Capgemini engineering

HOW UTILITIES CAN MOVE TO INSIGHT-DRIVEN OPERATIONS

USING DATA AND AI TO PREDICT DEMAND AND MANAGE ASSETS IN A FAST-CHANGING WORLD

CONTENTS

INTRODUCTION

Utilities deliver electricity, gas, and water to homes and businesses. To do so optimally, they need to understand changing demand in order to balance supply, and to deploy and manage a complex network of infrastructure to generate, transmit and deliver these services.

Insights derived from data make both demand forecasting and asset management more predictable. This is true even in stable times, but as climate targets, low carbon technologies and regulatory pressures fundamentally change how utilities operate, data becomes all the more important in navigating this far more complex, distributed, and nuanced world.

Electricity in particular faces huge external pressures to change thanks to new rules and technologies (see box). To thrive it will need more data driven insights into changing demand (eg from EV charging) and integrating new assets (solar, wind, batteries) into existing infrastructure. At the same time, they can use sophisticated data techniques to optimise their existing infrastructure – despite the fact most was not designed for the digital world.

Whilst slightly more insulated from immediate global disruption, water and gas can also benefit from the same approaches to gathering data-driven insights, reducing cost, improving service, and meeting changing regulations.

Three ways electricity is being disrupted:

- Production: by diverse energy sources, and digital energy management and trading systems
- Transmission: by microgrids and energy storage
- Consumption: by in home energy generation and storage, and the electrification of heating and transport

The data to do all this is increasingly available. Smart meters provide granular insights that can predict energy and water consumption. Connected Industrial Internet of Things (IIoT) sensors and actuators throughout generation, distribution and storage can give us detailed overviews and control of our assets. Diverse data sources, such as weather, environmental context, and consumer behaviour, as well as engineering data on asset design, can add layers of insight.

In theory, this should allow us to digitalize energy systems and move to informed data driven insights and decisions every step of the way.

But data does not automatically yield transformational insights. These are complex data sets reflecting physical and human systems, often incomplete, and from multiple sources. This data needs hard work to capture, combine and analyse to truly understand what it is (and isn’t) telling you.
If we get it wrong, the lights go out or the water stops. But if we get it right, we can build new agile data-driven utilities ready to deliver the promise of personalized, distributed, sustainable services.

Much is possible already. Utilities that act now can realize some significant quick wins in cost saving and consumer benefits. In doing so, they will gain learnings that set them ahead of the game in their journey towards becoming data-driven organizations.

As the business and regulatory landscape changes, data will allow utilities to profit from new opportunities. This whitepaper will discuss those opportunities and how to think about delivering short term value and long term transformation.

WHO IS THIS WHITEPAPER FOR?

This whitepaper will help utilities globally, including companies involved in electricity generation and distribution, and the supply of gas and water. In particular:

- Asset owners, Network Planning and Operations teams responsible for designing, building, running, maintaining and decommissioning water, gas and electricity infrastructure
- Demand management and pricing teams, as well as those involved in tackling fuel and water poverty
- Strategic decision makers looking for an overview of data-driven opportunities and challenges

Through examples across different utility sectors, we will highlight the value data can bring across the board as well as the commonalities in approaches to data-driven insights which get us there.

WHY ACT NOW

- Understand a future with different energy needs – ensure the water runs and lights don't go out
- Optimize the asset design, use and maintenance lifecycle, reducing OPEX costs and increasing profits
- Respond to changing regulatory pressures and energy data modernization programmes[1] to ensure criteria in national energy digitalization plans are met and exceeded
- Win public grants and tenders for innovative data-driven projects
- Keep up with customer expectations and avoid being outcompeted

1 Such as Ofgem and Innovate UK 's Modernising Energy data initiative and Ofwat's AMP7 in the UK

SECTION 1 WHAT ARE THE INSIGHT-DRIVEN BUSINESS OPPORTUNITIES FOR UTILITIES?

In this section we look at illustrative business challenges of where data can deliver value in (i) energy distribution management and (ii) asset management. Subsequent sections will explore the data challenges that need to be overcome to deliver this business value.

1. Energy Distribution Management

Electrification of transport and heating, alongside new energy sources, storage, and home generation, will change how we predict supply and demand and balance networks.

Demand-modelling using smart meter data insights
Smart meters are the obvious disruptor of energy management, offering lots of valuable insight – when people are home, what appliances they are using, whether they own an EV. If we can untangle this complex data, whilst respecting privacy, we can build detailed pictures of energy use to predict demand at a granular, personalized level.

This can inform energy storage and release needs based on probable demand and remove guesswork from network upgrade and reinforcement planning. It can also be combined with other data sets, such as local development plans to guide network upgrades.

Predicting sudden, local changes in demand
Data can help utilities make targeted investments in infrastructure, manage how they push energy to substations, and match generation with demand.

DATA CAN HELP UTILITIES MAKE TARGETED INVESTMENTS IN INFRASTRUCTURE

Large group behaviour shows up in statistics; we know people in the UK make tea during ad breaks. But local effects – down to individual home level – are harder to predict. An investment in charging infrastructure, or a vocal resident, could provoke a sudden local EV uptake, and a rapid change in electricity usage.

Smart meters and network data can allow us to extrapolate likely future behaviours with the right models. Western Power Distribution is working with

Electralink to analyse data which allows them to monitor the effect of electric vehicles and low carbon technologies on its local, Low Voltage (LV) network[2].

Capgemini Engineering worked with a major water utility to develop predictive models for water demand, using pressure drop as a function of flow to develop systems for precisely controlling water pressure across the network to minimise leaks.

We can be more sophisticated still by building this data into agent based models – ie making behavioural projections based on similar people – which can be particularly valuable for spotting the start of a trend in changing demand.

Modelling specific energy users
Large energy users, such as factories, are well understood. But there is little data on smaller scale users – such as bakers or hairdressers. Understanding the units that make up an energy system can create more granular modelling, and help predict load management and inform demand side responses.

2 https://www.electralink.co.uk/wp-content/uploads/2019/11/WPD_ElectraLink_IBM_VM_Data_Project_PR_press_version.pdf

UNDERSTANDING THE UNITS THAT MAKE UP AN ENERGY SYSTEM CAN CREATE MORE GRANULAR MODELLING

An interesting example of this approach was a client project that involved deploying sensors in a sample of small businesses, eg on fridges and ovens in bakeries, to develop models of typical energy usage patterns and propose energy efficiency measures. These local models can then be used to build more complex models of demand across different types of users. This is known as federated learning - learning locally and aggregating the learnings.

2. Asset and Infrastructure Management

Data can predict performance of infrastructure, from pipes and cables to reactors and substations, allowing predictive maintenance and optimization to extend lifetime and reduce costs.

Predictive maintenance and lifecycle management
By collecting data on performance, as well as contextual information such as meteorological conditions, sub-surface environment, and exposure ground movements and traffic vibrations, we can build models that detect signatures of degradation or impending faults. China shows us one of the most sophisticated examples. Its State Grid uses AI for fault detection in facilities, power grid demand forecasting to identify type and severity of grid breakdowns, and computer vision to identify defects in transmission lines[3].

THERE IS VAST POTENTIAL TO USE MORE SOPHISTICATED DATA TECHNIQUES TO DO PREDICTIVE MAINTENANCE ON INFRASTRUCTURE ALREADY IN THE GROUND

Around the world, this is making headway on above-ground assets, which are easy to augment with sensors. But there is vast potential to use more sophisticated data techniques to do predictive maintenance on infrastructure already in the ground.

3 Example from: https://www.eurelectric.org/media/5016/ai-insights-final-report-26112020.pdf

Proxy data is one solution. For one energy generation company, Capgemini Engineering modelled gas turbine performance using sensor data on conditions of operation and environmental factors, from which the underlying efficiency could be unpicked and used to monitor for maintenance.

Pipes are particularly ripe for improved monitoring using a mix of sensor and proxy data.

For example, The International Energy Agency estimates that up to 40% of fossil fuel methane emissions can be eliminated at no net cost for natural-gas operators (the investment in monitoring and repairing gas leaks, which will also reduce environmental damage, should be offset by the extra gas available to sell).

Optimising control systems
Utilities increasingly have data from operational technology and network control systems, such as SCADA. This can be harnessed to manage switching on and off with much greater efficiency and granularity than previously possible, so networks run optimally. Edge compute can be harnessed to perform analytics on site and deliver rapid at-source interventions should problems be spotted.

Modelling deployment of new storage and generation technologies
Batteries and other new energy assets are likely to play a growing role in energy systems, but major new assets have knock on effects.

Modelling can be used to understand all the implications of such deployments from their effect on energy management, to how they will integrate into existing infrastructure. This will support grid changes and inform decisions and design of control systems of the distribution network, allowing utilities to plan, and ensure deployment of untested infrastructure runs smoothly.

Managing hard-to-predict events
Utilities are affected by the uncertainty of the world around them. Weather can take out power assets, earthquakes can break pipes, and sewers can flood. The better these can be predicted, the more resilience can be designed in, both ensuring backup supply and upgrading assets.

A water utility we work with, for example, wanted to know why two similar cities faced very different flooding patterns. By digging into their data and models we identified that the particular weather patterns in one city drove flooding (a dry spell followed by heavy rain meant solid matter in sewers would sink, then be propelled in a way that led to flooding, which didn't happen in less varied climates). This allowed the utility to upgrade specific infrastructure in that location in a way that addressed the root cause of the problem.

DATA CAN ALSO BE USED TO UNDERSTAND THE CARBON FOOTPRINT OF INITIATIVES SUCH AS NEW BUILDS

Carbon accounting
Data can also be used to understand the carbon footprint of initiatives such as new builds. Models can combine diverse data sources and industry standards to quantify the impact of your bill of materials, energy use, and waste, as well as understanding the effect of offsetting initiatives.

SECTION 2

THE CROSS-CUTTING CHALLENGES TO DELIVERING VALUE FROM UTILITIES DATA, AND HOW TO OVERCOME THEM

To realise the opportunities above, and others like them, utilities need to access the right data, apply the right skills and technologies, and turn it into valuable insight. This may involve drawing on hard-to-access data and knowledge that spans multiple parts of the business, and deploying a combination of skills that do not currently exist.

In this section we look at common data and cultural challenges in utilities that could limit the success of data projects, and how to overcome them.

1. Data Challenges

'We've got lots of valuable data… in the filing cabinet'
Projects such as predictive maintenance of underground cables will need data on that infrastructure. But most was laid before we understood the value of data. You can't just dig it up to fit IoT sensors. This leaves us with the data assets created when it was installed – such as maps, blueprints and handwritten notes – which are not easy for machines to read.

Before we even start any sophisticated analytics, we must overcome the challenge of digitizing legacy data.

Such digitization is increasingly possible thanks to natural language processing and image recognition AI tools. These not only digitize text and drawings, but can be trained to spot key insights and translate them into intelligent systems. In a recent project for Equinor, Capgemini Engineering took such an approach to extracting data from maintenance logs, enabling the company to build a knowledge management system which told engineers how problems they were assigned to had been solved in the past.

'We don't have that *exact* data'
Data is often not quite what is expected. Frequently data is collected at end points – a smart meter, an address of a leaky flooding event. But the data need for your model is the precise location of the cable or the sewer that is leaking. Or, you might have an asset that breaks down once a decade, leaving you with limited data to predict the future with.

Sewers for example are often "inferred" rather than known to exist, making drawing conclusions from data very complex. If we don't really understand what the data means and how it was collected, it can easily render any model misleading or useless.

So you need to find other ways to interpret the environment. This may need you to combine data sets (eg maps, analogous assets, physics-based models) to understand what's happening in the real world. It will mean close collaboration between data experts and engineering experts familiar with these assets.

DATA OFTEN EXISTS IN SILOS, LOCKED WITHIN CLOSED DEPARTMENTAL IT ARCHITECTURES, OR EVEN LAPTOPS

'I'm sure someone has that data, but I couldn't tell you who'
Historically, utilities, like many engineering industries, focused on making each division focused and efficient, with advantages for specialisation. But in a more collaborative world this has left a legacy of barriers to data sharing. Data often exists in silos, locked within closed departmental IT architectures, or even laptops.

For example, if your model requires combining data on energy demands in Oxford, with data on Oxford's substations, you may need to start asking around the organization to find what you need for the model, which can slow progress.

In the short term, locating data like this may be a necessity. It will be made easier by clearly defining what you want to do and what data you need, so you can focus your search rather than trying to get as much data as possible. Doing this presents an opportunity to gradually map where data is stored, who owns it, and barriers to getting at it, and feed this into a roadmap towards data maturity (on which, more later).

'Can I do that under GDPR?'
Smart meter data will be the great disruptor of the energy industry. It can help balance demand, spot problems, provide a more personalized customer experience, and even cross sell products.

Smart meter data is owned by the Data Communications Company in the UK, which currently limits its availability to energy providers to manage their own billing. But there are growing calls to free this data up to wider uses, such as for comparison sites.

Greater access will open opportunities to build sophisticated predictive models that provide hugely detailed insights about current and future energy use to consumers and energy providers. This would be used to optimize production, adjust tariffs and cross sell energy saving devices.

It also comes with significant privacy challenges. If models are trained on smart meter data, hackers may be able to mount attacks which identify which data points were used in the training, and reverse engineer them to access confidential information about people or energy systems. Complex models trained on user data may even inadvertently reveal personally identifiable information in ways not considered by the modellers, with legal consequences (see image).

Credit: XKCD

Manipulating this data is tricky, especially on a large scale. Composite learning - tools that combine models to derive insights without combining data - offer potential solutions. Lessons can be learned from healthcare, which has become adept at manipulating highly confidential patient data without compromising privacy. Knowing what to do is hard. Knowing what not to do is just as important.

2. Cultural and Skills Challenges

'This is too complex for the data team'
Predicting energy demand or asset performance involves modelling real-world complexity. They require complex models of physical systems or human behaviour.

Many data experts are versed in spotting patterns in data and may default to these approaches if they do not have a background in utilities. They use data to predict what will happen when you turn a dial, based on what happened in the past.

But past data is not always a perfect guide. That dial is not just changing the data output, but making a physical change such as cooling a transformer. Understanding this allow much more nuance in models – and avoidance of risk. Building this understanding into models requires collaboration between subject matter experts and data experts.

'These data folk just don't get us'
To ensure the aforementioned collaboration, data teams and subject matter experts need to be able to communicate on the same level.

Subject matter experts know their data better than anyone and can be understandably protective of their data and defensive of their expertise. This has to be handled respectfully. Projects have been derailed by over-confident IT contractors telling subject matter experts they have a tool to suck up their data and give them all the answers.

Data scientists can improve models in ways that benefit the engineers, if the former can get the right data from the latter. Statistical models may need to be combined with physics-based ones, and real-world measurement data. The perfect models for predicting how an asset responds to a given scenario, will still be wrong if the data used to mimic that scenario is misrepresentative. The key to unlocking this knowledge is to engage at a comparable level of expertise. Having nuanced conversations about power engineering and physics-based models tends to be the best way to get experts excited to tell you about their data.

IN UTILITIES, DECISION MAKERS HAVE DEVELOPED STRONG INSTINCTS ABOUT WHAT IS RIGHT

'I don't trust this model'
In utilities, decision makers have developed strong instincts about what is right – and rightly so, a wrong decision can be costly. If the model output feels wrong, it will be ignored. This hasn't been helped by an influx of overhyped black box AI technologies that have left many burnt by unfulfilled promises.

Models need to be explainable. Users need to be able to dig into the data and understand what is driving the outcome, assess it in light of their expertise, and intervene effectively. This needs purpose built explainable models, or at the least, explainability tools such as SHAP which can dig into models and understand what is driving the outcome so they can see that the model is right (or reassess if it is wrong).

'This model just tells me what I already know'
A common complaint levelled at data teams is that a lot of work goes into building a model that just confirms something obvious. To overcome this, start with the business challenge and build models aligned to required value. Don't just look for correlations – many will have already been noticed.

That said, the obvious is sometimes worth investigating. Outside perspectives can open new opportunities. In an example discussed above, we presented our findings regarding the effect of rainfall on flooding. The subject matter experts told us they were obvious, and perhaps they were. But no one had thought to check for this until we started digging into their risk models.

'This works, let's deploy it everywhere'
On the other side of the coin, once models are shown to deliver value, teams sometimes get over-enthusiastic. Modellers default to an approach they know or have seen work in another context. Engineers want to roll out their favourite model everywhere.

But it is worth slowing down. An algorithm for pipe degradation might be a decent approximation for all pipes, but perhaps there is a better one for pipes in different geologies or which have more intensive use. Knowing what modelling options are available and being able to pick the best algorithm for the job can lead to incremental improvements that add up across the organization.

'We just don't have the capacity'
Most utilities have some data science function, but they tend to be relatively small, siloed, stretched, and overwhelmed when faced with decades of vast engineering data or a deluge of consumer data. Teams may have expertise, but not necessarily the governance frameworks or range of skills to make the best of data.

Data expertise is necessary but not sufficient. There is also a need for managers and translators that can allow for robust exchanges between subject matter experts, data scientists and the business, creating specialist teams that can understand problems and build bespoke solutions.

Trusting Your Data: Lessons From Oil and Gas

The oil and gas industry faces some similar data challenges to utilities. Both manage large assets and infrastructure, which require planning and lifecycle management. Both operate in highly regulated environments. Both need to make real time decisions where mistakes can be costly.

Oil drilling involves a major asset covered in sensors which collect data on physical parameters, such as time and depth series measurements on pressure, rock hardness, or drill angle. Taken together these can construct physical models of the drill dynamics and subsurface chemistry, which can be used to optimize the process and predict impending problems.

Similarly, utilities need to collect data on substations, pipes, and wind turbines to model how they are operating in the real world, in order to optimize, manage and spot problems.

In both cases, data comes from a wide range of sensors. These may be totally different datasets and formats – eg temperature and vibrations – that needs to be aggregated to give something meaningful. They may use different labelling systems for the data.

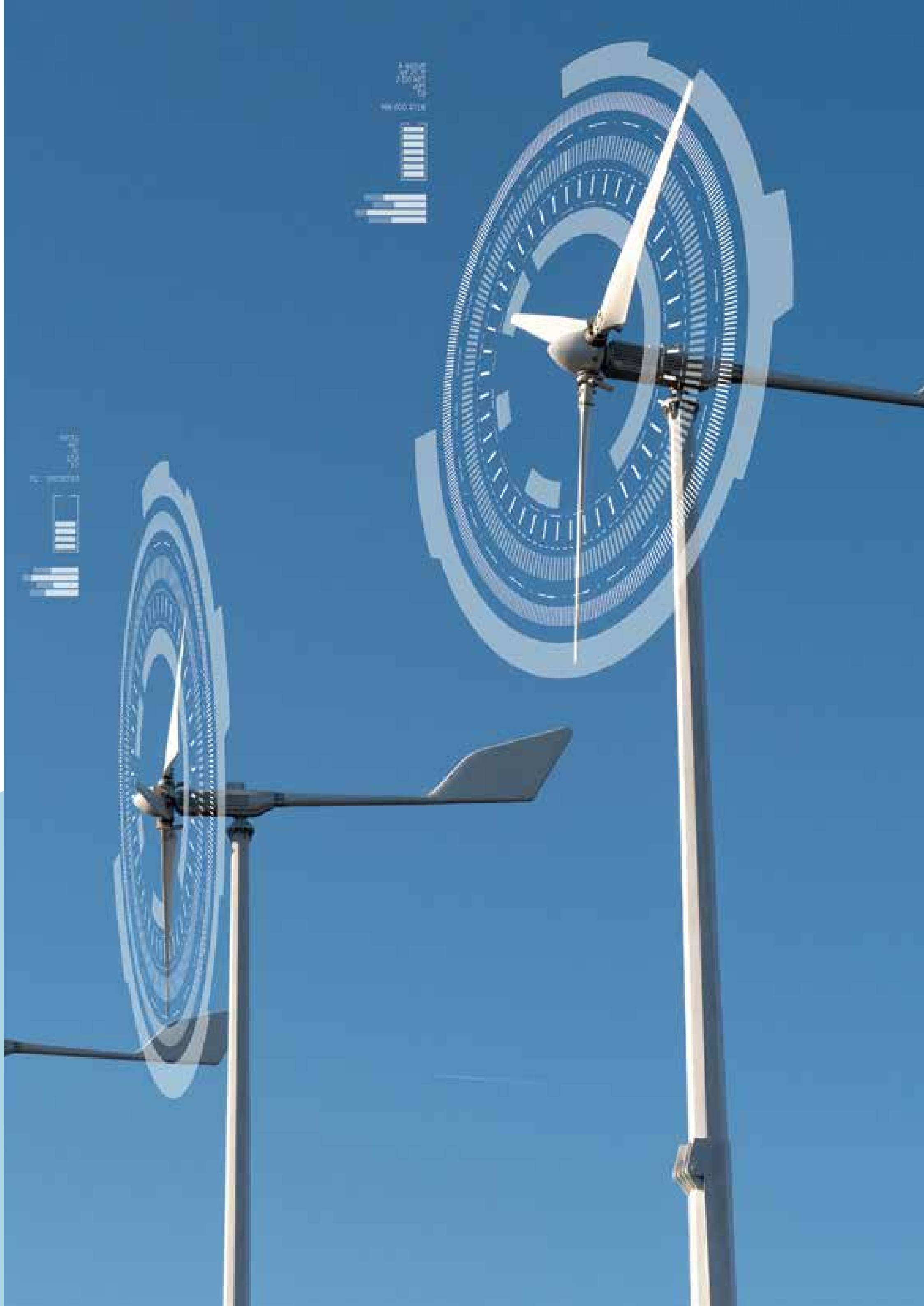
Sensors may be deployed at different times by different companies and collect data in different ways. Time stamps may be different – if one sensor is ten seconds out, any model will misunderstand what it is seeing. By the time the data reaches the model it may have hopped through several nodes – the sensor, the asset, the service provider, the company cloud – which may introduce delays to when the data arrives.

All this complexity needs to be dealt with before the data is fed into the model, so that it can build an accurate picture of the world it is predicting and provide reliable insights.

To make models work, the data coming in needs to be available and consistent. This means mapping incoming data into consistent formats. It may also need redundancy planning in case one node fails and takes out a critical data stream – if you can't get your primary feed, can you switch to a system that extrapolates what you need from other feeds, or do you wait?

To make these models work, we need to be aware of all problems data can throw up. A solid understanding of the physical systems they are measuring - rock chemistry, instrument vibrations, power systems, fluid dynamics – is vital to be able to label things correctly, manage variability in data, and spot and address anything that looks wrong. This domain experience, which ensures the data coming in is correct, is just as important as the maths and the software.

Only then can we be sure that the data feeding the model, or passed to the data science team, is robust.



SECTION 3

BRINGING IT ALL TOGETHER TO BUILD AN INSIGHT-DRIVEN UTILITY

The long term goal of all of the considerations outlined previously is to become an insight-driven utility. Such an organization would have oversight and mastery of its data. It would be constantly launching multiple data projects (such as those described in section one), and deploying resulting models into the business to provide reliable predictions and automation.

In this final section we move up a level and look at the organizational changes needed to become truly data-driven.

Create a culture of innovation and agility
Create a culture where lots of data projects are being tried. Encourage innovation and managed risk-taking, but have checks to make sure failing projects are spotted and stopped. Start small and build up – a changing energy demand model might start with a few houses to gather fast insights, then scale out to a small town, then do several towns in parallel, then link them all together, building accuracy and complexity as you go.

Create collaborative teams and structures
Ensure you have access to the skills you need. This may combine data management, data engineering and modelling. These will need to include people who truly understand the data and what it represents in the real world to get the most out of it. It also needs IT and software skills (MLOps) to operationalize the models.

HAVING SUBJECT MATTER EXPERT BUY-IN IS KEY TO ENSURING YOU CAN GATHER THE DATA AND COMBINE MODELS IN WAYS THAT ACCURATELY PREDICTS THE REAL WORLD

It must also include people who can bridge the gap between subject matter experts and data experts. Having subject matter expert buy-in is key to ensuring you can gather the data and combine models in ways that accurately predicts the real world. This needs people who speak their language and understand their data and challenges and who are good managers that the experts can trust.

Embrace Proof of Value over Proof of Concept
Require use of 'proof of value' to start each data project – an exercise which tests a range of data ideas and asks, if I did this would it be valuable to the business. If it is, the team goes and checks how easy it is to get the necessary data and assesses whether the cost outweighs the benefit. This is a great way to ensure only the most promising projects move forward.

Use governance frameworks to progress a portfolio of data projects
Mandating governance frameworks for data projects, such as Capgemini Engineering's RAPIDE, help progress multiple projects through a series of stages in a well-managed way. It helps keep projects on track, ensure the right things are done at the right time to set them up for success, and provide regular reviews which stop projects if they do not look like they will succeed.

DATA SCIENCE AND AI DEPEND ON GOOD DATA TO BUILD TRUSTWORTHY MODELS THAT PROVIDE VALUE

Get your data strategy right: Great Insight depends on good data
Data science and AI depend on good data to build trustworthy models that provide value. An effective data strategy is essential to create order.

ASSIGN DATA OWNERS AND STEWARDS WHO ARE RESPONSIBLE FOR ENSURING DATA IS COMPLETE AND UNDERSTANDING ITS LIMITS

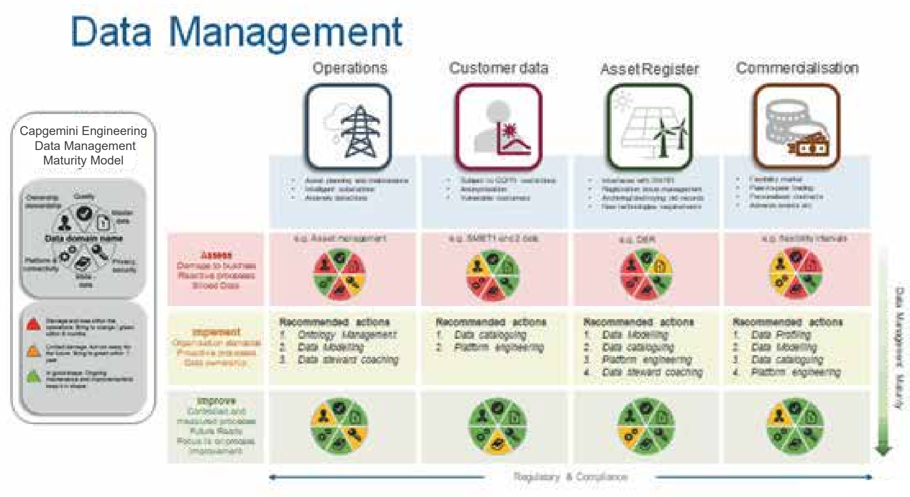
We recommend starting by mapping what data you have and where it is. Assign data owners and stewards who are responsible for ensuring data is complete and understanding its limits (provenance, transparency, etc). These people then become the go-to for anyone who needs that data – whether for building models or responding to regulatory issues.

To make things easier in future, set up processes to automate capture of data and meta data. Automatically document the modeller workflow, and preserve their working environments so they can easily revisit previous approaches.

Build a roadmap
As you mature, look at your data landscape, and the outputs and challenges from your portfolio of projects, and start identify where your organization needs to invest to seamlessly deliver the types of data projects you need. Look at skills, technologies, data management, frameworks, etc. Constantly update this as new issues and opportunities are identified.

Get buy in from the board
Ensuring a long-running, well-funded, and sucessful transformation programme needs board buy-in. Find ways to get the board excited about the possibilities – to get people talking and engaged. The board probably won't understand at a detailed level but it's important to bring everyone on the journey. Getting the business involved in proof of value exercises can inspire them and give them a top level understanding.

Data Strategy Dashboard
We sometimes use a 'trivial pursuit' approach (see image) where each 'wedge' represents an aspect of data maturity (data quality, privacy, etc) and each is awarded green/amber/red. This is often an effective way to both communicate and to start useful discussions (and sometime a bit of healthy competition between departments to improve).



CONCLUSION

The opportunities from data in utilities are huge, but so are the challenges of making sense of, and drawing insight from, diverse and complex data.

Capgemini Engineering can help you navigate the complexity and ensure you succeed. We understand this complex data, and we talk the language of the energy and water engineers capturing it, and the decision makers exploiting its insights.

We have proven frameworks for quickly deriving value from individual projects, and for building portfolios of projects which advance your digital maturity at a manageable pace with a combination of agility and appropriate governance oversight.

Contact us to discuss the insights from this paper and how we can apply them to individual data challenges or your digital transformation.

AUTHORS

Charlotte Brown
Sector Director, Energy

Matt Jones
Lead AI and Data Science Strategist

Danica Greetham
Analytics Consultant

Nick Cook
Consultant

Jeroen de Jong
Senior Data Science and AI Consultant

Chris Hawkins
Operations Manager and Consultant

Dave Dungate
Technical Manager