Unlocking Efficiency and Innovation: The Crucial Role of Data Governance and a Modern Data Platform in Orion's AI adoption

Peter Jaksons, Akin Atanda, Alika Ackroyd

Orion group NZ

1. Introduction

Over the last year, Orion has embarked on a journey to adopt artificial intelligence (AI). The rise of AI offers opportunities for optimising and empowering energy distribution operations and drive operational efficiencies across the organisation. Successful adoption of AI depends on data that is well governed, clean, relevant; and is supported by a modern data platform that can handle big data with speed and scalability. Figure one shows the interdependence between data governance, a modern data platform and analytics and AI. This diagram explains that data governance is the overarching framework that applies to all data sources (at Orion these include our asset management data sources, operational data, finance, etc), the centralised data and the end users of data products (such as other application, people and smart devices). The modern data platform is a cloud-based data environment with scalable computing power, which enables people to handle large amounts of data for reporting and analytics more easily and at scale.
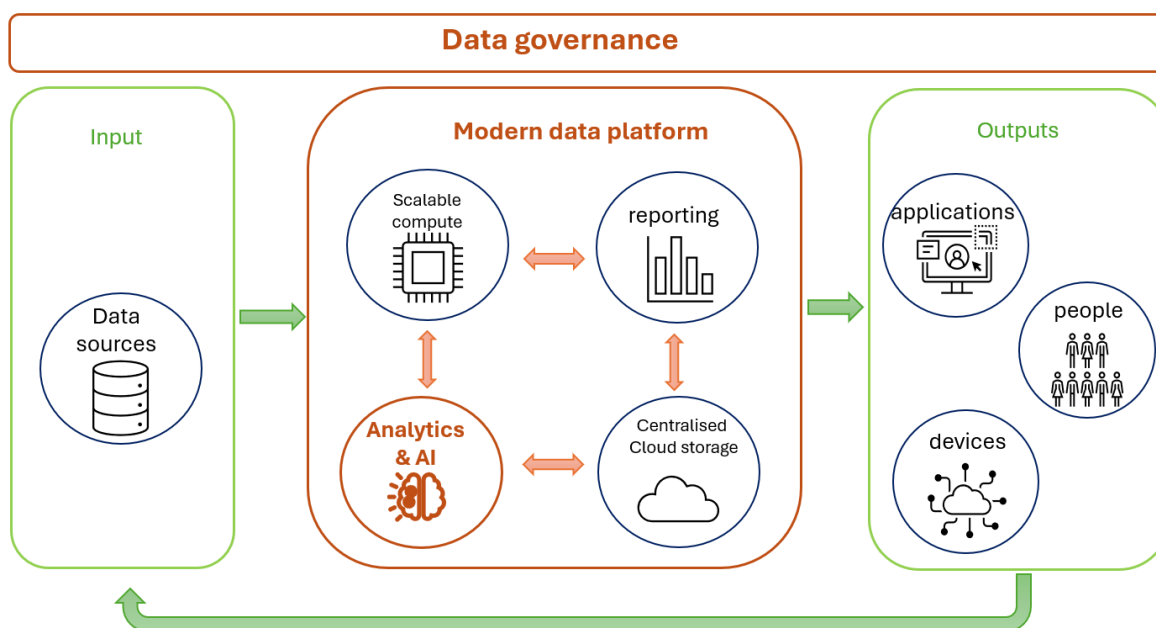

Figure 1: High level relationship between data governance, a modern data platform and analytics and AI.

This paper aims to show how Orion has used its data governance and modern data platform, explained in sections 2 and 3 respectively, to improve its efficiency, innovation, and integration. We will present examples of how our data governance programme has created a trustworthy, high-quality, and digitally enabled data environment for our organisation, and how our modern data platform has supported the processing and analysis of big data and fostered interoperability, which has enhanced our internal and external collaboration and innovation. These achievements, detailed in section 4, have helped us to advance our artificial intelligence (AI) solutions at Orion.

2. Data Goverance at Orion

Orion has access to vast amounts of data from smart meters, LV monitors and other IoT devices as well as many corporate and operational data sources. To take maximum advantage of our investments in data husbandry and to accelerate the digitalisation of Orion we have started implementing a data governance framework. As a result, over the last two years we have implemented clear policies, standards, and procedures to enable easy access to data while being able to audit usage and comply with privacy, security, and regulatory frameworks. Moreover, our data governance framework aims to foster a culture of data literacy, accountability, and collaboration among our employees and stakeholders, ensuring that data is used ethically, responsibly, and effectively for achieving our strategic goals and delivering value to our customers and partners.

Orion's data governance framework is loosely based on the principles outlined in the DAMA International Data Management Body of Knowledge. Our framework covers five key areas:

1. Metadata Governance: This ensures that we have clear and consistent definitions, descriptions, and standards for our data assets. We have started initiatives around data cataloguing, meta data tagging and are assigning data ownership to business owners.

2. Data Quality: This principle emphasises defining high-quality data, formulating a data quality strategy, identifying critical data and business rules, conducting initial data quality assessments, prioritising potential improvements, setting goals for data quality improvement, and developing and deploying data quality operations. At Orion many key data sources such as our asset data and GIS data have or are going through a data quality review which includes improving the data quality and the underlying structure of the data.

3. Leadership Commitment: This requires the active involvement and support of our senior leaders in driving data governance initiatives. We have established a data governance steering group that oversees the data strategy, policies, and standards, and allocates resources and responsibilities for data management. We also ensure that our data governance approach is aligned with our business goals and values.

4. Data Lifecycle and Lineage: Understanding the lifecycle and lineage of data is vital for managing it effectively. Data is a commodity that is created, used, transformed, and reused throughout an organisation, necessitating cross-functional and diverse skills for data management. At Orion as part of the roll out of our modern data platform we are creating 'gold layer' data assets. These 'gold layer' data assets are high quality, readily available data sources, that can be used for reporting, analytics or to feed into other applications. The

creation of these assets is under source and version control with visible lineage to ensure trust in these assets.

5. Sustainable Data Governance: This means that we continuously monitor and improve our data governance practices and outcomes. We do not see data governance as a one-off project, but as a long-term process that requires ongoing commitment and ownership from our business leaders and stakeholders. We align our data governance strategy with our business goals and values, and we measure and communicate our progress and impact. We also foster a data-driven culture and empower our data users with the skills and tools they need.

Data Governance mostly focusses on individuals and workflows, typically enhancing data projects incrementally, often without sudden large changes being obvious. Still, we'll outline two data governance endeavours that have had a noticeable impact to our ways of working at Orion:

1. **PowerBI data governance to increase trust and security.** One of the ways we have enhanced our data governance is by applying it to our PowerBI reporting framework, which has resulted in increased trust, higher security and faster uptake of our reports. To increase trust, we have implemented a peer review process to certify our datasets, data pipelines and reports, ensuring that they meet our quality standards and best practices. To increase security, we have visualised the access and sharing of our reports and data, enabling us to audit and control who can view or edit our data assets. Figure 2 for example illustrates how we track the access of data sets available to staff in PowerBI. The graphs show the distribution of access across different workspaces and datasets, while the table provides the same information in a more detailed way. This information helps us to quickly verify whether any reports or data are being shared or accessed by people outside the organisation or from unauthorised areas within the company. Finally, to increase uptake, we have tagged metadata and catalogued our reports and datasets, making it easier for our employees to find and use relevant and reliable data sources for self-service reporting. Figure 2 illustrates how we track the access of data sets available to staff in PowerBI.
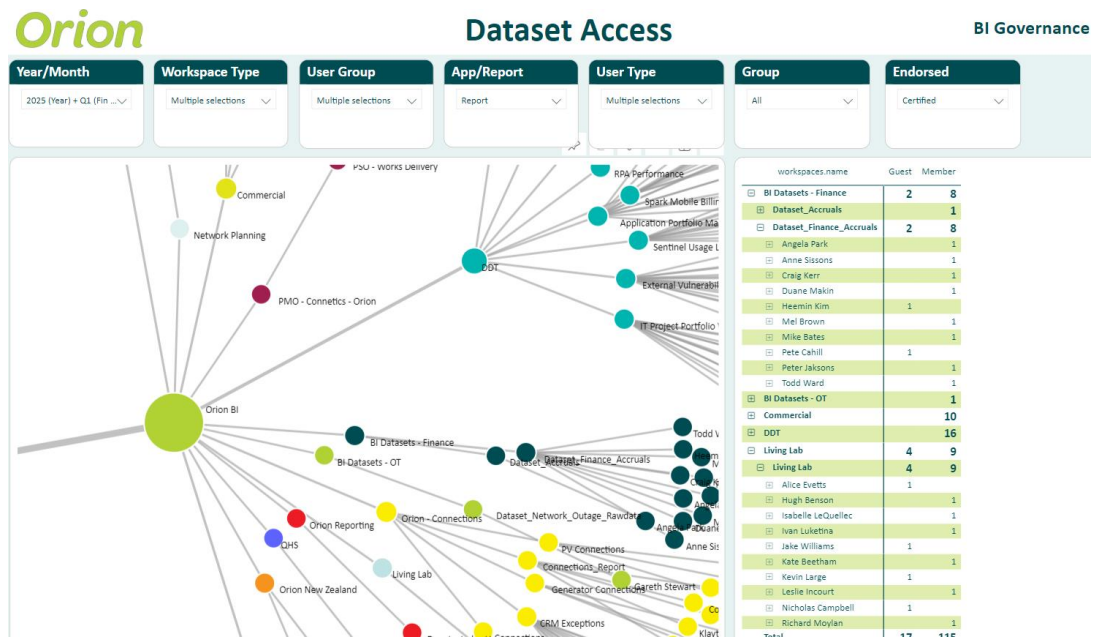

Figure 2: Example of our PowerBI data governance framework.

2. **Network data visibility:** Orion uses various devices, such as smart meters, sensors, and fault indicators, to collect data from its high voltage (HV) network down to the individual connection point (ICP) level. This data collection enables Orion to monitor the condition and performance of its assets, improve its operational efficiencies by optimising maintenance and outage management, and plan for future network investments and upgrades. However, until recently, Orion did not have a clear picture of how much data it had access to and how complete and reliable it was. By improving the data quality and cataloguing of different data assets, Orion has been able to integrate and analyse all the data sources in a unified platform and enhance its understanding of its current network visibility rate. It has also been able to identify where there are gaps in data coverage and prioritise areas for improvement. Figure 3 shows a visual representation of Orion's network data visibility, which indicates the percentage of network components that have data available for analysis at different levels of the network hierarchy.
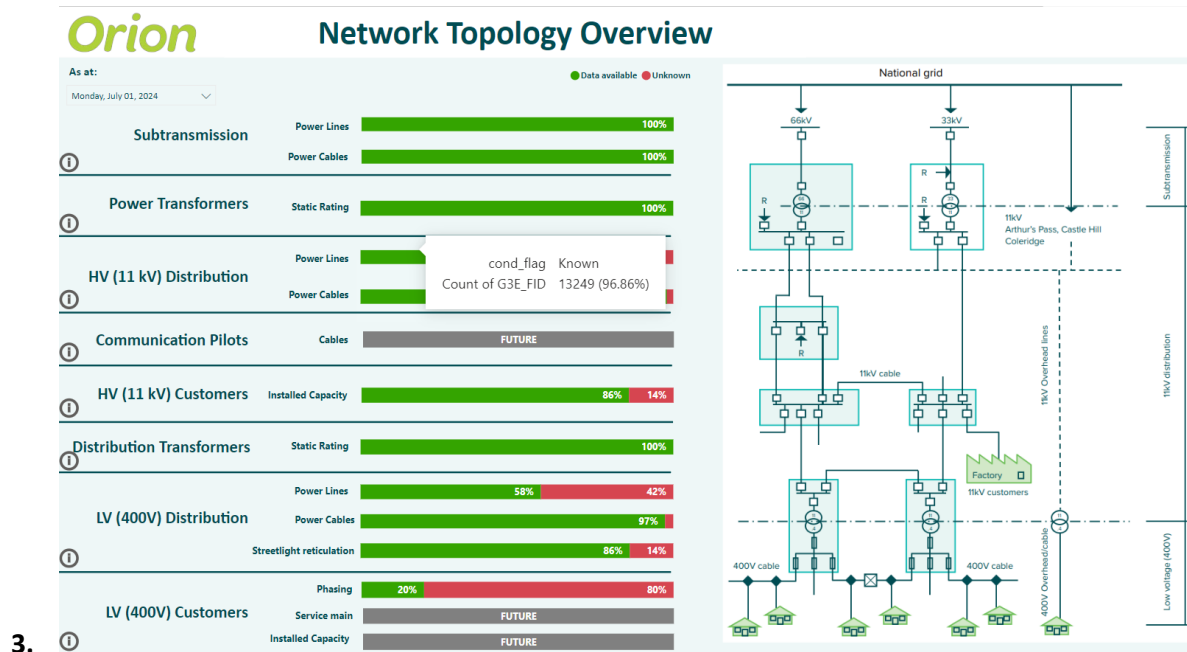

3.

Figure 3: Visual representation of our network data visibility from HV down to ICP level.

3. Modern data platform

To increase its network visibility and network digitalisation, Orion has been investing in datasets that capture various aspects of its electricity distribution network. Some of these datasets are of very high volume, such as the smart metering data, which collects 5-minute interval readings for approximately 185,000 ICPs, and the LV monitoring datasets, which collects 10-minute interval readings for approximately 1,000 devices. The size and frequency of these and other datasets required Orion to invest in appropriate IT infrastructure that can handle large volumes of data efficiently and at scale: a modern data platform. The modern data platform that Orion has invested in is a strategic asset that can help the organisation leverage data for insights, innovation, and competitive advantage. It is designed to handle large volumes of data from various sources, such as SCADA, AMI, GIS, CRM, and IoT, and supports a wide range of analytics and machine learning

applications, such as network planning, asset management, outage detection, demand response, and customer segmentation.

The platform offers several key benefits that align with Orion's data strategy and goals, such as:

1. Future-Proof Architecture: The platform is designed to be robust and flexible, able to handle current and future data needs. It supports various data sources and formats, both structured and unstructured, and enables seamless integration with other cloud services and applications.

2. Unified Data Management: The platform provides a comprehensive solution for the entire data lifecycle, from ingestion to actionable insights. It allows for data preparation, transformation, exploration, and visualisation, as well as the development and deployment of AI and ML applications, all within a secure and integrated environment.

3. Data-Driven Decision Making: The platform empowers Orion to become a data-driven organisation by providing tools and capabilities that facilitate better decision-making. It enables analytics across multiple data types and sources, both internal and external, allowing for a more holistic view of the data landscape and the generation of adavanced insights.

4. Scalable Compute and Storage: The platform can handle large and complex data sets, such as smart metering data and IoT data, with ease and efficiency. It leverages the cloud's scalability and elasticity to accommodate growing data volumes and demands, ensuring that the data infrastructure grows with the business.

5. Advanced Analytical Capabilities: The platform enables Orion to establish and enhance its analytical competencies in areas such as AI and ML, which can provide more sophisticated and predictive data analysis. It also supports a variety of languages and frameworks, such as Python, R, Scala, SQL, and TensorFlow, allowing for flexibility and innovation in data analytics.

6. Data Governance and Security: The platform comes with built-in governance and security features that help manage data access, encryption, and masking, ensuring compliance with various regulations and standards. It also provides audit logs, monitoring, and alerts, enabling visibility and control over the data activities and performance.

Figure 4 illustrates Orion's data platform architecture, which has four main layers:

- **Data Sources:** This layer collects and ingests various types of data, such as smart meter data, network data, SCADA data, customer data, and external data sources. The data can have different formats and protocols, such as CSV, JSON, XML.
- **Data Ingestion and Storage:** This layer ingests and stores the data in real-time or batch mode, using Azure Data Factory and Azure Blob Storage. Azure Blob Storage offers a cost-effective, scalable, and durable storage solution for large volumes of data.
- **Data Processing and Transformation:** This layer processes and transforms the data according to the business rules and logic, using Azure Databricks and Delta Lake. Delta Lake provides a reliable and high-performance storage layer for big data, with ACID transactions and schema enforcement.
- **Data Consumption and Visualisation:** This layer consumes and visualises the data by various users and applications, using Power BI, DataBricks and APIs for other application. Power BI provides interactive dashboards and reports. APIs provide a standard and secure way of data exchange with external applications and systems.
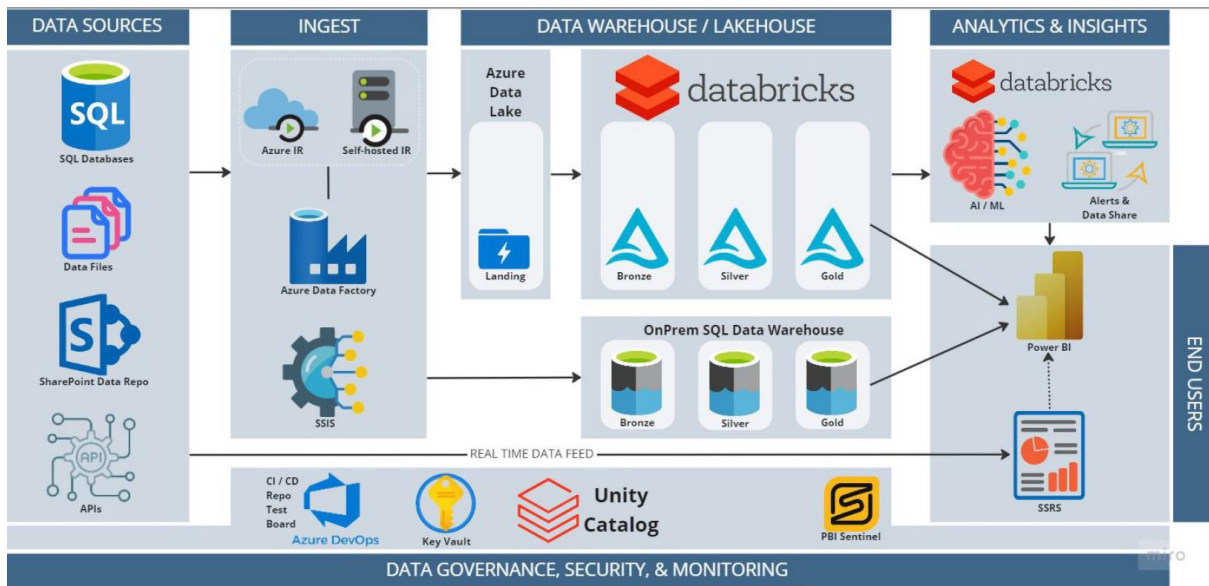
Figure 4: High level architecture of Orion's modern data platform.

We can showcase the functionally and success of Orion's investment in a modern data platform by the following two examples.

1. **Uptake of our data platform:** The success of our modern data platform can be illustrated by our platforms utilising metrics dashboard, as shown in figure five. Over the last 16 months we have created 65 schemas for different functional areas across the business, added 999 tables, onboarded 20 active users and have more than 25 billions observations processed and stored. Note that some schemas and tables are coming from new data sources such as smart meter data or Low Voltage monitoring, while most schemas and tables are ingested from existing storages systems into our data platform to provide centralised governance and access to data.
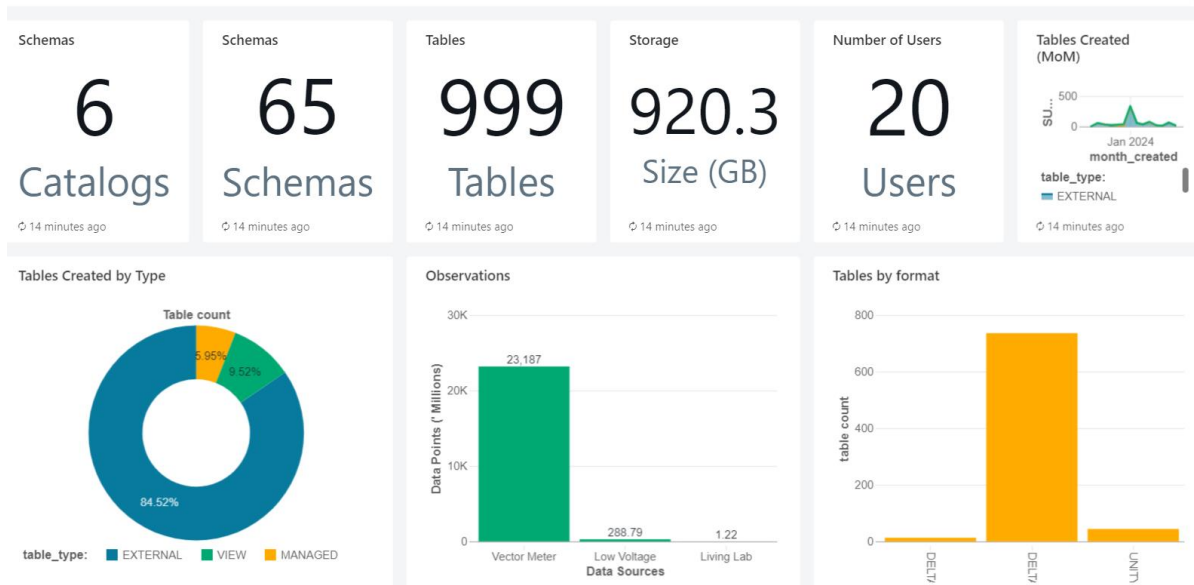

Figure 5: Dashboard of Orion's modern data platform uptake and utilisation.

2. **Processing speed and ease of use:** One of the benefits of our platform is its speed and usability, as shown in fig 6. The figure displays the average real power (kW) of all residential (RES) customers for whom we have smart metering data in the greater Christchurch region, by month and day of week (in this case we show the data for January). We can also segment the market by other categories, such as: Businesses, Irrigation, Telecommunication, Lighting, Priority Customers, and Customers with Generation capabilities. Using large volumes of smart metering data and the scalable computing power of our data platform, which can aggregate billions of rows daily, we can produce these graphs on demand often with higher accuracy compared with some of the predictive models. Our goal is to combine our knowledge of the network's composition based on customer segments and generalised load profiles for each segment, to create a more precise distribution power flow model. This would enable us to have a more realistic view of the network and make better decisions and more informed future planning.
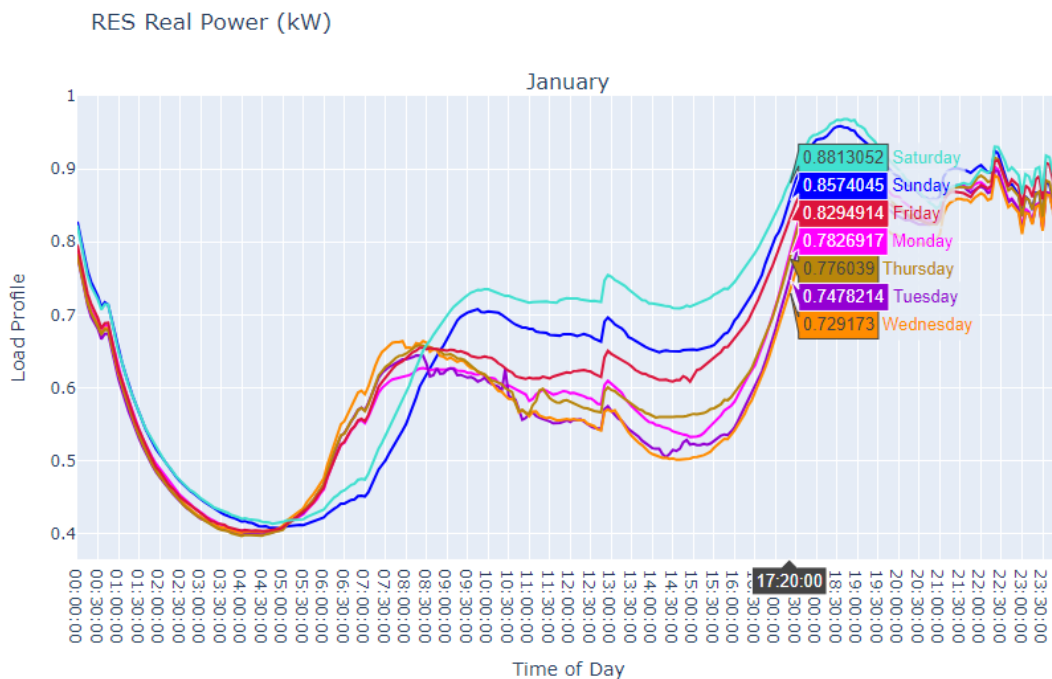

Figure 6: Aggregated load profiles by segments. In this case the segment is day of the week.

4. AI and Analytics:

Our commitment to data governance and the implementation of a state-of-the-art modern data platform have given us the tools and framework to create and develop AI and analytical pipelines more rapidly. Data governance provides assurance that our data adheres to high standards of quality, reliability, and compliance with regulatory demands. Our advanced data platform facilitates the integration, processing, and retrieval of data from a variety of sources and structures, empowering us to us sophisticated methodologies at scale. As figure 1 shows, the synergy between these two elements enables faster development of AI and analytical projects.

In the following sections, we will present two examples of how we leveraged our data governance and our data platform to create AI solutions. These examples showcase the value of our investments in data goverance and our data platform, as they were developed in less than a month:

1. **Customer segmentation:** In figure six, we aggregated the data numerically based on known customer segments (days and months). Using unsupervised machine learning methods, it is possible to cluster groups of households or business together based on their consumption profile. Figure seven shows the results of K-means clustering of average residential load profiles in winter. In this case the model was asked to partition the population of households into three clusters. The right panel shows the estimated load profile per cluster, where cluster two can be interpreted as the group of households with high energy demand (green line), and cluster 0 as the group of households with lower energy demand (blue line). The left panel gives a spatial representation of the same clustering process. Unsupervised machine learning methods are notoriously known for being both memory and computationally hungry algorithms. Running these models, even on very modern desktops, would not have been possible. Our modern data platform allows us to build and finetune these models more rapidly and at a relatively low cost. In this case rerunning the model takes less than half an hour.
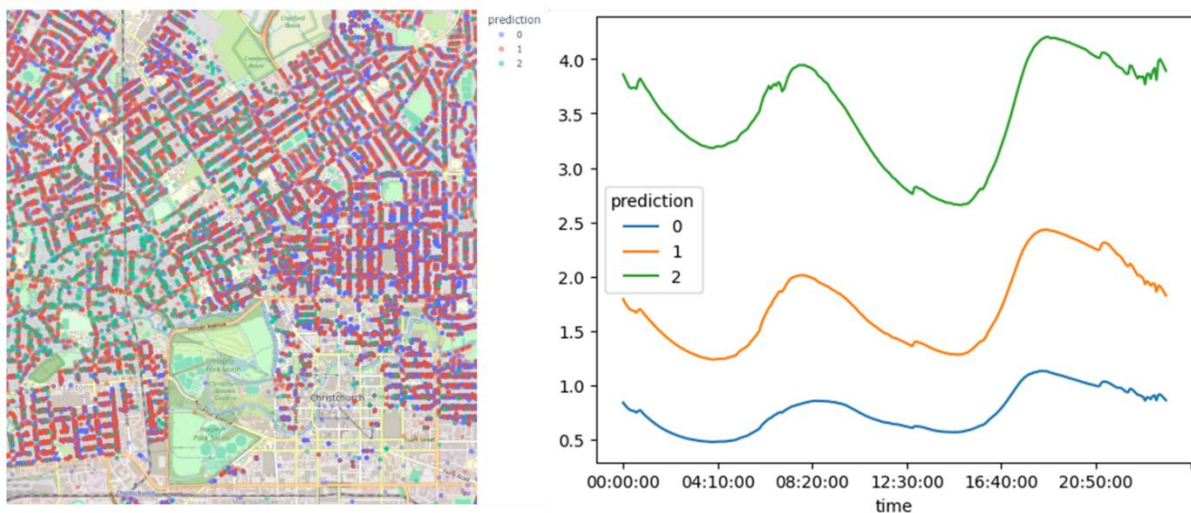

Figure 7: Unsupervised clustering of households. In this case the number of potential clusters is three. Right: Estimated load profile per cluster. Left: Spatial visualisation per cluster.

2. **Automated pole inspection:** As a proof-of-concept we have developed an AI tool that can automate the inspection and maintenance of power poles based on images. The tool enables users to upload images and the model gives feedback on the type of pole, lean angle, a maintenance-risk score, and additional comments with possible action points. This tool can lower the chances of failures and outages, enhancing the safety and efficiency of our operations and field workers. Figure eight shows an example of this AI pole inspection tool. In this case, the model scans the image and informing the user that it is a wooden pole with a 60-degree angle lean with a broken line. It assigns a maximum severity rating of five (min 0 and max 5), as it detects that the pole is near a busy road which requires the user to immediately investigate the situation, resulting in immediate intervention. This proof of concept demonstrates the potential and the progress Orion has made in developing AI applications. We aim to implement many of these applications into production environments soon.
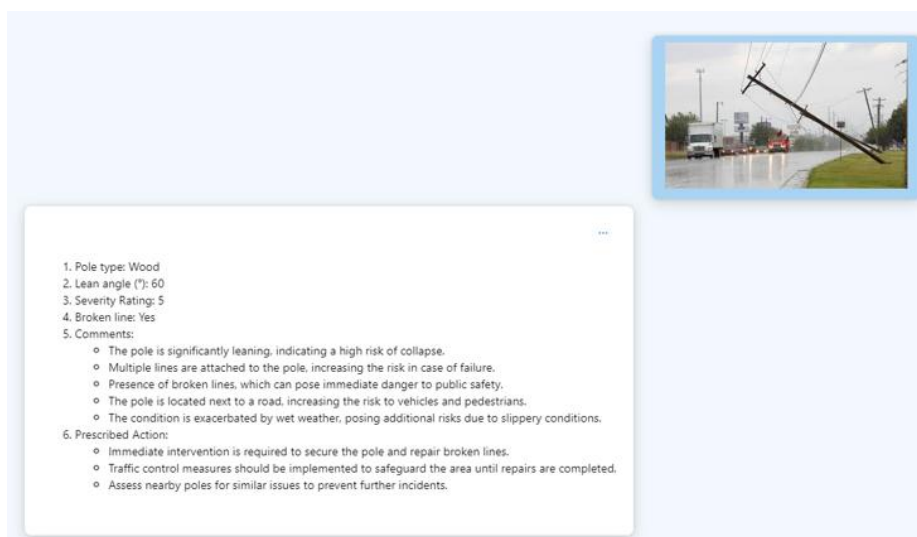
1. Pole type: Wood
2. Lean angle (°): 60
3. Severity Rating: 5
4. Broken line: Yes
5. Comments:
   ○ The pole is significantly leaning, indicating a high risk of collapse.
   ○ Multiple lines are attached to the pole, increasing the risk in case of failure.
   ○ Presence of broken lines, which can pose immediate danger to public safety.
   ○ The pole is located next to a road, increasing the risk to vehicles and pedestrians.
   ○ The condition is exacerbated by wet weather, posing additional risks due to slippery conditions.
6. Prescribed Action:
   ○ Immediate intervention is required to secure the pole and repair broken lines.
   ○ Traffic control measures should be implemented to safeguard the area until repairs are completed.
   ○ Assess nearby poles for similar issues to prevent further incidents.

Figure 8: POC screen of pole maintenance and inspection AI tool

5. Discussion

We are proud to demonstrate how our AI applications can enhance our business outcomes and deliver value to our customers. But we also want to emphasise that these applications are not possible without good data governance and a modern data platform that supports them. To become AI ready, organisations need to invest in both the quality of their data, good data governance and the appropriate IT infrastructure, in our case a modern data platform. Below, we share some of the key lessons we learned from Orion's journey towards AI readiness:

- Data governance requires constant attention and effort, but it pays off in the long run. It also needs strong support from senior leadership.
- Setting up a modern data platform is a big project that needs proper resourcing. It is vital to design and build the platform well, but also to optimise the data modelling after the platform is working. Do not just lift and shift the existing data into the new platform.
- Cloud-based platforms may not always save costs, but they offer more functionality and less maintenance. Many AI and machine learning models these days are ready to use out-of-the-box. Instead, the main challenges for developing AI tools and machine learning models are ensuring high-quality and well-governed data and having scalable computing power.

Special consideration should be given to the fact that AI applications are not magic solutions that can solve any problem without human input. They require careful design, testing, and validation to ensure they are fit for purpose and do not cause unintended consequences. Human oversight is essential to monitor the performance, accuracy, and ethics of AI tools and machine learning models, and to intervene when necessary. AI should augment human capabilities, not replace them.

Orion's journey to adapt AI has only just started but the future looks promising. Over the coming year the emphasis will be on further development of AI tools and on how to production these applications.