at 1250 Connecticut Ave NW, Suite 700, Washington DC 20036.

The purpose of this summit was to understand the power sector data landscape, discuss existing data strategies around the world, and identify gaps and opportunities to improve the U.S. data strategy. The following subsections summarize the key points discussed during the summit:

A. Data Strategies Around the World

Large-scale data strategies are typically set to a decadal horizon. For example, the U.S. Federal Data Strategy[2] was developed in 2018 by a 57-member committee representing 23 federal government agencies. The key goals of the 2018 U.S. data strategy are to:

- Improve the quality and usability of federal data,
- Increase public access to government data,
- Promote the use of data to inform decision-making, and
- Foster innovation and collaboration across federal agencies and with external stakeholders.

Similarly, the European data strategy aims to make the EU a leader in a data-driven society. Their goal is to create a single market for data that will enable free data flow within the EU and across sectors for the benefit of businesses, researchers, and public administrations. The EU also conducted several workshops in FY2022 covering the following energy data strategy related topics[3]:

- Efficient integration of the EVs into the electricity system and the data-sharing framework
- Priorities in the energy transition that require enhanced data exchanges at the EU level
- Data sharing in renewable energy: experiences, and incentives organized by EUREC

[2] https://strategy.data.gov/

[3] https://commission.europa.eu/events/workshops-digitalisation-energy-system-2022-02-16_en

- Promoting cooperation between digitalization of energy centers of expertise and innovation hubs
- Prosumer data exchanges to enable seamless flexibility integration into the TSO-DSO markets

The UK and Australia have also been leading the way in this area and their efforts are being viewed as potential best practices that can be leveraged in the U.S. energy sector. The Energy Security Board of Australia (ESB) is implementing a data strategy to unlock data as an enabler in the energy transition. It provides an overarching consideration of the energy sector's existing and future data needs, supporting the needs of consumers, industry, and policymakers in the energy transition. The ESB's data strategy focuses on workstreams to address critical issues related to energy data access and sharing and design options to address emerging data needs for DER integration.[4,5] The Energy Data Taskforce (EDTF) was commissioned by the UK Government, Ofgem, and Innovate UK to develop an integrated data and digital strategy targeted to unlock the opportunities of a modern, decarbonized and decentralized energy system for the benefit of consumers.

B. Discussion Areas

During the summit, the power sector data landscape was explored through a series of panel, breakout, and roundtable discussions. Participants shared perspectives on the emerging trends for power system data in an evolving energy sector with a forward-looking view toward the next 10 years. The general categories of power system data sets that were discussed in the summit are listed below:

- Grid models (T&D)
- Grid measurements and state (T&D)
- Meter and behind-the-meter data
- DER location, capabilities, settings
- DER telemetry
- DER and load forecast
- Outage data
- EV data
- Historical and event data
- Climate data

The participants further discussed pathways to establish a power system data strategy to support the evolving data needs of the energy sector. The key considerations that emerged from these discussions can be broadly categorized into five categories: (1) Data Needs, (2) Data Interoperability, (3) Open and Closed Data Sharing, (4) Data Governance, and (5) Data Privacy.

1. Data Needs

The participants identified different use cases for power sector data, such as planning, operations, automation/control, T&D and edge coordination, post-event analysis, forecasting, asset management, and research (algorithms, optimizations, AI/ML). These use cases can be further categorized into the following categories of emerging data needs:

---

[4] https://esb-post2025-market-design.aemc.gov.au/data-strategy

[5] https://www.aer.gov.au/networks-pipelines/guidelines-schemes-models-reviews/network-visibility

Extended Scope of Network Model Data

Various participants stressed the need for electric model data including topology and equipment parameters for transmission, distribution, and edge systems. Accurate network models and synchronization between planning and real-time operational models are critical for managing a more dynamic system. Spatial resolution mapping of multiple data sources was also identified as one of the key steps required for an effective analysis and support decision-making. The participants further discussed various challenges faced by different entities in maintaining an accurate electric network model, as listed below:

- A large number of distribution utilities do not have sufficient feeder-level data and an appropriate representation of dynamic load profiles and DER capabilities.
- There is insufficient knowledge of connected non-utility DERs for planning, operations, and safety (e.g., the type and location of connected DER, whether the resources are controllable, the inverter settings).
- An accurate distribution system electric model and limited as-operated visibility is lacking at the distribution level, e.g., underfrequency load shedding parameters.
- Portions or aggregations of model data may need a platform to enable data sharing across ownership/operator boundaries at the interconnection points between transmission and distribution, and distribution and edge systems.

Greater Behind-the-Meter Visibility

Increasing adoption of DERs and EVs is resulting in a more dynamic load. Fast-acting IBRs, and significant and sudden load changes associated with EVs drive the need for more frequent and granular data for operations and control. These characteristics are impacting both transmission and distribution, further emphasizing the need for shared data across T&D boundaries. Keeping these considerations in mind, the participants discussed the need to better understand, forecast, and share real-time meter data, all of which could potentially be achieved by an expanded view of operator visibility of load-side behavior including near real-time information from meter data, higher sample rate meter data, voltage, harmonics, and power quality data, DER and EV telemetry data, and disaggregated load information particularly for large loads such as HVAC.

Accurate Grid State and Higher Fidelity Measurements

Increasing IBRs, more variable renewable generation, and the changing grid-edge landscape, as noted above, are driving the need for more granular data to coordinate across a much more dynamic, coupled grid to ensure reliability and resilience. The participants identified the following considerations to highlight the importance of more granular data:

- High sample rate, point-on-wave measurements are increasingly critical to respond and better understand events such as oscillations.
- Transmission systems, distribution, and grid edge are becoming more coupled. The need to analyze behavior across interconnected systems and organizations increases the need for time-synchronized measurements and compatible measurement resolutions.
    - Potential use cases of high-fidelity monitoring at residence: new heat pumps that require 1.5 kVAR of reactive support when OFF, homes that switch from 3 kW supply to 10 kW load in 1 minute because of passing clouds and loads switching on.
- Measurements for a dynamic and coupled grid are increasingly becoming much more complex. Realistic synthetic data sets and associated models, and anonymized historical

data sets for postmortem analysis are required to support the research community and industry to facilitate development of the necessary approaches to support grid planning and operation needs.

Continual Demand-Side Forecast Data

The future grid with high levels of DER and electrification requires more complex and dynamic load-side models to support system planning and operations. The participants discussed a few characteristics of a dynamic load-side model as listed below:

- New load models that represent new demand profiles including DER and electrification
- Short-term updates to load models to support operational forecasts and operational decisions
- The level of uncertainty of a given forecast data to support decision-making
- More real-time data on the available demand response or flexibility capacity resources to support transmission or distribution operations

Other Data

The group also discussed a wide variety of data that is not directly produced from the connected electrical systems but can be used by the energy sector to support planning, operating, and coordinating across organizations, some of which are presented in Table 1.

Data Category	Examples/Use Cases
Environmental Data	- Real-time weather data and climate models to understand solar and wind resource behavior as well the operating context such as thermal limits and storms - Localized weather forecast data with more geographic granularity - Thermal data for equipment and line status - Historical and real-time weather data for operations, planning, and research with compatible resolutions - Satellite data, drones/camera/video feeds to understand the physical environment
Cross-Sector Data	- Electric vehicle and charging telemetry and growth projections; traffic patterns to understand where and when EVs will be connected - Communication network status to understand limitations in visibility/controllability - Models and test systems for interdependent infrastructures analysis and planning (e.g., gas and electric) - Social vulnerability data
Markets and Financial Data	- Types and costs of preventive and corrective utility measures to enhance grid resilience planning - Market data to formulate pricing for energy and capacity services - Grid side data to support pricing validity by energy service providers

Table 1: Example use cases of other data used for grid operations

Crosscutting Data Needs

The participants discussed various crosscutting data needs that are important to ensure data accessibility and validity. The group noted the importance of methods to validate the data, establish trust in the data source, and identify potential anomalies to establish data integrity for critical operations. The participants also noted that simplified data access and management is needed to allow large-scale use of data. Additionally, availability of spatial and temporal granular data as well as high licensing costs were identified as the main barriers to reliable forecasting.

2. Data Interoperability

The participants discussed various considerations associated with data interoperability between various segments of the electric grid, as shown in Figure 1. The group noted that often the "desired data" for sharing are just data from neighboring spaces as shown in the figure. Common variables/inputs in equations at different positions shown in the figure pose opportunities for interoperability given the continuity of physical laws across boundaries. Additionally, details of device access, topological context, standardized and common taxonomy for data and operator logs, data governance/CIM to enable further analytics (including large language models for noncritical tasks) are important issues to consider when thinking about data interoperability. The group also discussed two challenges associated with data interoperability:

- Desired high-resolution of data poses problems with storage of data for subsequent interoperability and analysis. Current practice often includes continuous erasure of historical high-resolution data, which also raises the issue of defining "interesting" data or information that needs to be stored for future analysis given the uncertainty associated with future data needs.
- Currently, customers and relevant players often do not want visibility of data lower down the hierarchy, as presented in Figure 1, as it could unnecessarily complicate their analysis. Because the grid is rapidly evolving, the customers still need to have access to these data, which could then be used to inform potential modeling approach, analysis, and operations in the future.

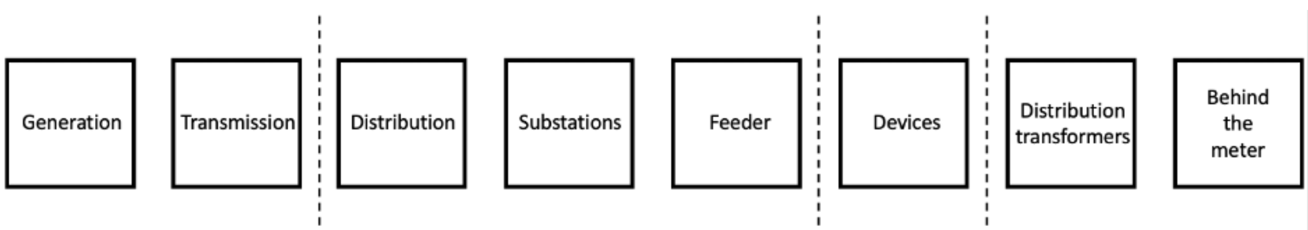

Figure 1: Flow of data from the utilities' point of view

3. Open and Closed Data Sharing

The group discussed the landscape of current data sharing, considerations for open and closed data-sharing environments, and the need for data governance. Table 2 summarizes key discussion points.

Topic	Key Points
General Data Sharing Needs	Traditionally, the utility has been the asset and data owner. With increasing integration of non-utility resources, microgrids, and interactive loads, there is a need to share data outside the utility boundary across many different stakeholders.
	Similarly, with increasing integration of grid edge technologies, there is an increased coupling between the transmission and distribution systems, driving the need to share more information across the T&D boundaries.

Topic	Key Points
	The research community needs access to model and large sets of realistic data to analyze system behaviors, characterize state-of-art innovations, and train AI/ML approaches.
	There is a growing need for cross-domain sharing of data sets with similar resolutions, e.g., electrical, transportation, environmental, social, and weather data.
	There is a need to identify solutions that allow stakeholders to gain necessary insights while also respecting ownership and sensitivity of the underlying data.
Outage Data Sharing	Outage data sharing from utilities has been successful through the DOE's Outage Data Initiative.
	Some non-utility data vendors scrape and gather data from sources that have not been vetted for accuracy. In many cases, news vendors subscribe to these data sources and risk spreading potentially erroneous data.
Edge Data Sharing	DER and EV data and telemetry sharing are mostly closed but critical to ensure reliable and resilient operations in future grid across transmission, distribution, and edge systems.
	DER Registries are emerging with more open data-sharing constructs around DER location, nameplate, services, or capabilities.
	Meter data has been mostly closed but there have been some instances of meter data sharing. - The Green Button Initiative led to more open meter data sharing, especially to customers. - Some states have policies regarding sharing of aggregated, clustered data (15/15-Rule: at least 15 customers aggregated with no one customer greater than 15% of group's consumption).
	One example of a need for more open data sets are capacity maps, showing where solar and batteries can be connected to the grid. This would require some level of electric network model information to inform the capacity map and anonymization to ensure end-user data protection.
Utility Data Sharing	Crafting data-sharing agreements between utilities often involves lengthy NDAs and a large amount of effort.

Table 2: Key points related to open and closed data sharing

4. Data Governance

Data Governance is needed to establish trust in data custodians for shared data spaces. In the UK, the assumption is that data are open and available but, in the United States, especially for IOU entities, all data are closed. There are uncertainties regarding the extent of data sharing allowed and associated risks which often results in utilities' reluctance to share the data. Hence, it is important to establish trust and ensure integrity of the shared data which can be achieved through a well-defined data governance structure. Improvements like regulatory policies from entities like FERC, NERC, and DOE, enforceable data measurement and quality standards, audit or calibration requirements, and metadata requirements to include data source and accuracy specifications are important to strike a much-needed balance between a fully open and a fully closed data-sharing environment in the United States.

5. Data Privacy

In addition to data needs and data-sharing requirements, the participants also discussed data privacy considerations to enable the value of data sharing while respecting the privacy of the data producer. Table 3 categorizes key discussion points into the needs, challenges, and solutions relevant to data privacy.

Data Privacy Topic	Key Points
Needs	Need for aggregated, anonymized, and load composition data.
	Need to ensure that data recipients protect data, control sharing, and do not use the data beyond agreed consent.
	Need for cybersecurity protection standards to be included as a part of data governance or data-sharing agreements for recipients of utility data.
	Need to understand the extent of information sharing that is possible/allowed, especially in scenarios involving data related to actual grid events that can be used to support dynamic modeling of the grid.
Challenges	It can be a challenge to derive, from aggregated data, continuity across spatial, temporal, and sensitivity scales.
	Grid operators can be reluctant to allow AI/ML models to be trained on their data due to privacy concerns, as the original data could potentially be recovered from the learned model.
Solutions	Derivative data products, that enable data analysis for the consumer without providing access to the underlying detailed data, could be one of the key approaches to balancing data needs and privacy considerations.
	Role-based access (like ERCOT) or highly detailed private and reduced public data or other models can enable access at the right level for the right user.

Table 3: Key points related to data privacy

V. Conclusion and Summary of Key Findings

In addition to raising awareness about research and industry advancement in power system data analytics, OE's workshop provided an opportunity for experts to identify gaps and future research needs in this space. Key takeaways from the webinar series and summit include:

- Accurate, accessible, and timely grid data are critical for grid reliability, affordability, and security.
- Power system data has immense potential that the sector is only beginning to tap into. It will be a challenge to harness these potential benefits while ensuring privacy.
- Standards and governance will be critical for vetting and standardizing data and facilitating exchange of useful data and information.
- Because the grid is rapidly changing, there is an increasing need for real-world and synthetic data to foster understanding of the conditions of the modern power system and facilitate faster innovation.
- A shared framework and roadmap for power sector data would build stakeholder confidence, accelerate innovation, and ensure the adoption of common technologies and standards.

U.S. DEPARTMENT OF ENERGY | OFFICE OF ELECTRICITY

DOE/OE-0042 March 2024