Integrated Energy Data Resource Stakeholder Comments

Submitted by:
Jess Melanson, Utilidata President & Chief Operating Officer
jmelanson@utilidata.com / 401.383.5800 / www.utilidata.com
225 Dyer Street / 2nd Floor – CIC / Providence, RI 02903

Utilidata greatly appreciates the opportunity to provide input regarding New York's Integrated Energy Data Resource (IEDR) plans. As a software provider with over a decade of experience in distribution grid operations, we have unique experience capturing, analyzing, sharing, and making use of primary system and grid-edge data, in service of many of the goals New York State is trying to meet, including distributed energy resource (DER) development, transportation electrification, building electrification, energy efficiency, and electric utility function.

We commend the Public Service Commission and NYSERDA for recognizing the essential role data access plays in the energy sector transformation. Providing data from utilities to DER companies, energy service providers and other innovators can help drive tremendous value for customers, the overall energy system, and the effort to quickly and affordably decarbonize. Sharing data can also help reduce the information asymmetry that currently exists between utilities and various parties attempting to interface with the utility grid.

However, the success of the IEDR initiative hinges on the ability to deliver "useful energy-related data." In this context, the term "data" should be used to describe precise, actionable information. In Utilidata's experience, turning raw system data into precise, actionable information is a very complex process, requiring a great deal of attention given to what information is captured, how it is analyzed, in what time scale it is transmitted, and how it is put into context with other grid data. To say that the devil is in the details would be an understatement. Most open data initiatives fail not for lack of good intentions by the parties involved, but because these technical complexities were not navigated successfully. Developing a platform that can share relevant and actionable utility system data with third parties requires extremely close collaboration between the utility, third parties, and the provider of the data platform.

We know this to be true because we have experienced these challenges first-hand and are working collaboratively with utilities, DER developers, and national labs to overcome them. Our real-time machine learning grid optimization software is used to process, contextualize, and share data from utilities. In New York, we are currently undertaking a project funded by NYSERDA with National Grid, the National Renewable Energy Laboratory (NREL), and Standard Solar to operationalize "open data" between the grid and a solar farm in Clifton Park. This project exemplifies the collaboration between utilities and third parties that is necessary to achieve the use cases NYSERDA has identified. Our experience with this project, combined with our decade of experience operationalizing real-time

system data, highlights key foundational elements that must be addressed to achieve any of the use cases NYSERDA has identified.

While the prioritization of use cases is a worthwhile endeavor, focusing exclusively on the data needs of individual use cases -- rather than the foundational system data that could be collected and analyzed to support a broad range of use cases -- may unnecessarily limit the scalability and effectiveness of the IEDR. Therefore, we recommend that in parallel to the use case evaluation, NYSERDA also select a number of substations, in addition to the one at Clifton Park, to pilot the data collection, analysis, and sharing process to ensure that useful and relevant system data can be made available in the IEDR.

Outlined below are lessons we have learned thus far that we hope will aid NYSERDA in developing an IEDR that can not only support an initial set of prioritized use cases, but quickly and easily scale to support a wide range of use cases.

1. **The majority of primary system data needed for most IEDR use cases is available via two utility data sets.** The first crucial data set of system measurements can be obtained from the utility's SCADA system. Nearly all of the essential system measurements (including demand, current, voltage, frequency, and VARs) from primary system devices (including substations, station load-tap changers, line regulators, capacitors, reclosers, and line voltage monitors) are available via the same data transfer path from a utility's SCADA system. The process of pulling demand data from the substation, for example, is the same as pulling current data from a capacitor. Therefore, NYSERDA would be well-served to collect all of these measurements at once, rather than prioritizing the collection of a single measurement to support a prioritized use case.

   The second crucial data set is the asset description and hierarchy, which provides information about the physical attributes of the grid and connectivity between assets. This data is available via a utility's CYME or SYNERGY model.

2. **Grid-edge data is also essential, but utilities have not integrated it with their primary system data, creating an information gap that must be addressed.** The grid edge is becoming exponentially complex, particularly as New York ramps up utilization of flexible demand and the integration of solar, electric vehicles, electric heat, and batteries. Data from grid-edge end points, particularly advanced metering infrastructure (AMI), provides the most accurate and complete understanding of what is happening at the grid edge.

   Each meter records about 25 million data points per month, including demand, voltage, power flow, and DER operational characteristics. For that data to be most actionable, it should be collected, processed, and managed by a system with powerful grid-edge computing capabilities, as we explained in our public comments regarding National Grid's Benefit Implementation Plan (BIP) in Case 17-G-0239. Existing centralized utility systems are not designed to process this data for operational use, let alone turn it into actionable insights for third-party stakeholders. For example, in our Clifton Park project, because National Grid's SCADA system does not process AMI data, we had to retrieve AMI data from the meter manufacturer's cloud.

New York is making massive investments in AMI. As early as possible in the deployment process, utilities should be planning to integrate that data into a software platform that can make relevant data available for the IEDR initiative.

3. **System data must be analyzed together to truly understand grid conditions.** System measurements can be misleading if not properly analyzed and contextualized, especially grid-edge measurements. A raw data dump of intermittent AMI voltage reads, for example, will not be actionable. AMI voltage measurements are highly variable, given that homeowners and businesses use different amounts of power throughout the day. This variable AMI data must be connected to circuit models and measurements from primary system devices to understand what's happening on the system. The asset hierarchy data is essential as it explains how the various devices are connected, enabling the determination of whether devices with similar measurements are physically connected or operating in the same area of the grid.

   The kind of grid-edge operating system that we described in our BIP comments can analyze and integrate AMI data to determine grid-edge conditions at any customer location, operate the grid more efficiently, and enable the grid and DERs to communicate and coordinate operations. This kind of real-time machine learning system can also combine AMI data with primary system data to provide a more holistic and accurate view of the grid.

   As part of the Clifton Park project, we used our machine learning software system to analyze CYME models, primary device data, and AMI data, as well as metadata describing the relationships between the various datasets. This allowed us to gain a detailed understanding of the interactions between solar inverters and power flow on the system. By collaborating with other stakeholders and engaging in an in-depth analysis, we gained a better understanding of exactly what types of data, data relationships, and data analysis functions would be needed to support our project. Before this process, the stakeholders involved in this project would not have been fully able to answer questions 7, 8 and 9 in the use case profile. We believe our experience can help support a number of other use cases in DER development, transportation electrification, building electrification, and utility function use cases.

4. **Existing system datasets often contain discrepancies or inaccuracies that must be reconciled with machine learning before the data can be useful.** CYME or SYNERGY models are inadequate for understanding real-time grid conditions and contain discrepancies that must be corrected to use the circuit data for any planning exercise, particularly those that are going to change power flow, like adding new demand, generation or modeling price-responsive customer behavior. Machine learning algorithms, coupled with an understanding of power system behavior, can quickly detect and locate inaccurate model descriptions.

   For example, most utility circuit models, including those at Clifton Park, assume ideal conditions, such as three phases that are "balanced" with the same line construction and impedance values, which is rarely a reality. Additionally, many models end at the service transformer rather than showing individual service points. Our machine learning tools allowed us to combine measured data from field devices with open-source GIS, satellite, and road

information, to connect the service points - which came into our system as AMI data - and approximate the length of the service drops from the service transformer to service points. We could then approximate the impedance of the service transformer and correct the circuit models, using real-time data. A machine learning model with a robust feedback loop can use real-time data to continuously improve, adjusting its parameters to ensure it remains as accurate as possible.

5. **Even with the most advanced software platform, implementing an IEDR effectively will require close collaboration with utilities to validate and resolve discrepancies in the data.** Under the IEDR initiative, utilities are being asked to analyze and provide data in a way that has never been done before. In our experience, every time data is evaluated in a new way, it reveals unknown issues and discrepancies. For example, for the Clifton Park project, National Grid provided Utilidata with 150 GB of data, including the 2019 CYME models, primary device data, and AMI data, as well as the metadata describing the relationships between the various datasets. Even though we were using powerful machine-learning models, we still needed close collaboration with National Grid to understand the discrepancies between the simulated and measured data, the hosting capacity analysis results and procedures, technical roadmaps for supporting smart inverters in New York, and selection criteria for solar interconnections as well as forecasting data for the next 15 years.

   Therefore, IEDR cannot simply be a mandate imposed upon utilities, or a passive exchange of information between the utility and third-party stakeholders. It must be implemented in a way that fosters collaboration among all parties and creates opportunities to validate assumptions and provide crucial context. Blindly relying on data without some kind of validation will not work.

NYSERDA will need to confront these realities no matter which use cases it prioritizes. However, once the utilities have a system in place that can process and contextualize the key data sets, tweaking outputs for new use cases should not be a heavy lift, because the platform has been built the right way, from the bottom up.

We believe fully incorporating the lessons learned from the Clifton Park project will best ensure that the state successfully executes the IEDR. National Grid and the other parties have been great collaborators in this process. We would welcome the opportunity to have the project team share the details from this effort thus far. We also believe it would be fruitful to identify a number of substations in each utility service territory where New York can quickly replicate this process, in parallel to the effort to prioritize use cases. Laying the correct system data foundation at different substations in each utility service territory will provide invaluable lessons and best position the IEDR to become an effective resource for third-parties, avoiding the fate of other well-intentioned but unsuccessful pursuits to share "system data."

We appreciate the opportunity to provide our perspective and look forward to continued collaboration as stakeholders across the state work to make the IEDR a success.