Avangrid A member of the Iberdrola Group

The Digital Transformation of Electric Operations

How Avangrid Built a Model for Internal Data Science and Analytics

Acknowledgements

Written and prepared by – Mark Waclawiak, Dr. Nicole Rosato, Brian Denney, Dr. James Healy, Genevieve Rowe

Additional contributions from the Operational Performance organization – Julio Fonseca, Mateus Sartori Marcurio, Flavio Faria do Nascimento, David Florin, Peter Kozun, Nathalia Pinheiro Marcia, Megan Ehrich, Vinson Chen, Joshua Harford, Brianna Coates-Strutz, Jason Serpe, Riann Higgins, Katie Wojnowski, Jackson Whitney, and Adam Steiger.

1. Executive Summary

A transformative change in the way electric utilities leverage existing and emerging datasets is necessary. Modern utilities are some of the largest generators of data and must learn to use it to overcome the numerous challenges posed to the resiliency and reliability of the electrical grid. Avangrid established the Operational Performance organization in 2020 to increase the internal analytical capabilities of electric operations. Through the recruitment and development of data science, engineering, and analytic professionals, the Operational Performance organization has transformed Avangrid's culture and usage of data by rebuilding the core underlying data infrastructure, thereby making significant improvements to data availability, quality, and reproducibility. These strategic and organizational gains were made using low-cost methods and tools that emphasized quality, speed and control while eschewing the traditional avenues of reliance on external vendors or consultants. The result is a foundational data infrastructure tailored for rapid growth and flexibility. Our infrastructure is managed and utilized by a team of internal data professionals with the knowledge and skillset to develop the models and analytics necessary for electric operations' needs.

2. Introduction

Many utility organizations are limited not only by the underlying data infrastructure's inability to facilitate data availability but also by a lack of personnel with the skills to perform advanced analytics. In 2019, prior to the implementation of the Operational Performance (OP) organization, reliability analytics at Avangrid was limited by both factors. By hiring data-focused professionals and investing time to build highly flexible database infrastructure, OP established the necessary foundation for a highly successful data science organization.

An effective data science and analytics organization is critical to the core functioning of electric operations. Utilities have significant reporting responsibilities for both internal tracking and external regulatory requirements - accurate and reproducible reporting is a necessity. Additional reporting needs and requirements are often requested with very tight turn arounds. These frequently include the fulfillment of ad-hoc and often unpredictable regulatory data and information requests or urgent questions from upper management about specific key metrics. If a utility has immature data infrastructure plagued by disparate data sources, disorganized Excel sheets, and fragmented processes, these requests become time-consuming, inaccurate, and hard to reproduce. The inability to effectively communicate a utility's data leads to ineffective responses to regulatory stakeholders. This has significant potential to damage the utility's reputation and relationship with regulatory authorities while fostering the damaging perception of data obfuscation.

The need for larger-scale analyses also emerges through the desire to drive strategic thinking with more effective use of data. More advanced data science can provide predictive information based on historical data, detailed prioritizations for programs and projects, and even system asset verification and tracking through imagery. Without a mature, developed database infrastructure, these types of analyses are challenging, if not impossible.

The infrastructure and personnel challenges a new data organization will face cannot effectively be solved through out-sourcing. The cost, complexity, and the narrow approach of "use-case thinking," leads to inflexible infrastructure that struggles to adapt to the next request or requirement. The outsourcing of infrastructure development also atrophies the internal data skillset, rendering end-users limited in their own control and adjustment of infrastructure to meet new requirements.

The inception of an internal data science team at a utility is an absolutely essential step for forward-thinking organizations. Our data science team continues to succeed at streamlining data processes and pipelines while providing ease of access to data to many organizations within the company. In four years, we have made a transformational change to the company's approach to data. We have repeatedly proven the strength of a data-driven approach to age-old utility problems like vegetation and asset decay, while also coming up with new and innovative ways to use the data that the company already collects. These models, prioritizations and reports are used every day by organizational leaders to make decisions about projects critical to the health of our grid.

In this paper we will lay out the original concept for the data science team, beginning with strategic hires and then blending those data-focused professionals with utility experts. We will outline the high-level details of the core data infrastructure, how we have brought about organizational and leadership buy-in, and how the data science team continues to benefit Avangrid.

3. Building an internal team

The philosophy at the center of the Operational Performance organization is simple; data analytics, data engineering and data science are core functions of a modern utility. Utilities generate massive amounts of data pertaining to the electrical grid; leveraging these datasets to extract maximum insight and value while minimizing expense and risk provides a significant operational advantage. Avangrid itself operates four electric utilities in Connecticut, New York, and Maine, with four operating companies: the United Illuminating Company (UI), New York State Electric and Gas (NYSEG), Rochester Gas and Electric (RG&E), and Central Maine Power (CMP). These companies each generate a massive amount of data every day that, until recently, was significantly underutilized. The Operational Performance team's primary focus is to leverage this data to the benefit of the entire Avangrid organization. We study data pertaining to resiliency and reliability risks on our network (for example, vegetation, aging infrastructure, and types of weather in the Northeast) and are always looking for new data sources and new, innovative ways to grow our core data infrastructure.

The Operational Performance organization consists of three departments:

- Data Science and Analytics is made up of data scientists, data engineers and data analysts from outside the utility industry, but with strong skills in database building, modeling, and machine learning. They are responsible for:
  - Development of underlying data infrastructure (databases/pipelines), customized models, algorithms, and computational methods, as well as more advanced AI models and platforms.
- Reliability Reporting is made up of data analysts with strong utility backgrounds. They are responsible for:
  - Reporting of reliability and resiliency metrics and KPIs to internal and external stakeholders.
  - Development of dashboards and automated reports for core reliability datasets.
- Reliability Engineering is made up of engineers with utility backgrounds. They are responsible for:
  - Development of reliability and resiliency prioritizations, projects, and tools utilizing the core data infrastructure.

Having an in-house data team alongside electrical engineers and analysts is beneficial to all three groups: the data science team improves the overall data culture and teaches their reliability engineering and reliability reporting colleagues programming/data science techniques, while reliability engineering and reliability reporting teams, in turn, teach the data science team technical aspects of electric utility operations.

Upon the team's inception, we considered the common utility practice of outsourcing data science and analytical functions while utilizing internal SMEs in either project support or end-user roles. We identified this to be overly inflexible, expensive, and found it often produces underwhelming results and minimal benefits. In its current state, by recruiting the technical data resources and pairing them with utility electrical engineers and analysts, OP is able to rapidly develop models and infrastructure to meet the day to day needs of the utility.

3.1. Pre-2021 - Legacy Data Processes

Before the implementation of centralized database infrastructure, the process for outage management and reporting was cumbersome, relying on manual data processing in Excel and Access. Excel manipulation and consolidation was at the core of analysis, severely limiting electric operations' ability to derive new insights from reliability data, never mind implement effective data-driven solutions or prioritizations.

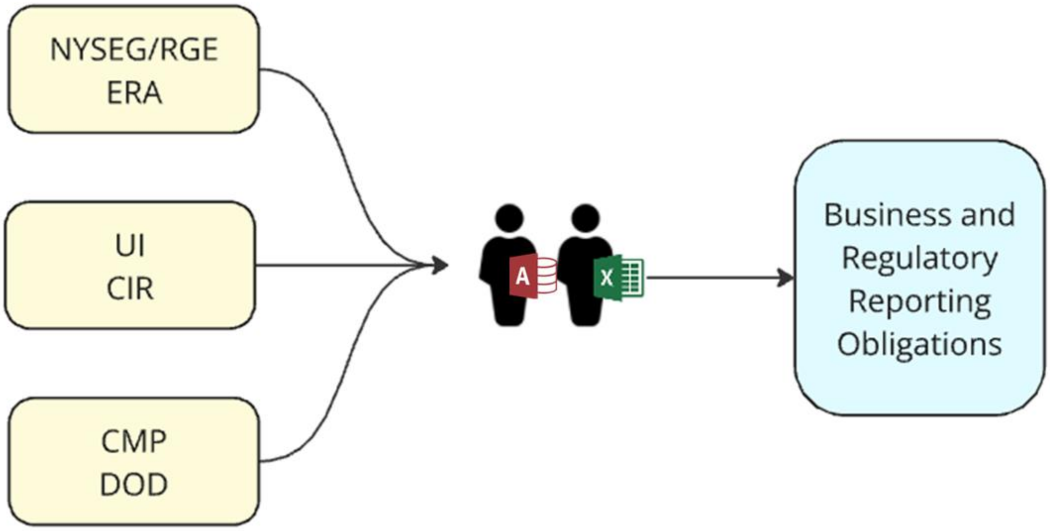


Figure 1: *Three operating companies' outage management systems outputs were copied into Microsoft Access and Excel databases. Regulatory and internal reports were built using these spreadsheets. This process took 1-2 days per week.*

3.2. 2021 - Initial UOD Build

In 2021, the Unified Outage Database (UOD) was commissioned and set up in a SQL Server environment. User permissions were established, and database construction began. We were able to connect the three Outage Management System[1] (OMS) databases, those used in the weekly reporting process (Figure 1), to our SQL Server database. We used those connections to build the first set of Extract, Transform, Load (ETL) codes. We then built SQL queries to restructure data from the OMS systems.

Our ultimate goal was to create one location to track all Avangrid outages; this required consolidating all outages from all four Operating Companies into one Outages table. The SQL queries were challenging; multiple disparate datasets had to be standardized since all three data sources have unique backend structures like column names and datatype conventions. For example, a field that is a number in one dataset may be a string in another even though both columns represent the same piece of information. It was necessary to compare datasets field by field to develop a standard set of columns that would eventually make up the core Outages table. The SQL scripts were incorporated into the ETL codes which were, initially manually, run every day to keep our database current with the OMS systems. Finally, we had

[1] *Outage Management Systems are systems used to dispatch and record outage data.*

our Outages table. This was the most significant data breakthrough in this first period; outages across OPCOs could finally be compared.

Other early database connections included links to supplementary outage data, for example, circuit data, customer counts, etc., and a large-outage monitoring database. We integrated each additional data source into our outages dataset, which provided a robust set of information for each outage.

TSD - The first major new dataset developed in the UOD: When a transmission line or a substation experiences a fault that results in an interruption to distribution customers, the outage records are created on a distribution circuit level. Historically, in the case of upstream transmission or substation faults, outages were attributed to distribution circuits regardless of where the fault actually originated. Inaccurate outage recording not only negatively skewed performance metrics for distribution circuits but obfuscated the impact of faults at the transmission and substation levels and led to misaligned prioritizations of transmission, substation, and distribution investments.

A long-standing goal of operations was to find a way to determine which network level had experienced the fault for a given outage without doing a manual inspection of each incident. Operational Performance had developed a cleaned set of outage data in preparation for our new database, the UOD. An interdisciplinary collaboration of OP technical data experts and electrical engineers used the cleaned data to develop, apply and test an algorithm to classify an outage as Transmission, Substation or Distribution (TSD). The UOD eventually provided the infrastructure necessary to apply this algorithm automatically to new outages. Upon production release, the algorithm was capable of reliably classifying 97% of outages. The remaining 3% was reviewed twice a week by representatives from the three Operational Performance departments. The algorithm has since improved, but outage review continues to ensure not only continuous testing of model assumptions but also allow for further refinement as new data comes in.

The implementation of the TSD algorithm has led to significant improvements in the prioritization of investment across the network for both reliability and resiliency. It has also provided support and justification in Avangrid regulatory dockets and rate cases.

PowerBI reporting: To facilitate more effective reporting, a PowerBI reporting environment was set up. The goal was to replace the historical Excel reports with new dashboards that could be connected to the UOD. Reliable data imports from the UOD allowed reliability reporting to tranistion from weekly to daily updates and eliminated the 1-2 day process of onerous and repetitive report construction (described in Figure 1). Static reports were replaced by interactive dashboards that remained current with the reload schedule.

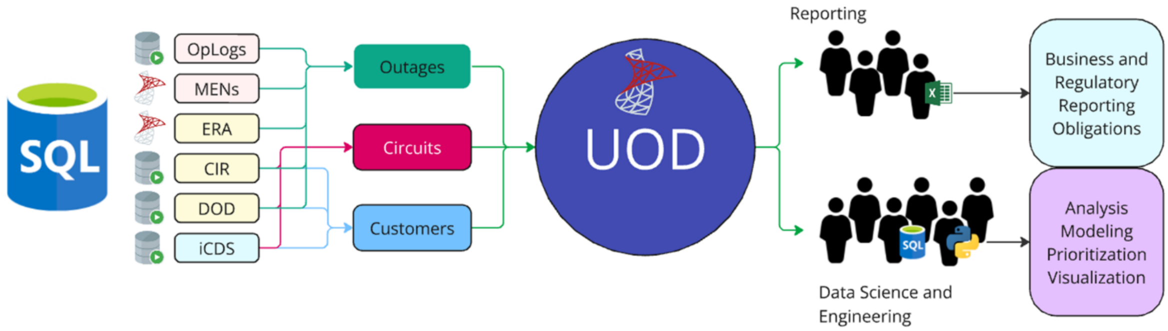


Figure 2: *By mid-2022, our UOD contained databases from six different backend databases and expanded its focus to Circuitry data and Customer data as well as Outages. The UOD facilitated more efficient and accurate reporting and paved the way for true data science to begin.*

3.3. 2022 and early 2023- UOD growth, UAD inception

During the next few years, the UOD grew dramatically in terms of complexity and data. As time went on, we found need for other data sources to be brought into the UOD including: circuit profile data, vegetation trim history, historical weather data, and many others.

A second relational database, the UAD (Unified Asset Database), was set up to manage data from our asset management source of record and GIS systems that pertained to assets and infrastructure. The UAD allowed us to combine two formerly disparate datasets, Outages and Assets, for analysis.

UAD inception

Prior to the creation of a Unified Asset Database (UAD), asset data was housed in two disparate systems that were neither designed for analysis at scale, nor easy to extract data from. As a result, asset analytics for the purpose of asset management required a multi-day process of extracting data for an entire asset class in batches then exporting data to dozens of Excel files. Data cleanup algorithms then had to be applied to each file to both add fields derived from existing columns, and correct known data quality issues. The process was manual, variable, and posed challenges around reproducibility.

Thus, the next core infrastructure development was the addition of a second major database. The UAD centralizes asset data from both systems to enable asset analytics for the millions of assets on our grid. Geographical Information System (GIS) data is extracted monthly, while an automated pipeline for asset data out of the asset management source of record is in development. This automated pipeline will extract 310 million asset records spanning 38 asset classes from the source system monthly, in the form of sixteen disparate extracts consistent with the backend structure of the database.

Using this central asset data repository, our team of data scientists, data engineers, reliability engineers, and reliability reporting analysts can pose and answer a myriad of questions regarding asset health, asset replacement, and outage drivers.

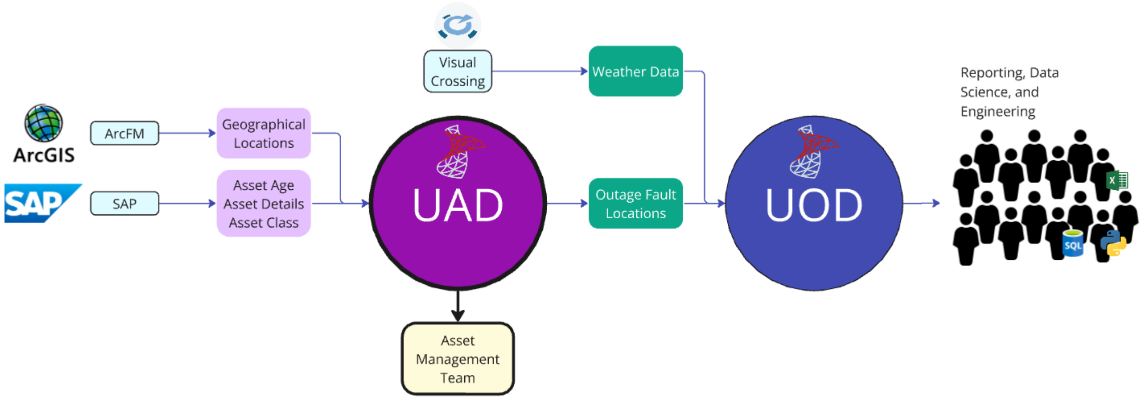


Figure 3: *The UAD added significantly to the core data infrastructure and allowed for analysis of asset data at scale.*

3.4. 2023 - Linux Server, Image Storage and Reloader framework

In addition to data science and analysis, the ability to move and process data at scale is a core necessity of an internal data science organization. As our databases grew in size, our traditional laptops did not have enough processing power to process the amount of data we dealt with every day. This led to Operational Performance commissioning a dedicated Linux server. This massive increase in computational power gave Operational Performance an opportunity to build a powerful mechanism for the loading of data. An infrastructure rebuild was required; we exchanged the ETL codes for a suite of Python libraries that run our daily reloads. These now run automatically, four times per day, to ensure we are always working with the latest data. This infrastructure has become the backbone of data processing and allows Operational Performance to build and maintain data pipelines quickly and efficiently. A new pipeline between databases can be built in a few minutes and incorporates robust and reliable functionality with standardized error handling. This allows our team to spend each day executing reporting and analysis and developing new infrastructure instead of loading data.

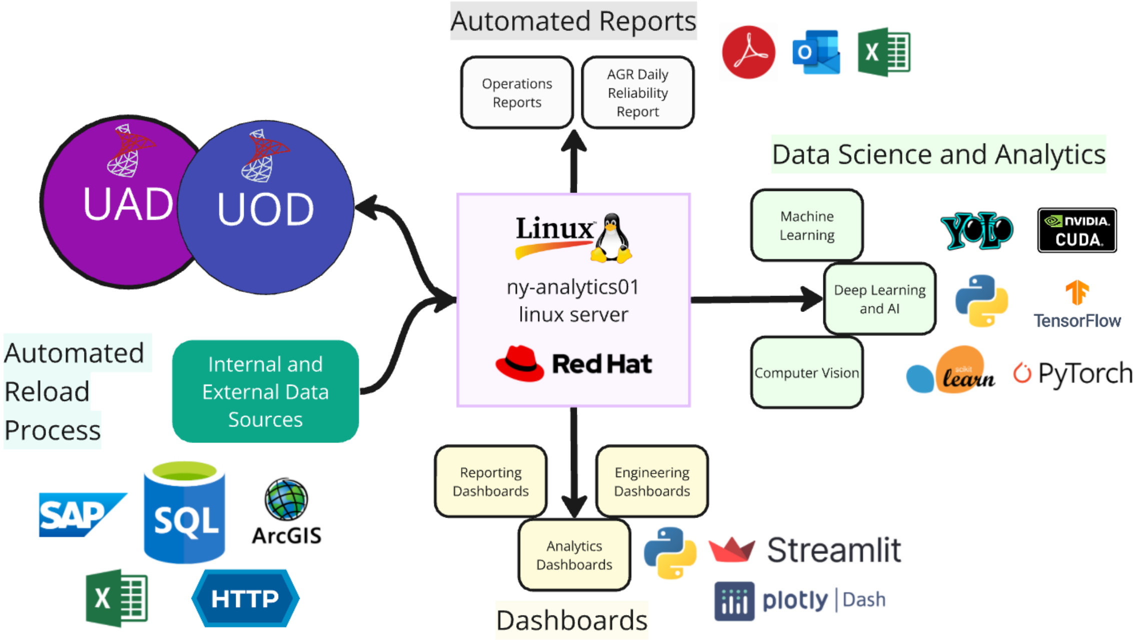


Figure 4: *The Linux server now powers all data moving from upstream databases into our SQL server environment, hosts dashboards used for reporting, runs data science models and codes, and generates daily automated reports.*

3.5. UPD, UFD, UGD

Three more specialized SQL Server databases were added in 2023 and 2024.

UPD – Unified Program Database

The Unified Program Database began as a replacement for disparate Excel spreadsheets that were used to house and manage program/project data.

UFD – Unified Forecasting Database

The Unified Forecasting Database is a repository for load forecasting data. The UFD includes data on electric vehicle registrations, electric vehicle charging, heat pump distribution, and customer load data to facilitate the next generation of load forecasting.

UGD – Unified Generation Database

The Unified Generation Database is used to house distributed generation data. Unified data from disparate systems allow for better understanding of distributed energy resources (DER) and the grid interconnection process.

In less than four years, the data science and analytics department has written more than 100,000 lines of code and manages 40+TB of data spread across 5 SQL server databases, a PostgreSQL database and a data storage server.

4. Cultural advantages of internal data science and analytics

Utilities do not often have a strong data culture that emphasizes best practices around data acquisition, storage, and management. Implementing a data science and analytics organization allows for the proliferation of knowledge and awareness of cross-industry methods that can greatly improve the utility's data operations.

Collaboration Across Corporate Functions – Though the Operational Performance organization is embedded within electric operations, the awareness of what is possible through the effective management and leveraging of data has produced a strong desire across other utility functions such as planning, customer service, and finance to adopt similar practices. Our data science and analytics department often participate as internal consultants on ways these other groups can implement data best practices to improve their own capabilities and open doorways for the new integration of additional datasets to operational analysis.

Promotion of Data Analytics – In addition to cross-organizational promotion of the benefits of data analytics, the Operational Performance organization has participated in external venues and publications to discuss how Avangrid has implemented these internal capabilities to improve programs, projects, and strategy. In addition to the positive reception of analytics in regulatory dockets and rate cases, we have presented and spoken at the Electric Power Research Institute (EPRI), the Northeast Power Coordinating Council (NPCC) and Google while being profiled by media groups such as S&P Global Commodity Insights, Latitude Media, and local news outlets. This external promotion has helped drive industry interest and internal excitement about innovative ways to approach grid operations.

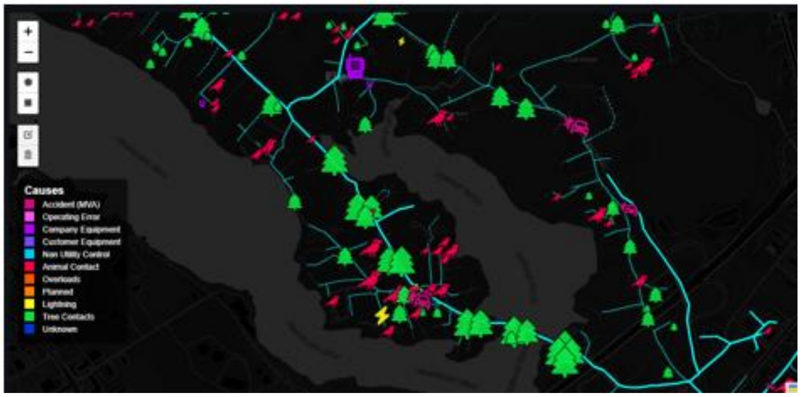


Figure 5: *Outage Location Mapping*

Deeper Functional Understanding - Data building blocks must be found, fully understood then integrated with other existing data to build a unified data science platform. The build process will be much longer and much more iterative than a single project or a single engagement but will yield significant advances in the understanding of core datasets. This knowledge will not be lost upon the completion of a project as it would with outsourced resources.

Cultural Awareness – Having data science and analytical professionals embedded in electric operations allows for increased awareness and understanding of what is possible through data. Leaders that have traditionally been in positions where poor data availability has required relying on subjective insights may not realize the potential efficacy of data-driven strategies. The immediate availability of internal data science and analytic resources to leadership, executing analytics daily that helps drive operational insights, allows for the rapid development of a data culture.

Alignment with corporate goals – The immediate result of a more robust data processing platform was that the reporting of key system metrics (SAIFI, SAIDI, CAIDI[2]) improved in terms of availability, frequency, accuracy, and reproducibility. These metrics are at the forefront of utility operations, and through our work, management and leadership can continuously analyze, track, and report at levels of increasing granularity for reliability and resiliency trends.

5. Major reliability projects

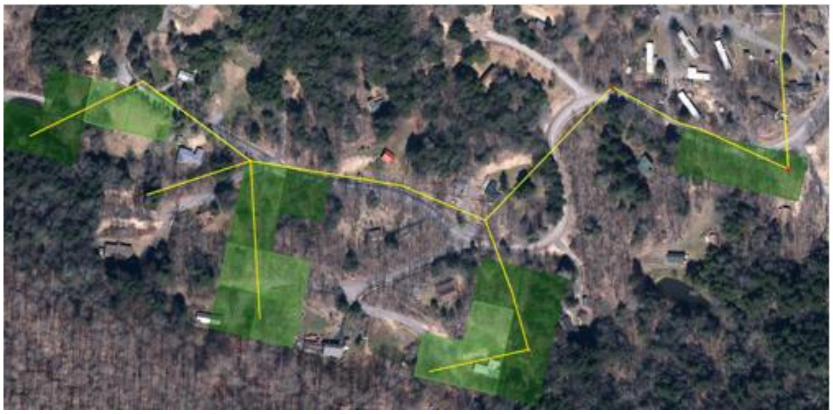


Figure 6: *Analysis of satellite imagery for vegetation prioritization.*

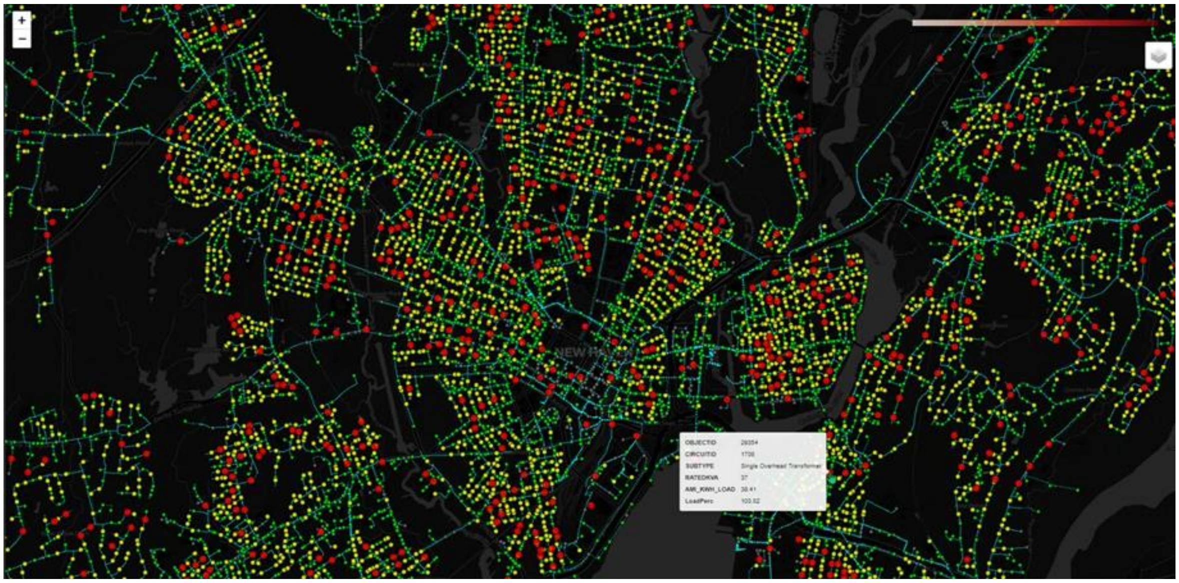
Vegetation Management – The primary focus of our reliability and resiliency analysis is on the effective management of tree trimming for the distribution network. In NYSEG's most recent rate case, data analytics quantifying the efficacy of the vegetation management programs drove the justification and eventual awarding of NYSEG's first fully funded cycle trim as well as a reinvestment of SAIFI penalties into reclaiming historically untrimmed circuits. A data-driven prioritization of the transition to a 6-year cycle that focused on impact per circuit mile allowed NYSEG to front-load highly impactful circuits for immediate customer benefit for reliability and resiliency. Data analytics on reliability and resiliency benefit also found success in CMP and UI's recent rate cases, and a similar prioritization was used for the implementation of CMP's new ground-to-sky trimming program. We are continuing to integrate new datasets such as tree species, tree decay, and canopy coverage to improve the ability to surgically plan and optimize our maintenance plans, as well as developing computer vision and other analytical methods for imagery evaluation.

[2] *SAIDI = System Average Interruption Duration Index. It is the minutes of non-momentary electric interruptions, per year, the average customer experienced.*
SAIFI = System Average Interruption Frequency Index. It is the number of non-momentary electric interruptions, per year, the average customer experienced.
CAIDI = Customer Average Interruption Duration Index. It is average number of minutes it takes to restore non-momentary electric interruptions.

Danger Tree Prioritization – The engineering team operationalized our vegetation management analytical work for the Danger Tree Program, which targets the removal of trees that are either dead or have a high risk of falling into the electric conductors from outside the right-of-way (ROW), to rank and prioritize more than 10,600 identified trees based on the quantified risk, customer exposure, and historical risk realized. Following the implementation of this prioritization, NYSEG and RG&E have reduced the costs of each tree removed to historic lows while maximizing the reliability and resiliency benefit for customers.

NYSEG SAIFI Target Change – In the most recent rate case for NYSEG, we developed a historical analysis using the original methodology from 2004 for setting regulatory targets. This allowed the company to successfully achieve a SAIFI target refresh based on annual customer growth rates.

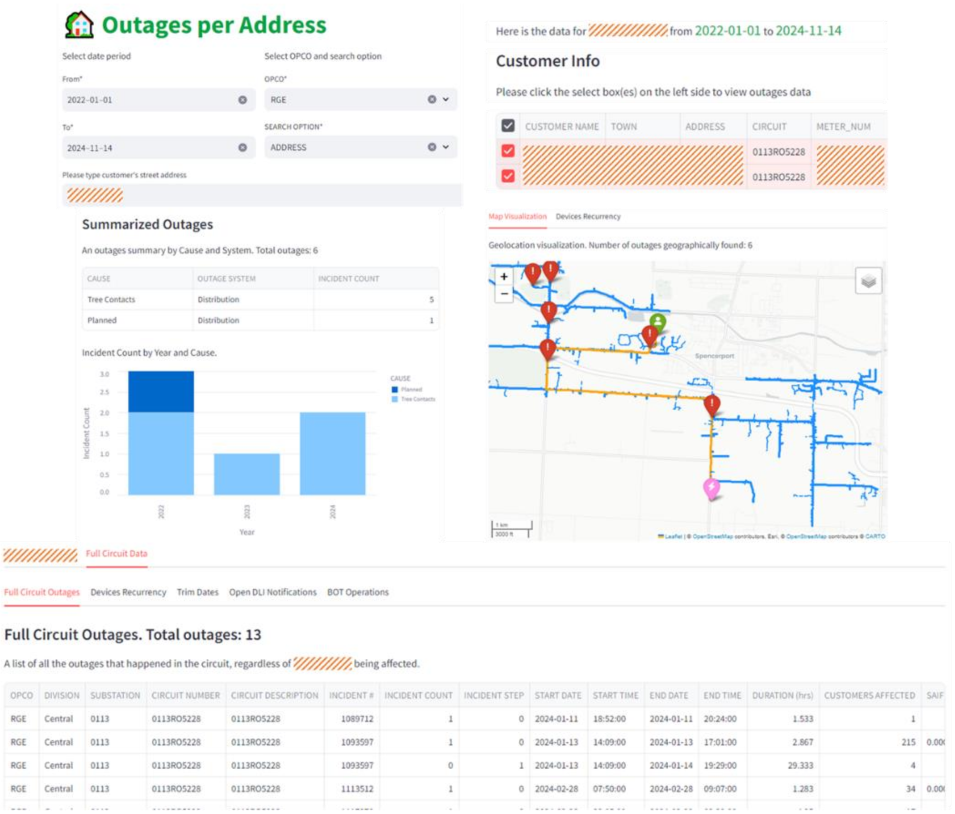
AMI Analytics – Utilizing AMI data and secondary transformer mapping at UI, we were able to map the secondary transformer load for all of UI and compare it to the transformer capacity and neighboring transformer capacity for potential load transfers to avoid overload failures. This analysis focused on a day of elevated temperature in July, with additional functionality planned upon the increased availability of full-year AMI load data.



OBJECTID	29354
CIRCUITID	1706
SUBTYPE	Single Overhead Transformer
RATEDKVA	37
AMI_KWH_LOAD	38.41
LoadPerc	103.82

Figure 7: *AMI data used to show distribution transformer loading in New Haven, CT.*

CEMI/CELID – Regulatory dockets in Connecticut requiring the reporting of CEMI (Customers Experiencing Multiple Outages) and CELID (Customers Experiencing Long Interruption Durations) necessitated a transformation of the underlying datasets. This analysis associated outages with individual customers as opposed to the traditional circuit and sub-circuit hierarchy. Utilizing the infrastructure that the data science and analytics team developed, our engineers were able to restructure the outage data across the utilities to calculate CEMI/CELID figures and develop front-end tools to look up customers by address to analyze the individual reliability and resiliency experience.



Outages per Address

Select date period

From*: 2022-01-01

To*: 2024-11-14

Select OPCO and search option

OPCO*: RGE

SEARCH OPTION*: ADDRESS

Please type customer's street address

[redacted]

Summarized Outages

An outages summary by Cause and System. Total outages: 6

CAUSE	OUTAGE SYSTEM	INCIDENT COUNT
Tree Contacts	Distribution	5
Planned	Distribution	1

Incident Count by Year and Cause.

Here is the data for [redacted] from 2022-01-01 to 2024-11-14

Customer Info

Please click the select box(es) on the left side to view outages data

☑	CUSTOMER NAME	TOWN	ADDRESS	CIRCUIT	METER_NUM
☑	[redacted]	[redacted]	[redacted]	0113RO5228	[redacted]
☑	[redacted]	[redacted]	[redacted]	0113RO5228	[redacted]

Geolocation visualization. Number of outages geographically found: 6

+
−
Spencerport
1 km
3000 ft
Leaflet | © OpenStreetMap contributors, Esri, © OpenStreetMap contributors © CARTO

[redacted] Full Circuit Data

Full Circuit Outages. Total outages: 13

A list of all the outages that happened in the circuit, regardless of [redacted] being affected.

OPCO	DIVISION	SUBSTATION	CIRCUIT NUMBER	CIRCUIT DESCRIPTION	INCIDENT #	INCIDENT COUNT	INCIDENT STEP	START DATE	START TIME	END DATE	END TIME	DURATION (hrs)	CUSTOMERS AFFECTED	SAIF
RGE	Central	0113	0113RO5228	0113RO5228	1089712	1	0	2024-01-11	18:52:00	2024-01-11	20:24:00	1.533	1	
RGE	Central	0113	0113RO5228	0113RO5228	1093597	1	0	2024-01-13	14:09:00	2024-01-13	17:01:00	2.867	215	0.00
RGE	Central	0113	0113RO5228	0113RO5228	1093597	0	1	2024-01-13	14:09:00	2024-01-14	19:29:00	29.333	4	
RGE	Central	0113	0113RO5228	0113RO5228	1113512	1	0	2024-02-28	07:50:00	2024-02-28	09:07:00	1.283	34	0.00

Figure 8: *A dashboard showing outages experienced for a selected customer location. Data is used to build CEMI/CELID analysis.*

TLI/DLI[3] Prioritization – the engineering team was able to optimize the reliability and resiliency impact of the Distribution and Transmission Line Deficiencies Inspection program, prioritizing more than 90,000 maintenance notifications generated by field inspections based on asset condition, risk of failure, historical reliability, and level of impact.

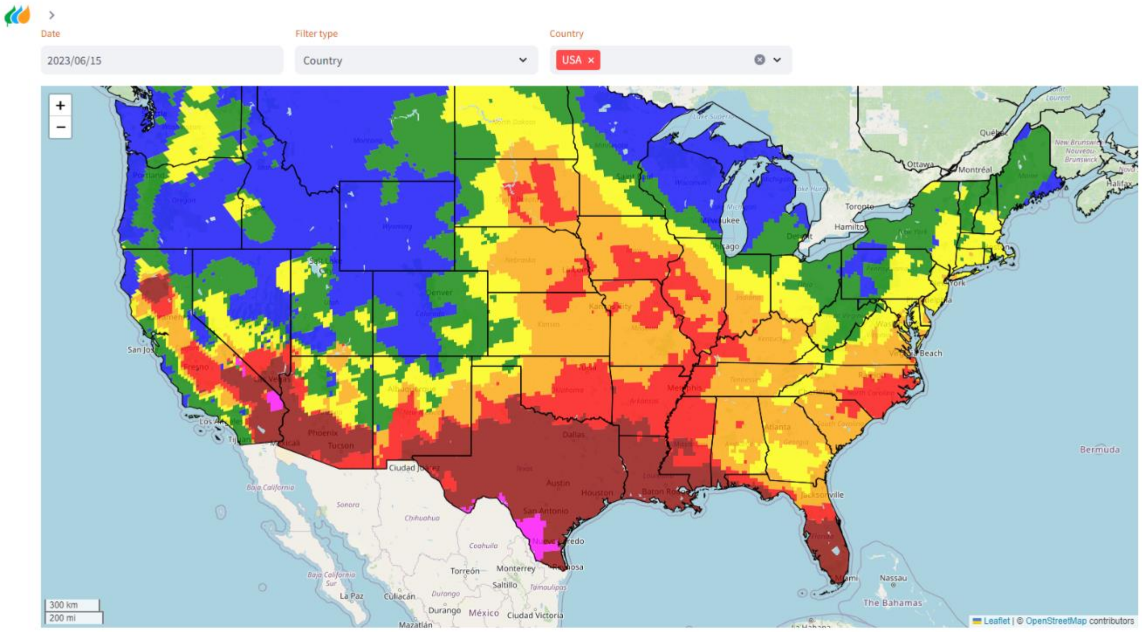
6. AI and data science

The advantage of the internal development of data infrastructure and deepening knowledge of core datasets is it allows for the utility to develop a focused vision of how best to bridge existing utility analytics into the future of advanced modeling and AI development. The Operational Performance organization is building and implementing these advanced technologies utilizing the same data science professionals that have been core to the utility's

[3] *DLI is Distribution Line Inspections, TLI is Transmission Line Inspections.*

digital transformation in areas such as geospatial weather modeling, computer vision, and machine learning.

Geospatial Weather Analytics – Weather is the primary external variable in reliability and resiliency performance, but often the geospatial aspect of weather impact and system risk is under-analyzed. The data science and analytics team pioneered a structure called GeoMesh which uses scalable polygons on a map to group datasets and do large-scale geographical analysis, providing a remarkable level of flexibility in weather analytics. This platform further allows for the iterative integration of novel datasets to drive new insights into optimized methods and strategies for reliability and resiliency investment that are tailored to historical weather trends and their impact on localized system vulnerabilities.



Date: 2023/06/15

Filter type: Country

Country: USA

Figure 9: *Temperature countours of the contiguous United States for a summer day in 2023 using the GeoMesh platform.*

HealthAI: Computer Vision for Distribution Assets– The data science and analytics team has developed our own computer vision technology to analyze millions of existing high-resolution photos of our street-level distribution assets. HealthAI has been trained to automatically identify assets in infrastructure imagery. The images are associated with particular assets in our GIS data using their location, then a holistic picture of the asset is formed. We integrate asset, location and maintenance data, into our backend PostgreSQL databases that are hosted on the Linux server. This information was used to develop a front-end "viewing" platform for our end-users to interface with the data. HealthAI will drive significant optimization to our maintenance programs for proactive asset replacement and reliability improvement.

Predictive Health Analytics (PHA) for Substation Equipment – PHA is data-driven approach to determining equipment's overall health and life expectancy based on numerous factors such as age, frequency of use, manufacturer and maintenance notifications. By integrating new operational data into these proactive asset replacement prioritizations, the company will be able to reduce large impact outages from substation equipment failure.

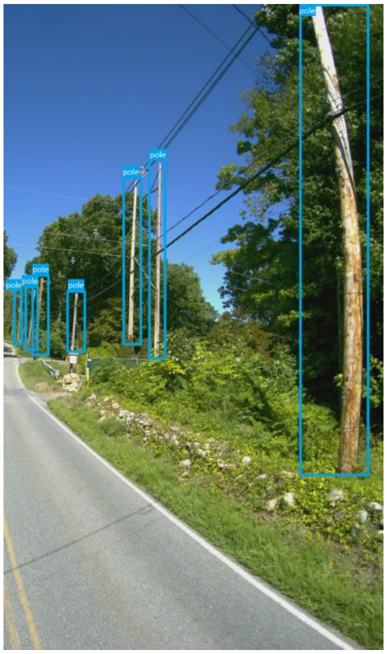


Figure 10: *These poles were labeled by the HealthAI model.*

7. Conclusion

Data science and analytics can be effectively leveraged by a utility to develop strategy and communicate the quantitative benefits of strategic decisions. Data-driven decision-making is a way to establish the vision of electric operations for effective management and investment in the grid. If utilities outsource the development of their foundational analytical infrastructure, they outsource the foundation of their own vision. This approach atrophies the core skillset necessary to develop flexible infrastructure to anticipate emerging problems and react immediately. Electric operations measures their deliverables in hours and days – a responsive analytical function is one that facilitates data availability to immediately respond to these requests with accurate and reproducible results.

The development of an internal data science and analytic organization and its accompanying infrastructure can be significantly more cost-effective for rate payers than vendor-driven use case development. It provides more tailored, highly effective results because it is embedded within the organization itself. Infrastructure does not need to be overly complicated or expensive. Simple on-prem database systems are often more than capable of accommodating effective digitalization with little cost in maintenance, development or access. This setup can accommodate organizational growth rapidly due to the widespread usage of programming languages such as SQL and Python across different industries. Training materials and programs for this type of skill development are ubiquitous due to their applicability and robustness, so employees can learn the skills they need on-the-fly. The style of database infrastructure, as we have defined it here, will quickly spread across organizational silos. If it is properly developed, it can unlock interconnections between previously isolated datasets.

The success of the Operational Performance organization at Avangrid is underscored by the central tenet that data science and analytics must be a core function of electric operations and the utility as a whole. By embedding data professionals into the organization who are focused on the development of flexible and agile data infrastructure, we were able, not only to anticipate and accommodate the daily needs of our organization, but also use and leverage data in unforeseen ways. Operational Performance executes advanced analytics for reliability and resiliency improvement, develops prioritizations for operational implementation, produces daily executive and external reporting, and provides successful, data-driven rate case justifications. OP, after only four years, is also able to develop more advanced AI models and data science platforms. The Operational Performance organization, at its core, is a bridge to the next phase of Avangrid's data-driven future by focusing on investment into its own people.