

DISRUPTIVE DATA SCIENCE

How governments and businesses can
unlock the true value of big data



WHAT TO EXPECT

Technology is driving an increase in the data available, collected and used by businesses and public bodies. Harnessing this data is key to developing competitive advantages and effective policies.

It can be challenging to navigate all this new information. That's why data science is more important than ever.

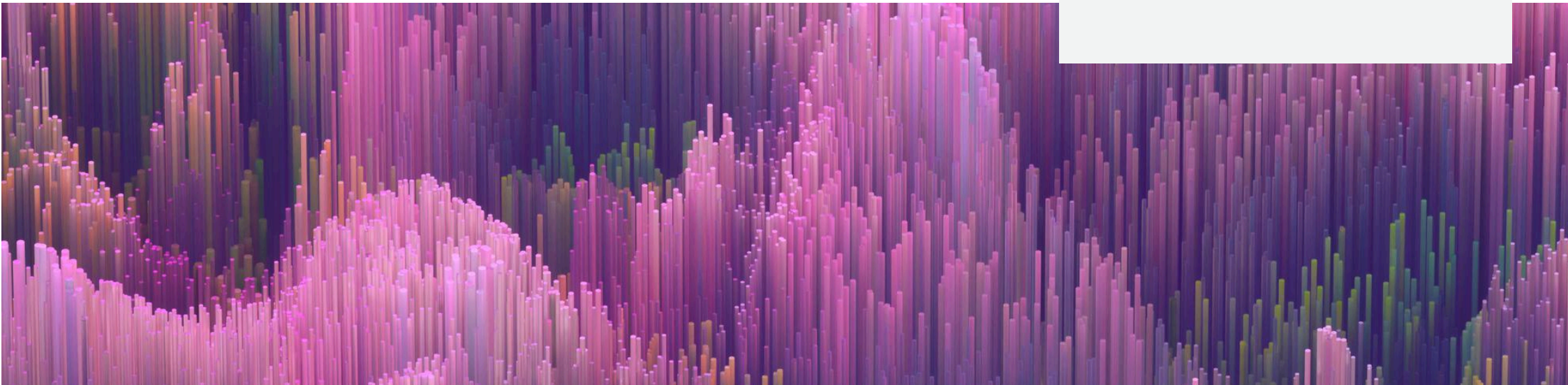
This handbook is all about data science at Frontier. It introduces our team, what we work on and how we generate value and solve problems for clients using data.

We want to stimulate a discussion, so we'd love to hear your views. Look out for the polls throughout the handbook, and feel free to get in touch.

How has your business's use of data changed in the last five years?

- Increased
- Stayed the same
- Decreased

POST ANSWER



WHAT TO EXPECT?

UNLOCKING THE COMMERCIAL VALUE OF DATA

REGULATING DATA: POLICY ISSUES FOR GOVERNMENT

IMPROVING REGULATION WITH DATA: THE EXAMPLE OF WATER

ACCESS DENIED: WHY COMPETITION AUTHORITIES ARE WORRIED ABOUT DATA

MEET THE TEAM





UNLOCKING THE COMMERCIAL VALUE OF DATA

Businesses collect more information on their customers than ever before, but clients often tell us it is difficult to know whether they are using data in the best way. In some ways, that's not surprising – collecting more data means more complexity and choice, and it is often difficult to see the impact directly. So how do companies ensure that more data leads to better results?

THE FULL PICTURE

Data can transform commercial performance when deployed effectively and as a result business models have adapted to become more data-driven. 'Tech giants' like Amazon, Google, Apple have emerged, but most businesses are increasingly investing in data. The outcome is a more targeted customer proposition, pricing is more reactive and supply chains can better predict risks.

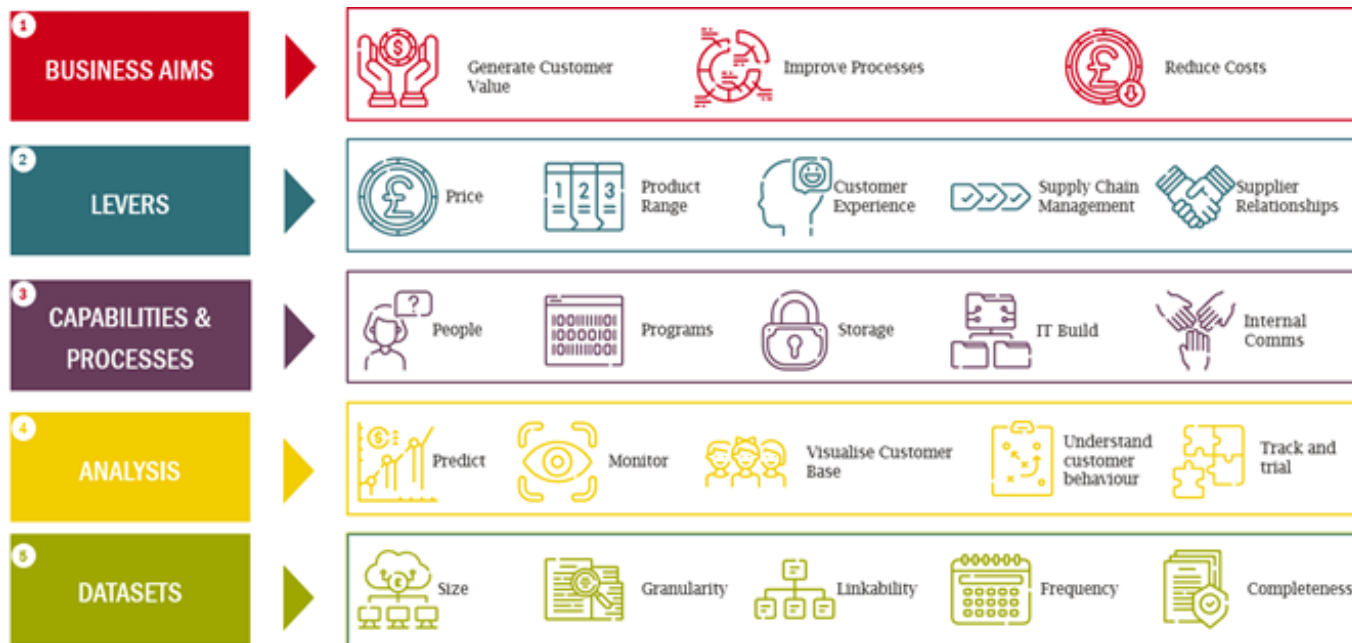
But while the potential upsides are clear, data-driven business models are more complicated. As an input to production, data is highly interdependent – generating value from it depends on what you use it for, and how you combine it with other inputs. There are also so many different options for investing in data, which can make it harder to identify the best use cases.

Businesses need a clearly defined data strategy to meet the challenge of generating maximum value from their data. The following sections outline how to approach a commercial data strategy, including the capabilities businesses need and three options for upgrading their datasets. The final section concludes on how to implement your strategy.

START BY MAKING A DATA MAP

A useful starting point is to make a data map, setting out how data supports the business model. Doing this early will help businesses see the bigger picture and the full range of investment options, while navigating some of the finer details on how data can yield better performance.

Every business will have a different data map, but here's a high-level overview.



This data map has five key elements:

- **Business aims:** What is the business trying to achieve? Why does it want to invest in data?
- **Levers:** What levers can be pulled to achieve these aims? How can decisions on pricing, product range and customer experience influence outcomes that are consistent with the business aims? How might data-based insights inform which of these levers to pull, and when?
- **Capabilities and processes:** What are the processes linking data insights to business levers? Are the business's internal capabilities fit for purpose?
- **Analysis:** Does the business generate the best insights from data? Does it perform the right analyses?
- **Datasets:** What datasets are collected? Is data collected on individual customers or at a more aggregated level? Can the business link together data sources to unlock new analyses?

When the data map is complete, businesses can flesh out their strategy by thinking from the top of the map downwards, starting on business aims and following the logic to finish on what datasets it needs to collect.

Here's some guidance on each of the key elements.

BUSINESS AIMS

First the business should consider why it wants to invest in data. As economists we try to understand how businesses can use data in two ways: 1) driving demand by improving the customer proposition or 2) creating more efficient supply by cutting costs and generating efficiencies. The examples we give in this article relate to how data can improve the customer offer.

These business aims will set the direction of your data strategy – so businesses should be as clear and specific about them

as possible. So if the aim is to generate customer value by changing the proposition, then which customers does the business want to target, and why? And what are the strategic reasons for these aims – are they purely opportunistic to create new sources of value for customers, or are they also for defensive reasons, for example mitigating competition risk in a growing market segment?

LEVERS

Next businesses should consider how data insights can inform changes to the customer proposition. This means understanding what commercial levers the business has to adjust the offer, and how data could inform choices on whether and how to adjust these levers.

Businesses have lots of options, so choosing which levers to focus on is key for prioritising their data investment. We suggest three activities to do this.

- When generating a list of ideas, businesses should be creative, including being open to business transformation by creating new levers through data. As a thought experiment, businesses could ask: 'If I had all the data I could possibly collect, what would I do with it?'
- The business should test the feasibility of each option with a wider exec group. As economists we like to do this by setting up hypotheses for each option, outlining what would need to be true for the option to be feasible. Then the exec can move on to asking what evidence exists to inform whether these points are true. It's helpful to be flexible on what counts as evidence to avoid holding up the process – anecdotal evidence from two or three members of the exec may count as "good".
- Businesses should assess which levers relate most closely to their strategic advantage and business aims. This is an effective way to prioritise which levers to focus on.

CAPABILITIES

A clear plan for how data supports business is important, but **businesses also need the right capabilities to deliver value in practice**. Two important capabilities are people and internal processes.

People and techniques are important in translating data into insight, which feeds into decision-making on your customer offer. Businesses need to have the right people, with access to appropriate programs and techniques for analysis.

It's important to use **appropriate programs** for data analysis - not Excel for very large datasets! Switching to R, Python or another suitable programs will require up-front investment but also pay off relatively quickly - they will dramatically reduce the time taken to process data and open up new options for analysis (like machine learning).

A more challenging question relates to whether analysis teams are investing in the

right techniques. The temptation is often to hire experienced data scientists to use more advanced methods. But it's important to have a clear understanding of why you're doing this, and how advanced methods will yield better insight. If businesses can follow the logic of their data map - starting with a vision of what business performance changes to achieve, what choices exist to affect them, leading through to how insight from data analysis could inform those decisions - that will provide a solid understanding of what data insight is helpful. In turn this can guide businesses to appropriate techniques, which will inform the type of team they need.

It's also important that **internal processes** efficiently connect data insights to decision-makers, since value from data is realised when insights are used to inform business decisions.

But 'efficiency' looks different to different businesses. Consider a food-to-go retailer responding to hourly shifts in demand.

Their analytics team has insight that a combination of bad weather and sports events increases the likelihood of fast food ordering, and bad weather is due in an hour. To be useful, this insight has to reach the decision-maker with control over pricing and promotions very quickly.

In contrast, a large supermarket might design their offer according to changes in behaviour across weeks or months. These businesses can afford a longer lead time from insight to decision. But data insights still need to be connected to the right people at the right time.



ANALYSIS

Insights from data analysis will influence decisions on whether, how and when to pull business levers.

These insights are useful across business functions. Business intelligence teams benefit from a richer understanding of their own customers' behaviour and monitoring competitor performance across the wider market. Product teams can improve the proposition by monitoring whether product changes improve performance, and the possible reasons why. Marketing teams will better know how their customers behave and why, which will inform business communications and adverts.

At Frontier we think about creating data-based insight in two main ways.

The first is **identifying and visualising** data. Techniques include exploratory data analysis or unsupervised machine learning to identify patterns – for example, to group similar types of customers together. Visualising data is also important to uncover a clear story. Interactive dashboards or mapping geospatial data in Python are effective options.

The second is **explaining and predicting**. Businesses will know what worked in the past, but understanding why, and what is likely to happen in the future, are key to commercial decisions. Statistical techniques like econometrics and machine learning can help businesses understand the causes of customer behaviour and predict the likelihood of future events.



DATASETS

Finally, **businesses need to collect the right data** to perform analysis.

We've noticed that businesses can sometimes focus on quantity of data – this is important, but there are substantial benefits to be gained from collecting the right quality of data. This means data with the right **characteristics**: granularity, 'linkability', time and frequency.

The **granularity** of data relates to the level of aggregation for each data point. For a UK retailer, this could range from very granular data on individual transactions, up to more aggregated data points for organisation-wide metrics.

Getting the granularity of your data right could support both 'identify and visualise' and 'explain and predict' analyses. For example, a business with separate data points for product groups in individual stores can better *identify* differences in product-level performance within and

between stores, than if it did not collect data at the store level. The business could also perform follow-on data analysis to *explain* the drivers of these differences – particularly if it can link in additional data on other relevant variables.

But it's not always true that investing in more granular data is worthwhile. More granular data will mean a larger number of data points, which are often time-consuming to collect. That means it's important to ensure that collecting data at greater levels of detail is useful for commercial decisions.

'Linkability' is the degree to which data points in separate datasets can be matched and merged. Organisations with linkable customer datasets will often use unique customer identifiers, where every data point for a given customer will be consistently associated with that customer, even for datasets in different business departments. These datasets are easily merged.

Using customer identifiers enables organisations to create larger customer-level datasets, covering different customer characteristics and behaviour. This can unlock more powerful forms of 'explain and predict' analysis, which feed into improved proposition design. A good example is where a retail bank uses customer identifiers to match data on its customers' current account and savings balances. This enables it to form a more complete understanding of customer behaviour across a wider set of products, including how customers move money across savings and current accounts to limit using their overdraft.

While some degree of 'linkability' is important, significant up-front investment is required to fully roll-out customer identifiers across an entire business's datasets and to coordinate their use consistently. Intermediate forms of linkability are also possible, such as businesses using definitions of local areas or customer characteristics (age ranges,

socioeconomic status) which align with external sources like the ONS. The case for investing in improved 'linkability' is strongest when businesses are confident their systems can identify customers efficiently, using automated processes.

A **time series** of data – where data is collected on the same customers at different points in time – improves the quality of 'explain and predict' insights and unlocks new options for analysis. In particular, having a time series is necessary for conducting trials where outcomes for two groups are recorded.

The **frequency** of time series data reflects the time interval between data points – for example, on an hourly, daily, weekly, monthly basis. Collecting higher frequency time series data can also strengthen 'explain and predict' insights, by enabling adjustments to be made to avoid bias in results. Businesses can also better monitor customer behaviour, by quickly identifying possible changes in behaviour and then

tracking their development to determine if the change is temporary or permanent.

However other factors limit the usefulness of time series and high frequency data. Datasets will need to be 'linkable', since collecting data on customers across time requires identifying the same customer in different periods. Very high frequency datasets also contain lots of data points, so your business will need the appropriate automated collection methods and storage systems in place to streamline data collection and processing.

BUILDING YOUR DATA STRATEGY

Finally, to effectively implement your strategy, make sure responsibility clearly lies with teams. And find ways to monitor whether it is generating real value, iterating further as appropriate.

How often does your business use data?

- We use data at scale to inform most key decisions and monitor performance
- We use data periodically and appreciate its value – with capability to scale up further
- We use data periodically and appreciate its value – but have reached the limit of what we can use data for
- We don't rely on data at all

POST ANSWER

AUTHORS



LAURA PETSCHNIG
MANAGER



ROB WOODHARD
CONSULTANT

[CLICK HERE TO MEET OUR DATA SCIENCE TEAM](#)





REGULATING DATA: POLICY ISSUES FOR GOVERNMENT

Data is a critical part of modern society, for both businesses and consumers. It has the power to drive productivity, boost innovation and increase inclusion.

But Governments worldwide have an important role in ensuring the right conditions are in place for society to realise the full value of data. Indeed, there are many initiatives that have this goal, from the UK's National Data Strategy to the European Commission's proposed Data Governance Act. This article explores how data is different to other physical inputs like labour or capital equipment, and what this means for data policy.

HOW IS DATA DIFFERENT?

From our work with industry and government, there are three main ways that data differs to other value creation inputs.

1. DATA IS RARELY TRADED THROUGH MARKETS

Unlike most inputs, data is rarely traded through markets. Instead companies will generate their own data and use it in-house, or transfer it without an explicit price – for example as part of a wider relationship between suppliers and customer organisations. Some data markets do exist but are often 'thin', with a small number of buyers and sellers.

We have observed three reasons why data is rarely traded in markets.

- **Data ownership is often unclear**, which affects whether data can be sold. Who owns data? Is it the individual, the company collecting it, or the country in which activity takes place? This is an issue for data markets, since clear identification of a buyer and seller is a pre-requisite to creating markets for products.
- **Data can create a competitive advantage**, which means organisations may not want to sell it. They might want to protect this advantage, rather than sell their data to other companies who will

use it to compete more effectively with them.

- **It's hard to judge a dataset's quality** before purchasing it, which reduces the likelihood of buyers and sellers agreeing a price. Data is often generated by a seller, who will have an incentive to overstate the data quality. Buyers understand this and also recognise the limits of their own understanding – both of which will make them less inclined to purchase the dataset.

2. THE VALUE OF DATA IS CLOSELY LINKED TO HOW IT IS USED AND GENERATED

Compared with other inputs, the value of data is more sensitive to the context in which it is generated and used. There are several reasons for this.

Data can have multiple concurrent use cases. Data is easily stored, so can be moved and combined with lots of different inputs. It is also non-rivalrous – if data is used for one activity, that does not prevent it being used for other activities and malleable. Datasets can be updated, reshaped and merged. With the right identifiers, they can easily be linked to other inputs.

The added value from data is closely linked to how it is processed and analysed by humans. In particular, there is a wide set of possible ways to analyse data, and the quality of any given analysis could vary significantly across individuals. As a result, the value of data outputs is not mechanistically related to the nature of data inputs

3. DATA FLOWS ACROSS BORDERS MORE EASILY

Unlike other inputs, data easily and invisibly flows across borders. This means the same dataset can be used in different sectors and countries, which makes data flows hard to measure. As a result governments need to think differently about how to quantify data and its associated economic activity.

We have observed several reasons why data flows more easily than other inputs.

- Data can move instantaneously, unlike many physical inputs. It is easily stored, can be moved quickly through file transfer and not subject to other physical constraints.
- Data is generated from a wide range of activities, which are often dispersed across sectors and countries. This means data is often generated in multiple locations at once, but these locations may be in a different place to where the data is needed, which is where it will ultimately flow.

- There is pressure for data to flow easily. A single dataset is often more valuable for any firm or government when it inputs into a larger number of uses, and combines with more inputs.

WHAT DOES THIS MEAN FOR PUBLIC POLICY?

The unique nature of data creates specific reasons for government to intervene in markets, and challenges to designing these interventions effectively. We consider the implications for three common policy-making activities.

1. ESTIMATING THE ECONOMIC VALUE AT STAKE

Governments need to understand the size and shape of markets and how they work before moving to design policies. This is particularly important for data, since data markets and ecosystems are often nascent and less defined. It's hard to estimate the size of economic activity associated with data for the following reasons.

It's challenging to define what economic activity is associated with data, due to the nature of how data flows across sectors. Most UK markets have a known definition according to a Standard Industrial Classification (SIC) code, and ONS data on GVA by SIC code is available. But for data, governments may need to define data ecosystems – rather than markets – where the same sources of data link across traditional markets to generate value.

It's hard to disentangle the value of data from traditional services, due to how closely data is linked to other inputs as part of production. Consider a supermarket using a dataset of customer transactions. By itself, this data is of limited value. But suppose a team of data scientists combine it with information from loyalty programmes. Insight from that combined dataset feeds into better commercial decision-making on pricing, which translates into increased sales revenue. How can the pound value of using a different transaction dataset be disentangled from the value of other datasets, the analysis performed by data scientists, and wider decision-making processes?

Assessing the economic value of data ecosystems is possible using a combination of economic frameworks to define an ecosystem, and data science techniques to estimate its size. In particular, innovative use of appropriate analytical techniques and new data sources is helpful to account for the complexity in data ecosystems.

Recently, we've completed projects estimating the size of the [UK geospatial data](#), [UK data assurance markets](#) and the economic impact of cross-border data flows on [the EU's economy](#). In the geospatial study, we used economic theory to define the demand and supply sides of the market, and then triangulated across multiple innovative data sources to scale aspects of geospatial activity.

2. REASONS FOR INTERVENING

Governments need to diagnose if, how and why markets are not working before designing policy to affect change. But with data, the relevant market may not yet exist, or exhibit specific characteristics, for the reasons we outlined above. As a result, governments often design data policy with two aims in mind.

AIM 1

To stimulate the creation of a market, possibly by removing barriers to market transactions. This could include incentivising data holders to make their data available where the social gain of doing so is larger than the benefit of keeping it private. Where data is commercially sensitive, developing a public data intermediary – an organisation that acts as a go-between for data providers and users – may be justified.

AIM 2

To improve the functioning of existing data markets to deliver better outcomes and unlock economic and social value from data. This could include overcoming market failures related to coordination, by mandating the use of certain standards which boost interoperability and sharing. For example, the ONS has set out its data standards – these may encourage other organisations to adopt similar approaches, enhancing [interoperability](#).

Government may also take actions to increase trust in nascent markets. Trust needs to be created and maintained between the parties who share, collect and use data. Mistrust reduces data sharing and the economic and social value that can be generated. Trust could be boosted by the establishment of regulations and standards which clearly set out who can access data and on what terms – such as the Open Banking reforms.

Regardless of the aim of the policy, government needs to take a careful, framework-driven approach to ensure that data policy is well targeted and avoids unnecessary consequences. [Our work for DCMS](#) has identified issues that may prevent access to data, and a set of levers that Government could use to remove or mitigate these issues. Our economic framework considers the potential causes of market failures in data markets – externalities, imperfect information, heuristics and biases, market power – honing in on the specific issues before

considering options for intervention.

3. VALUING POTENTIAL BENEFITS AND COSTS

The impact of the policy actions described above would have to be carefully assessed. The costs of any intervention could include direct financial costs to the public sector, displacement of commercial activity, or stifling of innovation. But the benefits could include facilitation of new markets, supporting faster economic growth or enabling socially valuable use cases.

Standard economic evaluation approaches may need to be adapted. Disentangling value in this context can be difficult, and articulating a counterfactual absent any intervention may be challenging given the rapid evolution of these ecosystems. Careful attention should be paid to the following issues.

- Whether sufficient **private incentives to share data** in a given context exist;
- What the **potential use cases** might be and the likely value at stake, both private

and social; and

- Whether significant **unintended consequences** from intervening exist, such as undermining incentives to invest in data collection and quality.

Separately, the proliferation of data flows can also provide useful tools for evaluation too – for example, web scraping can be used to gauge the size of ecosystems based on firms’ own descriptions of their activities.

CONCLUSION

Data has distinct characteristics. It’s rarely traded via markets, its value varies by use case and it easily flows across sectors and borders. These features can generate new rationales for public policy intervention, but they also pose challenges for governments. That’s why flexible and innovative approaches will be needed, to ensure the value of data is fully unlocked and widely shared.

What is the biggest barrier to your business using more data?

- Collection or storage costs
- Culture or lack of belief in use cases
- Lack of analytical skills
- N/A - it's just not relevant to the business model

[POST ANSWER](#)

[CLICK HERE TO MEET OUR DATA SCIENCE TEAM](#)

AUTHORS



ROB WOODHARD
CONSULTANT



SARAH SNELSON
DIRECTOR



NICK FITZPATRICK
MANAGER



FEDERICO CILAURO
MANAGER



IMPROVING REGULATION WITH DATA: THE EXAMPLE OF WATER

Technological progress has created exciting opportunities to gather and analyse data on a scale that was unthinkable a few years ago. And over the next decade we expect data collection and analysis to evolve further, to provide faster and better insights.

These developments have the potential to improve how we regulate economic sectors. But how might that work in practice? The water sector provides a useful example of how data can improve regulation.

NEW DATA CAPABILITIES

Data analytics are more sophisticated and user-friendly than ever before, and tools like machine learning and AI allow constant optimisation of operations.

With the continuing improvement of capabilities like these, we can envisage a world where companies in the water sector have the tools shown in Figure 1 at their disposal.

The opportunities created by the rise in data capability raise two key questions for regulation in the water sector, and

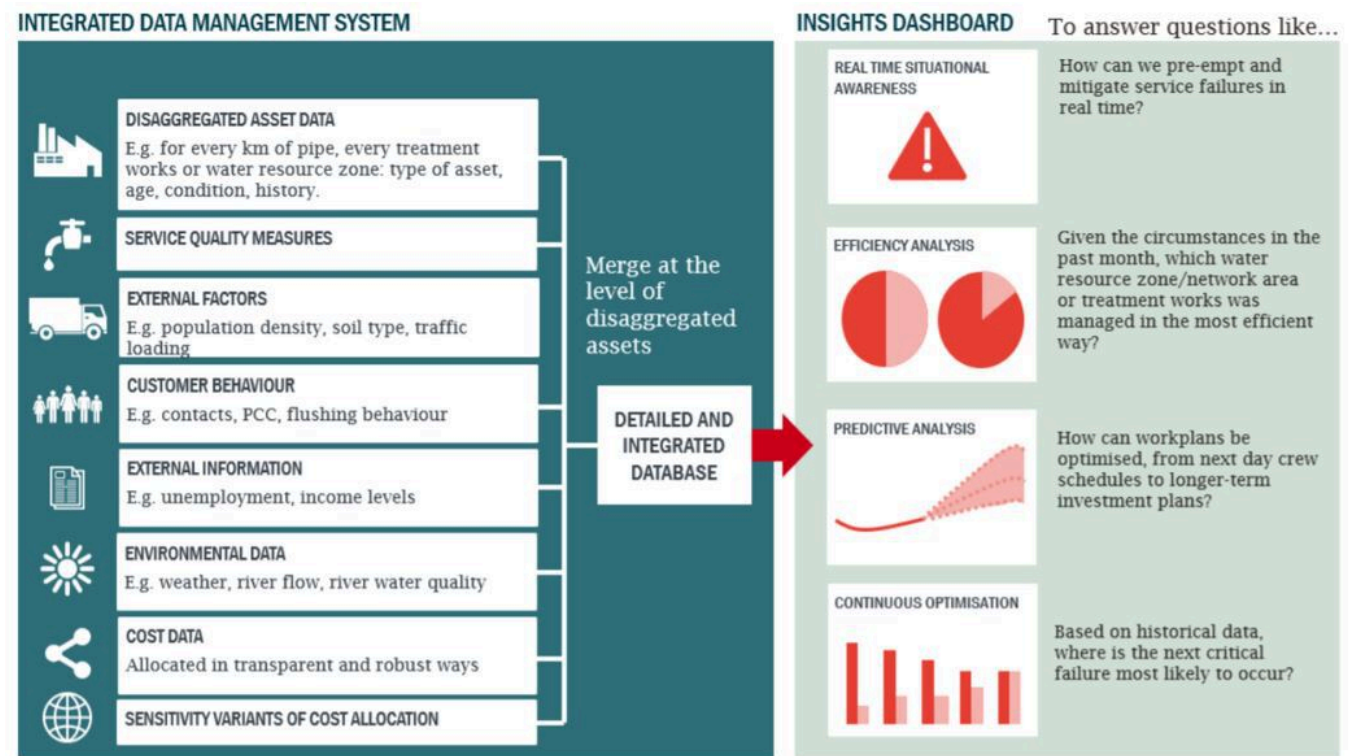


Figure 1: Future water company data capabilities

regulation more widely:

- How can companies and Ofwat use more and better data to improve economic regulation in the future?
- Is the current regulatory approach incentivising the right amount and type of investment in more and better data?

HOW CAN COMPANIES AND REGULATORS USE DATA TO IMPROVE REGULATION?

We've identified three areas where data can transform the current approach (for more detail, take a look at [our long-read article on data in the water sector](#)):

1. **Companies can draw on more and better data to improve the quality of their business plans.** There are great opportunities to develop better evidence on customer behaviour and views, efficient opex, cost and service special factors, enhancement projects and service quality targets. To achieve

companies' objectives at each price control, the evidence needs to be

supported by a clear regulatory data strategy.



Figure 2: Why companies need a regulatory data strategy



Figure 3: Steps to develop a regulatory data strategy

2. Companies can apply economics and behavioural science to more and better data to improve efficiency. There are three key opportunities - and companies that harness data and apply new analytical techniques can gain a significant advantage.

1. Benchmarking efficiency both within water company boundaries and between companies. This will generate genuine insights and therefore help drive efficiency.
2. Analysing how customer behaviour affects your costs to help shape initiatives to change people's habits. Technological progress opens up a range of opportunities to track actual behaviour via apps and device-based technology. This has the potential to transform the way companies engage with customers and to reduce costs.
3. Analysing customer views on real-time operations to optimise customer satisfaction.

3 **More and better data will create opportunities but also risks for Ofwat's benchmarking.** A data-rich world provides opportunities to improve our understanding of how external factors influence companies' costs and service quality. More and better data opens up new possibilities for integrating benchmarking of costs and service. New approaches can also improve accuracy and precision and therefore reduce the risk of misallocating cost allowances, leading to greater confidence in the results. But there will be risks, and ways to mitigate these need to be considered – such as increased transparency and avoiding cherry-picking results.

ARE REGULATORS INCENTIVISING THE RIGHT INVESTMENT IN DATA?

Investment in more and better data requires substantial costs and effort over multiple AMPs. Similarly, the efficiency gains from better insights will be realised over the course of several AMPs. With a challenging PR19 Final Determination and the problems created by the Covid-19 pandemic, investment in data will be difficult to justify.

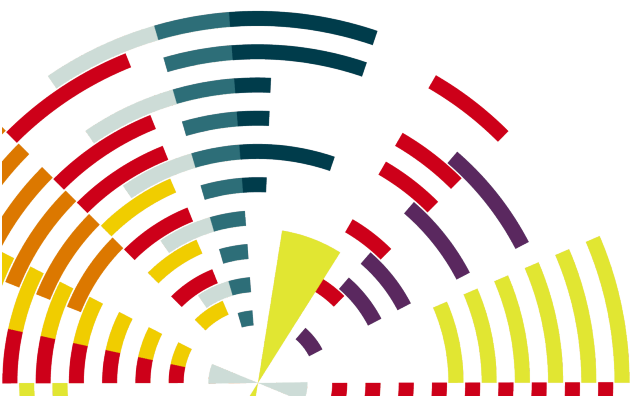
The current regulatory approach creates short-term incentives that may not be compatible with investments and benefits that stretch over several AMPs. However, in our view, more and better data is essential to drive long-term efficiency, which in turn is critical to the legitimacy of the water sector. The regulatory approach should capture the long-term value of more and better data.

WHAT NEXT?

While the future world of more and better data will not be realised in one price control period, it's important for the water sector to have a clear vision of how data will be used in the longer term.

In particular, the water sector needs to develop a high-level vision of how costs and service should be benchmarked at the next price controls review. If it doesn't, there's a risk that data is not comparable, that the incentives to collect relevant data are not sufficient and that it will be too difficult to adopt a new approach. A longer-term vision for benchmarking costs and service would create a way forward, the next price control could be approached in this context.

For more detail on how the data revolution might impact the water sector, take a look at our long-read piece [here](#).



What is your businesses biggest barrier to using more data / investing in better data?

- High costs (such as data collection methods and cost to manage and store)
- Understanding the need (such as it being unclear if it is beneficial or worthwhile to invest in)
- Lack of incentives

[POST ANSWER](#)

[CLICK HERE TO MEET OUR DATA SCIENCE TEAM](#)

AUTHORS



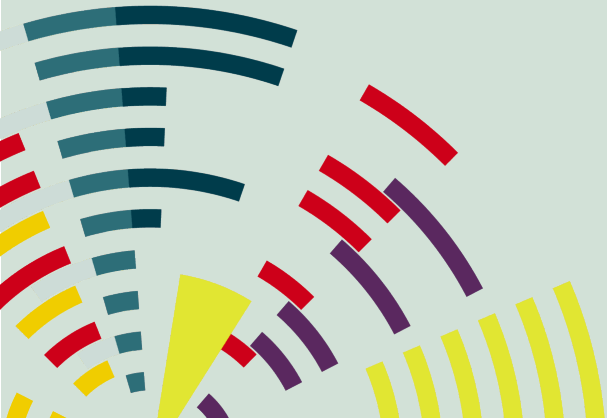
ANNABELLE ONG
DIRECTOR



CHRIS CUTTLE
MANAGER



FULVIO BONDILOTTI
CONSULTANT





**ACCESS DENIED: WHY
COMPETITION
AUTHORITIES ARE
WORRIED ABOUT DATA**

Data assumes an ever more central role in economic activity and day-to-day life. Talk abounds that it has become the new oil: useful and essential for businesses to function, but vulnerable to weaponisation by organisations who want to exert control and advance their own interests.

Debates about the control and abuse of data touch all parts of society - from concerns about protecting personal privacy to issues of national security. In the sphere of economics, competition authorities have been taking note. Worries about the control and use of data have increasingly coloured antitrust watchdogs' thinking in market studies and merger investigations, and the EU's competition commissioner, Margrethe Vestager, has spoken about access to data as a critical factor that can make or break the fortunes of small companies.

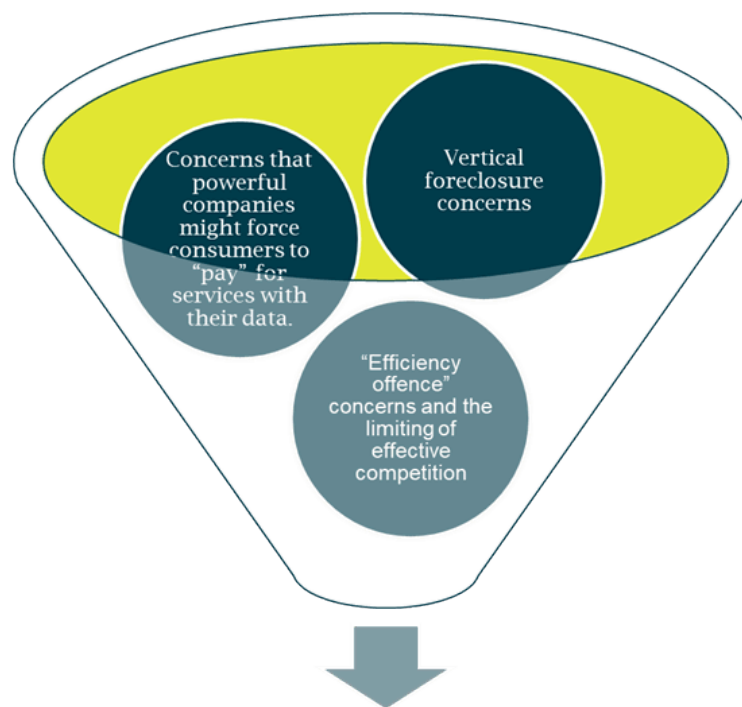
These concerns are beginning to enter competition authorities' formal guidelines. The UK Competition and Market Authority's 2021 merger assessment guidelines are strewn with references to data access issues - which didn't feature in the previous guidelines at all. And, perhaps most tellingly, in 2019 Commissioner Vestager was promoted to lead the EU's drive to create 'A Europe Fit for the Digital Age', alongside her traditional responsibilities.



ALEXA, WHAT'S THE PROBLEM?

In recent merger investigations, competition authorities' worries about data have fallen into three broad categories:

1. **Companies might force consumers to 'pay' for services with their data.** Antitrust watchdogs have suggested that powerful merged companies could demand that customers share more personal data as a precondition for using services. The authorities assume consumers would be unhappy about this, just as they would be unhappy about paying higher prices - although the evidence on this is mixed.
2. **'Efficiency offence'.** A newly merged company might have access to a richer combined body of data than any of its rivals. This could be a good thing if it allowed the merged firm to improve its offering, but some authorities have expressed concerns that it might make it harder for data-poor rivals to compete, leaving consumers at the mercy of a single company.



Competition authorities' worries about data

3. **Vertical foreclosure.** These concerns may arise when an 'upstream' firm that generates data merges with a 'downstream' firm that makes use of this data. Authorities worry that the upstream firm's data is an essential input for not only the downstream partner but also its rivals, and that the upstream firm could potentially foreclose these rivals - to the benefit of its downstream partner - by impairing or refusing access to its data.

While the first two concerns can be difficult to substantiate, they are being increasingly widely discussed. Nonetheless, much of the focus of competition authorities in recent investigations into data-heavy services has been on the third category of concerns – vertical foreclosure.

In such investigations, competition authorities first need to establish whether the firms in question would have both the ability and incentive to withhold access to data and – if so – what effects this would have on competition and consumer welfare.

Such concerns are increasingly featuring as the main theory of harm in merger investigations where the companies in question operate at different levels of a supply chain. Data access concerns were a central theme in the European Commission’s probe into the \$27bn acquisition of fintech data and analytics business Refinitiv by the London Stock Exchange Group, as well as a spate of recent investigations in to mergers between

digital services (such as Google/Fitbit and Apple/Shazam).

NOT REVOLUTION, BUT EVOLUTION

While it’s clear antitrust watchdogs are worried about the role of data in markets, it may not be necessary to rewrite the rulebook to safeguard effective competition. The concerns around data outlined above are not fundamentally different to those the authorities have addressed for more conventional inputs. Nonetheless, competition investigations have shown that data has features that distinguish it from other types of input, which can influence the way these debates play out.

For more detail on how these features have influenced investigations, see our long-read article [here](#).

Will the growing quantities of data drive or hinder competition in your sector?

- Drive competition
- No change
- Hinder competition

POST ANSWER

AUTHORS



JAMES BAKER
ASSOCIATE DIRECTOR



[CLICK HERE TO MEET OUR DATA SCIENCE TEAM](#)

MEET THE TEAM

MEET THE TEAM



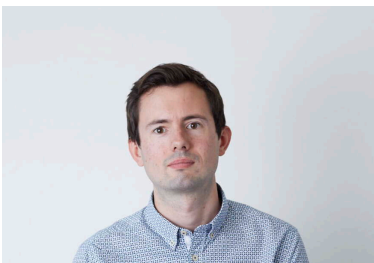
IAIN BOA (HE/HIM)
ASSOCIATE DIRECTOR



FEDERICO CILAURO (HE/HIM)
MANAGER



THERES LESSING (THEY/THEM)
CONSULTANT



ALEX WHITTAKER (HE/HIM)
MANAGER



GET TO KNOW US MORE

IAIN BOA

HOW HAS THE USE OF DATA CHANGED IN YOUR WORK FOR CLIENTS?

Client datasets are much larger now and they cover a wider range of topics. This means it's possible to gain new insights into old questions, but it's also easy to get lost on what to do next. Increasingly, conversations are about how we can get insights from their data, rather than what data they have.

THERES LESSING

HOW IS FRONTIER CHANGING THE WAY IT USES DATA TO GENERATE INSIGHTS FOR ITS CLIENTS?

We're increasingly using tools like APIs, machine learning and big data analysis. Tasks that used to be impossible, or were done manually, can now be automated quickly and precisely. For example, using the Google Maps API, we can run geographical analysis on datasets with incomplete store addresses.

HOW HAS THE USE OF DATA CHANGED RECENTLY IN YOUR WORK FOR CLIENTS?

I'm increasingly working on projects that are scripted from start to finish. This means we can re-run analysis for a different year if new data becomes available, or easily move the analysis up or down a level of granularity. This allows for more flexibility in what we can present to our clients. For example, in a recent retail project, we were able to toggle between analysis at the level of individual products or of higher aggregations, depending on what question we were answering.

FEDERICO CILAURO

WHAT ARE THE OPPORTUNITIES AND RISKS FACED BY SOCIETY IN RESPONSE TO GROWING DATA USE?

In principle, growing use of data can improve almost any area of human activity. Learning more about what works and what doesn't in education is particularly exciting, and so is using AI in a collaborative way.

I'm an optimist about digital technology, but there are risks. Data is not the same as knowledge, and often data on its own has no value. There's a risk that as more data becomes available we focus on data points that don't actually matter. For example, many have spent the last year over-interpreting daily fluctuations in Covid case numbers. Covid case data was very useful, but only once you start looking at patterns beyond daily change.

WHAT ROLE COULD GOVERNMENT POLICY PLAY IN ENABLING OPPORTUNITIES AND MITIGATING RISKS?

Frontier worked on a report for the UK Department of Digital, Culture, Media and Sport that looked at this. One interesting aspect is the 'limited excludability' of data. This means that, as a company, it's hard to police how others use data once you've shared it. This is unique to digital goods and assets that can be replicated and transferred easily. A result we see is that data organisations either do not share their data (or only share it with a limited few), or they only share it through lengthy licence agreements.

We're starting to explore how governments can help overcome this problem. This could include providing standardised data licensing agreements, or supporting trustworthy institutions that work as intermediaries between data providers and users.

HOW HAS THE USE OF DATA CHANGED IN YOUR WORK FOR GOVERNMENTS?

Two changes have happened recently. First, there are new ways to generate data used to inform public policy. For example, we've worked with collaborators using AI to identify and categorise companies active in emerging sectors.

Second, because most countries are grappling with digital technology issues, I've found myself working increasingly on international projects. For example, we've recently used data on online harms in Germany to draw lessons for proposed regulation in the UK.

ALEX WHITTAKER

WHAT DATA ARE FIRMS IN THE ENERGY SECTOR COLLECTING?

Over the last decade, the amount of data generated across the energy value chain has increased tremendously. Most visibly for consumers, smart meters allow suppliers to understand how much electricity and gas is being consumed, hour by hour. Devices like smart thermostats and appliances are also being increasingly adopted. But the energy networks themselves are also producing ever greater amounts of data. For example, while local electricity distribution networks used to be quite passive, smart grid systems allow the network operators to really understand how their assets are working and how to make the most out of them.

HOW ARE THESE COMPANIES USING THEIR DATA, AND HOW IS IT IMPACTING THEIR BUSINESS MODELS AND COMMERCIAL DECISION-MAKING?

On the consumer side, smart meters and half-hourly settlement are enabling some really interesting business models, which can also help with decarbonisation - different types of smart time-of-use tariffs, for example. But at present these mainly appeal to early adopters, so a challenge will be figuring out how to engage other customers too. Like any retail organisation, energy suppliers collect a lot of consumer behaviour data, which can be used to help with this.

Networks and generators are using techniques such as digital twins to predict where and when maintenance or investment should be carried out. For example, an electricity network might use models to forecast which parts of its network may come under stress in the future, and then consider whether this can be overcome using 'smart' interventions like demand-side response or energy efficiency, in addition to traditional reinforcement.

IS THERE ANY NEWLY AVAILABLE DATA WHICH COULD HAVE AN IMPACT ON BUSINESSES OR REGULATION?

While smart meter data in the UK is available to suppliers, other entities like networks and government are not able to access even anonymised customer-level data. These datasets may be useful for networks to better understand the loads on their systems, or for policymakers to estimate the distributional impacts of interventions.

We're also seeing a variety of open data initiatives. Regulators are pushing organisations towards taking a 'whole system' view (considering the impact of actions across the energy value chain) and the availability of new data will help make this a reality.

THANK YOU FOR READING

Visit
bit.ly/data-science-page
Contact
media@frontier-economics.co.uk