THE VALUE OF DATA ASSETS

A report for the Department for Digital, Culture, Media and Sport

December 2021

Frontier Economics Ltd is a member of the Frontier Economics network, which consists of two separate companies based in Europe (Frontier Economics Ltd) and Australia (Frontier Economics Pty Ltd). Both companies are independently owned, and legal commitments entered into by one company do not impose any obligations on the other company in the network. All views expressed in this document are the views of Frontier Economics Ltd.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The National Data Strategy (NDS) identifies opportunities for government action to unlock the benefits of data to the UK economy. Frontier Economics was commissioned by the Department for Digital, Culture, Media and Sport (DCMS) to support it in understanding the value of data and how changes in the characteristics of data can affect value. The work focused on identifying, appraising and evaluating the impact of potential policy interventions under the NDS on the value of data, with a particular emphasis on the "data foundations" and "data availability" pillars. For the purposes of this work, we define "value" as "private value", that is, the benefits generated by the data for the organisation which uses the asset. Wider benefits of data to other individuals and organisations ("spillovers" or "positive externalities") are outside the scope of this study.

This is challenging from both a conceptual (the concept of value can be interpreted in a variety of ways) and practical (empirical information on the value of data is limited) perspective. In light of this complexity, this report provides:

1. An assessment of the advantages and disadvantages of possible methods for assessing the value of data assets in the context of policy-making; and
2. A framework which describes how to approach the valuation of changes in data assets that could be brought about by policy action (e.g. improvements in the interoperability or accuracy of the data).

1: Methods for assessing the value of data assets

Through a non-systematic literature review and interviews with 12 experts and businesses, we identified the following categories of methods for valuing data:

1) **Cost-based methods**: valuing data according to the costs incurred to collect, store and analyse the data;
2) **Market-based methods**: using the market prices of data or market valuations of companies which use data intensively;
3) **Use-based methods**: a broader group of methods which aim to estimate the value to businesses (in terms of profits or productivity) or to consumers (in terms of willingness to pay) of using data.[1]

Each method can provide robust valuations depending on the aim of the valuation exercise:

- **Cost-based methods** are best for measuring the **value of data across the entire economy**, because of their relative ease and feasibility of implementation on a large scale.
- **Market-based methods** are often used to measure the economy-wide value of assets, but data is sold and bought relatively rarely. The application of market-based methods to data is therefore limited but could be used to assess the value of:
  - data which is exchanged through market transactions;

[1] These are sometimes referred to in the literature as "income-based" or "revenue-based", normally when applied to businesses.

  - data which has relatively close comparators that are exchanged through market transactions; and
  - data assets held by data-intensive organisations, from market valuations.
- **Use-based methods** can take account of the contextual factors of data use such as:
  - how the data is used;
  - in what context; and
  - for what purpose.

Use-based methods are therefore the most broadly applicable. However, this flexibility can also mean that it can be relatively resource-intensive to apply them.

The literature we reviewed and our interviews with stakeholders indicate that very few organisations currently estimate the value of their data. Therefore, assessing the value of data (e.g. the value of data used in a particular sector) cannot rely on existing information.[2]

2: A framework for assessing the impact of changes in data assets

There is no particularly meaningful way to think about the value of one (or more) specific characteristics of data in isolation. This is because:

- The characteristics of data assets are best valued as a "bundle", rather than in isolation. Organisations rarely value their data assets and they find it difficult and not particularly meaningful to think about the value of a characteristic in isolation;
- The value of a data asset is not driven solely by its characteristics. For example, the use of recognised recording standards for data is vital for creating value in some circumstances and not in others;
- Changes to one characteristic of a data asset typically also alter other characteristics. For example, investing in the accuracy of information could also change the extent to which it is interoperable;
- The value of data assets and their characteristics depends heavily on the purpose for which the data is used. Therefore, assessing the value of a change in data assets requires an understanding of how the data is used and whether the change will enable new uses of the data; and
- Extracting value from data generally requires complementary investment (in terms of people/hours and other resources) – in particular, investment in other intangible assets ranging from research & development to design and business process engineering – and a set of organisational characteristics.

Based on these findings, we propose a conceptual framework to assess the benefits and the costs of changes to data assets, focusing on the questions highlighted in the Figure below.

[2] For example, surveying a representative sample of manufacturing companies and asking them to report anonymously their valuation of the data they hold.

A. Changes in data characteristics

1. How could this change affect the value of the data asset? Does this change affect:

- the **demand for / productivity of** the data asset;
- the **supply of or the availability of** the data asset;
- the **sensitivity and the confidentiality** of the asset?

2. Which organisations and data assets are affected?

Changes in data assets can be:

- **Intrinsic,** where the change affects a specific data asset and occurs within the boundaries of an organisation; or
- **Extrinsic,** where the change may affect several data assets and involve interaction between different organisations.

B. Use cases

3. How is value currently extracted from the data asset?

- **To sell or license** data (the organisation sells the data itself without any analysis);
- To sell entirely **new data-related products** (not selling the data itself but a product fully based on the data);
- To **improve or enhance existing** products; or
- To **improve overall productive** capacities and efficiency.

4. Does the change enable new ways to extract value?

Changes in data assets can be:

- **Incremental** (relative to existing use cases).
- **Innovative** (enabling completely new use cases).

C. Organisational characteristics

5. What are the characteristics of the organisations that use (or could use) the data?

In particular:

- Size of the firm: a key determinant of the capacity to extract value from data assets;
- Level of productivity in the firm: cause and effect of high-value data assets;
- Organisational culture: an enabler and a multiplier of the value of a data asset.

Value of changes in data characteristics

By answering the questions in the figure, the framework helps policy-makers to navigate the most effective way of valuing the change, using one or more of the methods above.

Opportunities for further research

Valuing data assets and their characteristics poses distinct challenges compared to other asset types and public policy areas. Further research on this topic would help with understanding how these challenges can be overcome. It would also provide new evidence on the value of data to the UK economy and on the value of investing in data assets.

There is potential for research into the deployment of the methods outlined in this report:

- Cost-based, market-based and use-based methods could all be deployed to obtain a more up-to-date and comprehensive picture of the value of data in the UK economy (although it should be noted that market-based methods would need to be applied to particular types of data and/or industrial sectors where data is sold and bought).
- Use-based methods could be applied to quantify the impact of groups of characteristics on the value of data assets – for example, evaluating the impact of increases in the quality, quantity or usability of information included in the data.[3]

There is also potential for research that expands or builds on the evidence gaps outlined in this report:

- As described in the above framework, it is useful to define the economic purposes for which the data is used. However, relatively little is known about the size and features of economic activities which involve the sale of data and of data-driven products. The concept of data-

[3] Usability can refer to how easy it is to use the data within an organisation and/or to how easily the data can be exchanged through market transactions.

driven products is frequently used in the economic literature on the value of data, but, to the best of our knowledge, there is no operational definition of a "data-driven product" that could be used in applying our framework. Further research could aim to provide this definition, helping to distinguish cases where a data asset is essential to the provision of a product from cases where data helps to improve the quality or reduce the cost of the product. The distinction is important because, if the data is essential, all or at least a significant part of the revenues from the sale of the product could be considered as economic benefits generated by the data. If, instead, the data is not essential, the benefits generated by the data amount to the quality improvement or cost reduction that can be achieved from using the data.

- In light of the importance of the characteristics of organisations which use (or could use) the data discussed in section 3.6, research on the impact of organisational characteristics on the value of data assets could be a priority in this policy area and could focus on:
  - Size of the firm;
  - Level of productivity in the firm; and
  - Organisational culture.

1.INTRODUCTION

1.1 Context

The UK Government's National Data Strategy (NDS), published in September 2020, set out five priority areas for the government to support the use of data across the UK economy. The government's response to the NDS consultation, published in May 2021, identified a number of actions under each priority area. This report aims to support the Department for Digital, Culture, Media and Sport (DCMS) in defining, appraising and evaluating policy interventions around data, particularly those relevant to the priority area, to "unlock the value of data across the economy". This study specifically aims to improve the DCMS's understanding of how to value data assets and changes in the characteristics of those data assets.

Economists and policy-makers have developed a good understanding of value in important policy areas like health (where Quality Adjusted Life Years and other measures are regularly used), education (with various estimates of the economic returns of different levels of educational attainment) and transport (where measures like the value of statistical injury are often used). However, to date, we do not have the same level of understanding of the value of data. This report aims to reduce this evidence gap and to explore:

- The characteristics of data assets;
- The other drivers of data use;
- The marginal cost/benefit of changes in those drivers; and
- How these complex and interlinked dynamics can be quantified and valued.

1.2 Objectives

The wider aim of this work is to support DCMS in developing a framework to assess the impact of its policy interventions on data. The study does so by beginning to answer the following questions:

1. How can the private value of data assets be assessed?
2. How do different characteristics of data assets affect their value?
3. How can we assess the impact of policy interventions which may cause changes in the characteristics?

Given the breadth of data assets, data uses and potentially relevant policy interventions, this study cannot fully answer these three questions. Instead, it seeks to:

4. Provide an up-to-date review of existing evidence on the valuation of data assets; and
5. Build on this evidence by providing recommendations for:
    - policy appraisal and evaluation
    - further research.

This research was conducted in parallel with another project commissioned by DCMS in the same period (DCMS reference number: 102220), which sought to understand the externalities of data use. Although the two projects were delivered independently, there were opportunities to discuss, coordinate and align the two reports to ensure they covered complementary issues, without unnecessary duplication of messages and recommendations.

1.3 Methodology

Given the very abstract and conceptual nature of the objectives of the report, and the lack of suitable data and quantitative information on the topic, the methodologies adopted in this report are mainly qualitative:

1. **Literature review:** covered grey literature, sector research and academic literature in the English language. The review which underpins this report can be classified as non-systematic.
2. **In-depth expert interviews:** conducted with academic and sector experts from the Open Data Institute (ODI), the Office for National Statistics (ONS), the Organisation for Economic Co-operation and Development (OECD), the Bennett Institute for Public Policy at the University of Cambridge and Frontier Economics (i.e. experts in regulated sectors, such as telecoms and energy, not directly involved in the production of the report).
3. **Proof-of-concept interviews** with a small sample of organisations, covering a range of different sizes and sections and active at different stages of the data value chain.

1.4 Key terms

Before discussing the details of data assets, their characteristics and their value, it is important to clarify some key data-related terms which are used throughout the report. A full glossary of definitions is provided in Annex C.

Data is defined as characteristics or information, usually numerical, which are collected through observation. Data is the physical representation of information in a manner suitable for communication, interpretation or processing by human beings or by automatic means.[4]

In this report, we are interested in data as an asset: that is, cases where data can provide an economic benefit over a period of time, as opposed to cases where data only provides an immediate benefit (data as a good). However, it is challenging to infer where data is best characterised as an asset rather than a good. Therefore, we generally use “data” and “data asset” interchangeably throughout this report.

1.5 Structure of the report

Section 2 includes our findings on the valuation of data assets:

1. We define three categories of valuation methods: cost-based, market-based and use-based methods;
2. We describe how the literature has applied these methods, where and how these methods can be applied, and their key advantages and disadvantages; and
3. We make recommendations for further research on valuing data for policy-making purposes.

[4] This is the OECD definition of data (OECD, 2006).

In section 3, we focus on the characteristics of data assets and of the organisations which use them. We present:

4. A comprehensive list of all the characteristics which drive the value of a data asset identified by the literature;
5. A way of classifying changes in these characteristics which can help with assessing their potential role in determining the value of a data asset;
6. Guidance on how these groups of characteristics can impact the value of the asset and what needs to be considered in this context; and
7. Guidance on what valuation methods might be better suited to what circumstances and why.

In section 4, we apply this consolidated conceptual framework to two examples of hypothetical data policies:

1. A data trust which integrates and analyses data from several contributing organisations in a given industrial sector; and
2. A grant to fund the digitisation of paper documents.

These examples are purely hypothetical policy interventions which could be taken forward under one of the "levers" for intervention identified in a recent report commissioned by DCMS to identify how government could increase the availability of data in the UK economy (Frontier Economics, 2021). The application of our framework in section 4 is not intended to provide a prescriptive approach to appraisal and evaluation of data policies and it does not reflect official government policy.

Finally, section 5 presents our conclusions and highlights areas for future policy-relevant research.

2. VALUATION METHODS FOR DATA ASSETS

Key findings:

- Cost-based methods are best suited to very broad valuations of data (e.g. across the whole economy), cases where the value of data assets is less context dependent and relatively stable over time, and cases where the data is *not* created as a by-product of economic activity. As they focus on the costs of creation rather than overall benefit to the organisation, they likely provide a lower-bound estimate of the value of data.
- Market-based methods are a reliable option for valuing data assets when there is a significant volume of transactions. However, market prices for data are often not available and it is difficult to disentangle the value of data from other digital and intangible assets. These methods are therefore likely to provide an upper bound for the true value of data to an organisation.
- Use-based methods are the most appropriate for assessing the impact of data policy changes because of their flexibility and ability to account for differences in dataset values at a more granular level. However, the current empirical literature remains under-developed and will need to grow further before use-based methods can be consistently applied to appraising a range of data policies.

This section presents the benefits and the limitations of different methods for assessing the private value of data.[5] This overview is relevant to DCMS from two different policy perspectives:

1. For the measurement of the value of the data assets owned and used by different organisations across the economy, which could in turn inform a wide range of data policies; and
2. For the appraisal of specific data policies, affecting a subset of datasets and organisations in the economy.

How to value data across the full range of datasets and use cases is a complex question that has not yet been answered comprehensively by existing research. There is no empirical work which applies different valuation methods to the same assets, and so it is difficult to compare results across methods. This is, at least in part, because different methods are applicable to different assets and different circumstances.

From a general perspective, there are two main challenges in assessing the value of data, compared to more "traditional" assets.

First, the value of data is highly context-specific and closely tied to the process by which the data was generated and collected and how it is used. Even for a specific use case, for example using data to target pricing offers to consumers, the data may be more valuable to organisations which sell their products online (where the prices displayed can vary depending

[5] This includes research by practitioners and in the academic and grey literature. Where relevant, we also draw on the interviews with private sector organisations conducted as part of this study.

on who is visiting a website) than to organisations which sell offline (where the prices displayed cannot vary depending on who is in the store).

Second, there is often greater uncertainty about the future value of data assets compared to other types of assets. Patents or machinery, for example, are expected to depreciate over time as the asset becomes obsolete. Conversely, the value of data assets may decrease or increase[6] as use cases change over time.

Against this backdrop, it is worth noting one key finding from our interviews: very few organisations confirmed that they currently value their data and those that do are exclusively data-enabled businesses – that is, organisations where using data is a core part of their business model. These data-enabled businesses value their data assets mainly through use-based methods.

2.1 Categories of valuation methods

The literature on data valuation sets out a variety of methods for valuing data, which can be categorised into three groups:

1. **Cost-based methods**: valuing data according to the costs incurred to collect, store and analyse data;
2. **Market-based methods**: using the market prices of data or market valuations of companies which use data intensively; and
3. **Use-based methods**: a broader group of methods which aim to estimate the value to businesses (in terms of profits or productivity) or to consumers (in terms of willingness to pay) of using data. The literature often refers to these methods, when applied to businesses, as "income-based" or "revenue-based" methods.

This grouping is broadly consistent with categorisations proposed by academics and practitioners.[7] Annex G gives some examples of how each method has been applied in practice to measure the value of data assets. The following subsections summarise the main benefits and limitations of each approach.

2.2 Cost-based methods

Cost-based methods are widely used to value assets in companies' financial statements and in national accounts. This involves valuing assets according to:

- **Historic cost:** the cost incurred in producing the asset; or
- **Replacement cost**: how much it would cost to reproduce an asset at the time of the valuation.

[6] As stated in Li et al. (2019).

[7] For example, Nguyen and Paczos (2020) propose the breakdown between revenue-based, cost-based and market-based methods, although these are presented and discussed in a slightly different way. Revenue-based methods specifically focus on data monetisation rather than inferring the value of use cases for data. This study also concludes that there is no "one-size-fits-all" method for measuring the value of data and that more research on this topic is needed. Coyle et al. (2020) also list the three methods mentioned above for valuing data, with the slight difference that they refer to income-based methods instead of use-based methods. On balance, we consider income-based methods to be a subset of use-based methods as data can be used to generate income but also for a variety of other purposes. See also, Global Partnership for Sustainable Development Data (2018).

In the context of data, these methods would involve valuing datasets according to the costs at all stages of the data lifecycle.[8]

The central issue for both cost-based methods is understanding which costs should be counted towards the value of an asset and how they vary over time.[9]

These methods generally use organisations' financial and accounting information about their costs. However, the stakeholder engagement undertaken as part of this project indicated that organisations rarely value their data. Therefore, applying cost-based methods to valuing data would often require gathering new information on the costs of collecting, manipulating and analysing data. This information could be collected through surveys or could rely on labour cost data from online sources, as described in focus box 1 in Annex B.

Interviews with organisations which use data or with experts could be useful for understanding how costs should be allocated across different assets. For example, a company building a bespoke customer relationship management (CRM) system could be investing in creating and maintaining data on its customers as well as in building new software.

Benefits and limitations

Figure 1 presents the benefits and limitations of using cost-based methods to value data, based on a review of the literature and interviews with academic and regulatory experts.

Benchmarking across comparable organisations (e.g. businesses of similar sizes in the same industrial sector) could help with assessing the efficient proportion of labour costs specific to data. The estimated proportion could then be applied to labour cost data in an industry to produce cost-based estimates of the value of data in a given sector.[10]

Summary

Cost-based methods are best suited to the following cases of valuation:

- Very broad valuations of data (e.g. across the whole economy). Using data on labour costs from surveys or other sources such as data on online job postings could be a fruitful avenue for further research;
- Cases where the value of data assets is less context dependent, either because assets have more limited use cases or existing use cases differ less in terms of value;
- Cases where the value of data assets is likely to be relatively stable over time; and
- Cases where the data is *not* created as a by-product of economic activity.

Conversely, cost-based methods are less well suited to the following cases of valuation:

- Where the goal is to compare the value of data assets between different organisations, types of data or sectors of the economy, particularly when the collection methods differ across these units. This means that cost-based methods are generally less useful than other methods in assessing the impact of policies which aim to determine changes in the characteristics of data assets;
- Where the goal is to assess the potential value of data for new applications; and

8 Including collecting, storing, processing and analysing. See, for example, Government Data Quality Hub (2020).

9 More details on variation over time and amortisation are provided in Annex B (focus box 5).

10 Annex B (focus box 2) provides an example of how benchmarking can be used to value water company assets.

Where it is particularly challenging to apportion costs to the creation of a data asset (e.g. where the data is collected as a by-product of economic activity, or where it is collected along with the creation of other intangible assets such as bespoke software).

Figure 1: Benefits and limitations of cost-based methods

Type of benefit or limitation	Description
	Benefits
Widely applicable	Cost-based methods are widely applicable to different types of industry and datasets, as shown for example in Goodridge & Haskel (2015).
Consistent with valuation approaches for other assets	These methods are already used to value assets extensively across regulated sectors such as energy, telecommunications and water.
	Limitations
Challenging to define relevant inputs (allocating costs specific to creation of data asset)	It can be challenging to determine which costs have contributed to the creation of an asset. Goodridge & Haskel (2015) and Statistics Canada (2019) rely, respectively, on expert input and on subjective assumptions to assess how much time digital occupations such as “Software professionals” or “IT operations technicians” spend building data assets.
Challenging to value replacing assets that were created as a by-product	Cost-based methods can significantly underestimate value when data is created as a by-product from other activities and therefore the cost of producing the data is very low.
Does not allow value to differ by use case	Assets with similar costs but very different use cases would receive a similar valuation. For data where the context of a dataset’s use case and the process of generating data are important for generating value, cost-based methods may significantly misstate asset values.
Limited precision (likely to underestimate the current value of data)	Organisations are only likely to invest in an asset if it will generate greater value than it costs to create. As a result, cost-based methods are likely to provide a lower bound and understate the value of data.
Potentially a weak indicator of future value	Cost-based methods tend to focus on how much it did cost to generate the asset or it would cost to replace an asset today. This does not allow for changes in value that could occur through changes in future costs or use cases. This issue is exaggerated for data assets whose future use cases tend to be more uncertain than for other assets.
High informational requirements	May require detailed firm-level information that can be commercially sensitive.

2.3 Market-based methods

The literature has identified two main market-based approaches:

- **Using market prices of data:** This is a “bottom-up” method for valuing data based on the market price for datasets; and
- **Using market value of companies:** “Top-down” market-based options include attributing part of the market value of data-intensive companies to data or calculating the difference in the market value of data-intensive and non-data-intensive companies. A further option is to review the price paid for data-driven businesses in recent transactions.

In terms of data collection and sources, these methods mainly rely on transaction data. Some of this data is publicly available (e.g. data on the market valuation of stocks listed on financial markets or in the case of companies which publish prices on their websites). However, the prices charged for data and valuations of equity exchanged outside of financial markets are not always publicly available.[11] They are often confidential, as firms consider the prices at which they sell and buy different data assets as well as the value of private transactions to be commercially sensitive information.

Benefits and limitations

Figure 2 presents the benefits and limitations of using market-based methods to value data, based on a review of the literature and interviews with academic and regulatory experts. As there are different types of market-based methods, this table includes an additional column (labelled "approach"), which is not present in Figure 1.

Furthermore, in contexts where there is a market for data assets, it is likely that these assets will be excludable (fully or partly); if data was fully accessible to everyone then there would be no market. If the objective of a policy intervention is to reduce the exclusivity of a data asset and to increase open accessibility, a market-based method is unlikely to capture the impact of the policy on the private value of the data assets involved. Indeed, if the market value of the asset is associated with the competitive advantage that it generates, a change in its availability/accessibility will decrease (or eliminate) this competitive advantage and, hence, materially change its market value.

Summary

When applicable, market-based methods are a reliable option for valuing data assets when there is a significant volume of transactions. However, market prices for data are often not available and market valuations of data-intensive companies are only available for a limited number of businesses.

Furthermore, because of difficulties in disentangling the value of data from other digital and intangible assets, these methods are likely to provide an upper bound for the true value of data to an organisation, as shown in focus box 3 in Annex B.

When it comes to assessing the effects of data policies, market-based methods can be useful:

- They can help target interventions which could focus on a particular data market. Simplifying, an intervention that could lead to an X% increase in the value of a £1 billion market would have a larger impact than the same X% increase in a £100 million market. Having market-based estimates of the value of two data markets enables this comparison if both markets are sufficiently liquid.
- They can provide benchmarks for the potential impact of market-making interventions. If the objective of a policy is to help facilitate market exchanges of certain types of data, the value of data exchanged in other markets could provide a useful benchmark. However, this benchmarking exercise may not always be possible or useful, depending on whether and how the new market which the policy targets is comparable to existing data markets.

[11] Indeed, our own work on the geospatial data market shows that prices for geospatial data are often not disclosed publicly by geospatial data providers (Frontier Economics, 2020).

Figure 2: Benefits and limitations of market-based methods

Approach	Benefit / limitation	Description
		Benefits
Market prices	Simple to calculate	Using market prices, if available, may be the simplest way to calculate the value of datasets.
Market prices and market value	Reflects benefits of data use	Compared to cost-based methods, these methods reflect the value the market assigns to data. Therefore, these methods are less likely than cost-based methods to underestimate the value of data.
Market value	Relatively simple to calculate	Provided that identifying data-driven businesses is a simple exercise, comparing the market value of firms is a relatively simple method to execute.
Market value	Wider scope of use cases	In principle, this approach accounts for the full variety of ways that data adds value to a business. For example, if data indirectly adds value by complementing other inputs to production, the contribution of this indirect benefit would be valued in an analysis of market value.
		Limitations
Market prices	Not widely applicable	Not all data is exchanged through market transactions. In fact, this is likely to be relatively uncommon.
Market prices	Limited precision (may overestimate or underestimate value of data)	Because market transactions involving data are relatively infrequent, prices may be volatile due to the low number of buyers and sellers. Therefore, prices taken at a given point in time may not reliably reflect data value.
Market value	Limited precision (likely to overestimate value of data)	Comparing the market value of data-driven and non-data-driven firms is only a reliable estimate of the value of data if equity markets are fully liquid, and the sample of data and non-data-driven firms for which market values are compared is representative of the wider population of these types of firms across the economy.
Market value	Challenging to define relevant inputs (defining and identifying data-driven businesses)	In practice, defining and identifying whether a company is "data-driven" can be challenging. Company data use varies significantly and continuously along a spectrum.
Market value	Challenging to define relevant inputs (disentangling data from other assets)	These methods have been used to value companies' IT assets as a whole. It is feasible, in principle, to apply these methods specifically to data assets, but it would be challenging to disentangle the value of data assets from the value of other digital assets, and from other intangible assets (e.g. R&D).

2.4 Use-based methods

Use-based methods attempt to estimate the actual or potential return from using data (for businesses) or the willingness to pay for data or data-intensive goods (for consumers). The literature has identified four main types of use-based method:

- **Relative performance of data-intensive firms**: Econometric analysis can be used to compare the performance of more/less "data-intensive" firms. Performance measures can include productivity, margins, sales and measures of product development. This is different to the top-down market-based methods described above because it focuses on performance (e.g. profits or value of contracts won in a certain segment) and not on market value (e.g. how much the company was bought for).
- **Value of the specific benefits enabled or generated by data use**: Data also has a range of other specific use benefits which may not fit neatly into firm performance – for example, geospatial data can reduce travel times. In these cases, quantitative modelling and case studies can be used to assess the size of the use benefits (e.g. comparing the time savings from travelling with or without access to geospatial data).
- **Contingent analysis**: This focuses more directly on the benefits of data according to how it serves end consumers' demand. Contingent analysis uses surveys to ask consumers and data users about their willingness to pay for or accept changes in an asset. The relevance to valuing data is that these techniques can also be applied to assess the final value of products or services which use data as an important input.
- **Real options:**[12] it has been suggested that this method can be used to estimate the value of data where the future benefits of its use are uncertain. In these cases, the value of data can be estimated as a function of the current benefits of data use, the variance of these benefits and the cost of developing or accessing the data. This method estimates the option value of data – that is, the value of being able to purchase it in the future. In theory, this method should implicitly account for both the opportunity cost of purchasing the asset earlier and the future benefits generated by the asset.

These methods often use econometric techniques to estimate the value of an asset according to its uses. The key challenge for these methods is whether the standalone impact of an asset can be isolated from other contributing factors to firm/consumer outcomes. A potential solution is to involve stakeholders in the valuation process, as discussed in more detail in Annex B (focus box 4).

In relation to real options, one of the benefits of this method is that it is flexible and may complement other use-based methods, which means that the valuation of a dataset is more closely aligned to the context where the data is used. For example, calculating the option value of data could be a useful additional valuation approach, complementing use-based methods which value datasets according to their expected uses. However, this method is highly complex and difficult to implement due to a lack of data (and relevant markets).

In terms of data collection and sources, use-based methods can often rely on internal accounting and financial data (e.g. to measure the relative performance of data-intensive firms) as well as on surveys and interviews (for contingent analysis).

[12] In general terms, a real option is the right (not the obligation) of an individual or an organisation to undertake certain business initiatives, such as deferring, abandoning, expanding, staging or contracting a capital investment project. For example, real options valuation could examine the opportunity to invest in the expansion of a firm's factory and the alternative option to sell the factory (Locatelli et al., 2020). It is referred to as "real" because it typically references tangible assets (such as machinery, land, and buildings, and inventory) instead of a financial instrument (such as a stock or a bond). Real options therefore differ from financial options contracts as they involve real (i.e. physical) "underlying" assets and are not exchangeable as securities.

Figure 3: Benefits and limitations of use-based methods

Type of benefit or limitation	Description
	Benefits
Flexible and widely applicable	By focusing on value addition through the data use cases, these methods are more flexible than cost-based approaches. They allow for value to vary more between firms, according to how those firms use their datasets, and what the corresponding value of each use case is to end customers.
Does not systematically underestimate value	Similarly, by focusing on the value of uses of data, these methods will calculate added value from data, beyond what is simply the cost of supplying data. Contingent methods, particularly when the asset under analysis is not combined with other inputs and can be conceptually isolated by survey respondents, can accurately capture the value that individuals and organisations attach to certain assets.
Can disaggregate value of different datasets and their characteristics	Econometric studies will often be used to operationalise use-based valuation methods. These studies make it possible to estimate how the value addition of data varies by dataset and between different dataset characteristics. This is a useful exercise for organisations and government when determining what type of data policy to invest in.
Can assess potential value	The flexibility of use-based methods means that they are better able to assess the potential future value of data than cost-based methods (which do not take account of returns) and market-based methods (where returns are based on limited market information).
	Limitations
Complexity	These methods involve a greater number of calculations than cost-based or market-based methods. The additional complexity of use-based methods may require greater time and effort to tailor their application to the context of specific organisations.
May be challenging to compare across firms	It can be challenging to compare the outputs of these methods across firms as their flexibility can lead to very different approaches and outputs.
Current evidence base is limited	Operationalising use-based methods requires parameters on the impacts of data, which will usually be taken from the literature. For use-based methods to be flexibly applied, the empirical evidence used should describe how parameters vary by type of dataset, its use cases and by different types of firms. The evidence base required to operationalise use-based methods does not yet exist, although it is likely to develop in future.
Limited precision (challenging to disentangle the value addition of data)	In practice, given the complexity of the data value chain, it is challenging to determine how much of a company's performance is driven by its data, compared to other factors like intellectual property or brand. The same applies to contingent methods, where individuals (either customers or members of businesses) may find it difficult to disentangle the value they attach to a specific data asset when it is used in combination with many other inputs.

Benefits and limitations

Figure 3 presents the benefits and limitations of using use-based methods to value data, based on a review of the literature and interviews with academic and regulatory experts.

Summary

Use-based methods aim to estimate the impact of data on business outcomes (revenues, productivity, profits), individual wellbeing and other outcomes without relying on market prices for data. Therefore, these methods can be applied relatively widely. These methods are likely best suited to:

- Cases where the use of data leads to specific outcomes that may not be fully reflected in costs or market prices (e.g. time savings);
- Cases where cost-based or market-based methods are not feasible (due to the greater flexibility of use-based methods) – that is, for example, when market prices do not exist and the cost of collecting data is zero or minimal; and
- Cases where the goal of valuation is to compare the value of data across different organisations or types of data, and where the same use-based method can be applied to the different organisations or types of data.

Conversely, use-based methods are less well suited to:

- Estimating the value of data across a broad range of organisations or types of data (e.g. the entire economy). Econometric methods which estimate the impact of data on firm performance are a partial exception: these can be applied relatively broadly but they typically require the use of data to be measured through surveys. Because of this, they have been applied to specific sectors (manufacturing in the case of Brynjolfsson et al., 2011) or types of data ("online data" in the case of Bakhshi et al., 2014); and
- Cases where estimates of the value of data need to be produced relatively quickly and/or with limited access to specialist valuation expertise.

In principle, use-based methods are the most appropriate methods for assessing the impact of data policy changes because of their flexibility and ability to account for differences in dataset values at a more granular level. Furthermore, by assessing the final value of products which use data as an input, these methods will likely overstate the value of data as the entire value of the product is unlikely to be attributable to data.

However, while the current empirical literature on how data affects firm-level performance across different use cases is growing, it will need to grow further before use-based methods can be consistently applied to appraising a range of data policies.

3. A FRAMEWORK TO ASSESS THE VALUE OF CHANGES IN DATA ASSETS

3.1 Introduction

Our review of the literature identified a large number of characteristics of data (listed in Annex H), ranging from the accuracy of data to its interoperability and content (e.g. relating to whether or not the data is personal). However, this literature offers limited insight on the impact that a change in data characteristics has on value. Data characteristics can change for a variety of reasons, including investment decisions (e.g. allocation of people-hours and other choices of resources) by firms, a changing regulatory framework or because of policy interventions by government. This section aims to develop a framework for conceptualising and quantifying, where possible, the impact of changes in data characteristics on the value of data assets. The framework reflects three key findings from our literature review and interviews:

- The characteristics of data assets are generally best valued as a "bundle", rather than in isolation. Organisations rarely value their data assets and they find it difficult, and not particularly meaningful, to consider the value of a characteristic in isolation. Moreover, there can be trade-offs between different characteristics: for example, it may not be possible to achieve an increase in the timeliness of data without a decrease in the accuracy of the data. As a result, the starting point of our framework is to **group the characteristics of data according to how they might change the value of a data asset** – rather than defining a long list of data characteristics and investigating the value of each characteristic in isolation.
- The value of data assets and their characteristics depends heavily on the purpose for which the data is used. Therefore, assessing the value of a change in data assets requires an **understanding of how the data is used and whether the change may enable new uses** of the data.
- Extracting value from data generally requires complementary investment (in terms of people/hours and other resources) – particularly investment in other intangible assets ranging from research & development (R&D) to design and business process engineering[13] – and a set of organisational characteristics. Therefore, the framework takes account of the **characteristics of organisations which use the data**.

The framework is illustrated in Figure 4 (also presented in the Executive Summary) and described in detail in the remainder of this section. This framework could be developed further to guide policy appraisals in a similar way to the sections on land and asset valuations presented in section A1 of HMT's Green Book (HM Treasury, 2020).

[13] Haskel & Westlake (2018).

Figure 4: A framework for assessing the value of changes in data characteristics

A. Changes in data characteristics

1. How could this change affect the value of the data asset? Does this change affect:

- the **demand for / productivity of** the data asset;
- the **supply of or the availability of** the data asset:
- the **sensitivity and the confidentiality** of the asset?

2. Which organisations and data assets are affected?

Changes in data assets can be:

- **Intrinsic,** where the change affects a specific data asset and occurs within the boundaries of an organisation; or
- **Extrinsic,** where the change may affect several data assets and involve interaction between different organisations.

B. Use cases

3. How is value currently extracted from the data asset?

- **To sell or license** data (the organisation sells the data itself without any analysis);
- To sell entirely **new data-related products** (not selling the data itself but a product fully based on the data);
- To **improve or enhance existing** products; or
- To **improve overall productive** capacities and efficiency.

4. Does the change enable new ways to extract value?

Changes in data assets can be:

- **Incremental** (relative to existing use cases).
- **Innovative** (enabling completely new use cases).

C. Organisational characteristics

5. What are the characteristics of the organisations that use (or could use) the data?

In particular:

- Size of the firm: a key determinant of the capacity to extract value from data assets;
- Level of productivity in the firm: cause and effect of high-value data assets;
- Organisational culture: an enabler and a multiplier of the value of a data asset.

Value of changes in data characteristics

3.2 Measuring the impact of changes in characteristics

Our research indicates that it would not be feasible or particularly informative to quantitatively measure the impact that each characteristic has on the value of a data asset. This is because:

- Many characteristics drive value together as a “bundle”;
- Value will depend on many other relevant factors, such as organisational characteristics and use cases; and
- Data sources which could be used to measure the impact of different characteristics on the values of a wide range of data assets do not currently exist.

We illustrate this in Annex A through a worked example which assesses the public policy interventions that affect the characteristics of data assets.

In other words, as many of the organisations involved in our proof-of-concept exercise mentioned, data characteristics such as consistency and size are like cake ingredients: it is difficult, and often not particularly meaningful, to disentangle the impact of the flour from the impact of the eggs in determining the final quality of the cake.

Against this backdrop, valuing individual characteristics could be useful in understanding how different users value specific aspects of specific datasets. Annex G gives limited examples of successful characteristics-based valuation exercises.

3.3 Changes in data characteristics

How could this change affect the value of the asset?

This component of our framework involves analysing **how a change in the characteristics of a data asset could affect the value of the asset.** Does this change affect:

a. The **demand for/productivity of** the data asset: for example, more accurate data could increase demand for certain assets if they are sold or make them more productive if they are used within the organisation);
b. The **supply of/availability of** the data asset: for example, some datasets have more restrictions on their use than others, which makes it more costly to manage them and reduces their availability in the market; or
c. The **sensitivity and the confidentiality** of the asset: these affect the regulation to which the data is subject and the cost of complying with that regulation. However, they also indicate the value that others in society place on the data. For example, some data assets rely more on personal data (such as names, surnames and addresses) than others.

This distinction is relevant because it helps the value of the data asset to be considered from the perspective of the user of the data (demand/productivity) separately to the perspective of the owner of the data (supply/availability) and other actors in society beyond the user and the owner (sensitivity/confidentiality). In many cases, all three groups need to be considered together to understand the impact of a change in a data characteristic. However, this breakdown helps to articulate who the change in characteristic may affect and how, and therefore what makes sense in terms of valuation.

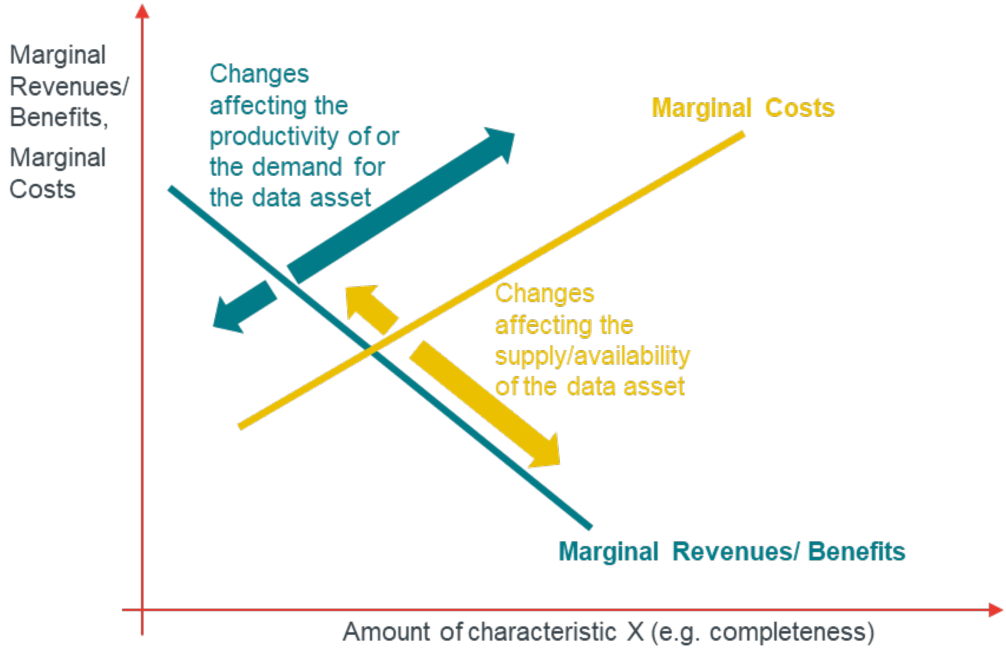
This grouping is not mutually exclusive. In fact, changes in characteristics which increase the demand for data are also likely to involve changes in its supply and availability. For example, one would expect that linking two datasets together from various parts of a healthcare system is, all else equal, more costly than in other sectors, due to the sensitivity of personal information related to health.

This classification can also be conceptualised (as shown in Figures 5 and 6 below) using a simple microeconomic framework where changes which affect the productivity of or the demand for the data asset are represented by a **marginal benefit/revenue curve,**[14] while changes affecting the supply and availability are represented by a **marginal cost curve**.

In this framework, a change which affects the supply/availability of the asset is expected to have an impact on the marginal costs curve, while a policy which changes characteristics affecting the demand for/productivity of the asset will shift the marginal revenues/benefits curve. In principle, a change in the sensitivity/confidentiality of the asset can have an impact on both marginal costs and marginal benefits.

[14] In this framework, graphically depicted in Figure 5 below, marginal revenue curves are downward sloping, under the assumption that every additional unit of the characteristic under analysis will generate a slightly lower increase in revenues/benefits. Similarly, the marginal costs curve is upward sloping, under the assumption that every additional unit of the characteristic under analysis will generate a slightly higher increase in costs. Lastly, the optimum level of each data characteristic (i.e. the equilibrium of this simple framework) is where the two curves intersect (i.e. where the marginal benefits generated by an increase in one characteristic equal the marginal costs of the same increase).

Figure 5: Simple microeconomic framework representing the impact of different changes in data characteristics on the value of data assets



Because of the challenges in measuring the impact of specific characteristics on the value of a data asset, it is often impossible to accurately estimate the marginal benefits/revenues and marginal cost curves associated with each characteristic.

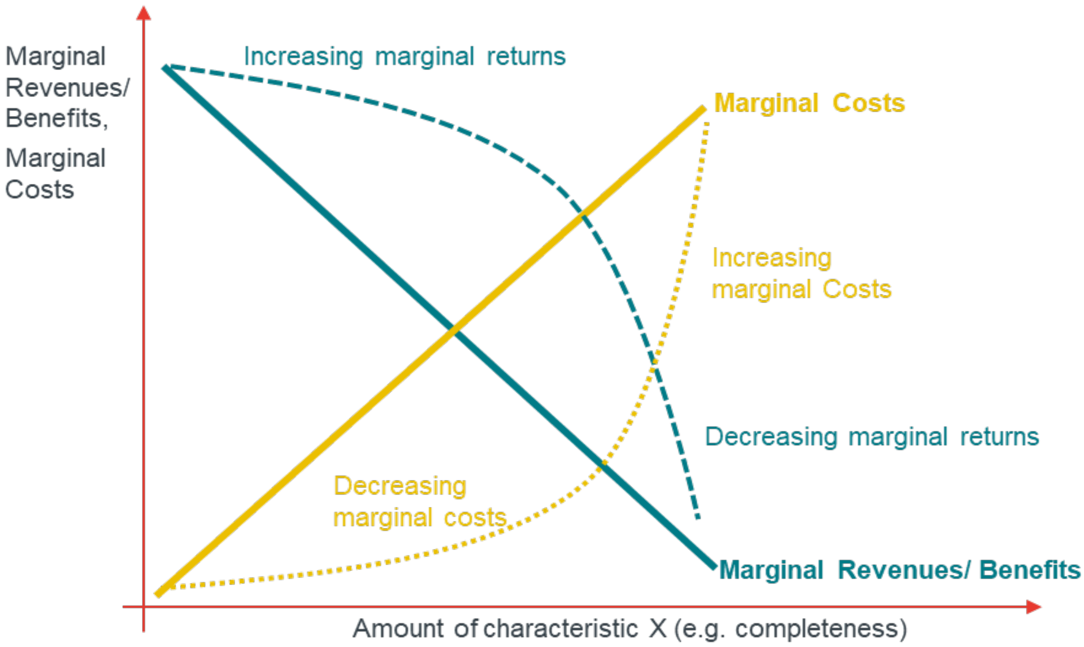
Assessing the extent to which marginal costs are decreasing, increasing or are constant is often easier and still informative for decision-making purposes.

Indeed, the relationship between data characteristics and value is not necessarily linear (e.g. every additional variable collected in a data asset does not have the same impact on the productivity/demand or on the supply/availability of the asset).

As Figure 6 shows, where the curves are flatter, an increase of Y% in the characteristic under analysis will generate an increase in benefits/returns or costs of more than Y% (increasing marginal returns). Conversely, where the curve is steeper, the same increase in the characteristic will generate an increase of less than Y% (decreasing marginal returns).

In some cases (e.g. changes in data accuracy or consistency), a positive relationship is expected (i.e. more is better and less is worse). In other instances, (e.g. changes in timeliness), the relationship is less intuitive as more timely data could be more or less valuable depending on the context and the use case, especially when the characteristic is associated with material trade-offs (e.g. more timely but less accurate).

Figure 6: Microeconomic framework with non-linear marginal revenues and costs functions



For most characteristics, one could expect decreasing marginal returns (e.g. the additional value generated by an increase in accuracy is higher for data assets characterised by a low degree of accuracy than for a dataset which is already 99% accurate). In other words, halving missing values from 10% to 5% is generally a bigger improvement than going from 0.5% to 0.25%. However, there are also reasons why this may not be the case – for example, there could be step changes where, at some point, the data becomes "good enough" for a new purpose, which leads to a jump in its value. This could be:

- A new use case (e.g. daily data is sufficient for an existing service but collecting hourly data enables the development of a new service); or
- A new method for analysis (e.g. a relatively small increase in the size of data could make it possible to apply more sophisticated statistical methods).

Moreover, marginal returns could be increasing in some cases (e.g. larger datasets increase the quality of products based on data, which leads to greater use of the product and even larger data assets).

Annex E provides more details on how certain characteristics can be measured or assessed and on the expected nature of these mechanisms, focusing on characteristics which affect the demand for and supply of data assets. In this context, it is important to flag that there is no agreement across the literature on what the relevant characteristics of data assets are, and more work is required to establish a common language and synthesise them for different purposes.

Demand/productivity-related changes

Changes which affect the productivity of or the demand for the data asset are associated, directly or indirectly, with the **benefits or revenues** that the data user can extract from the asset. In particular:

- **Demand:** Where data products or services are sold or licensed directly to customers, this affects the willingness of customers to pay for these. The same applies to assets used to create new products or to improve existing products. In all these cases, the value is “monetised” directly (by selling/licensing data) or indirectly (by using it in new or existing products).
- **Productivity:** If the data assets are used to improve an organisation’s operations and business activities (e.g. to manage its supply chain better), these changes have an impact on the productivity of the asset. In these cases, the “monetisation” of the data occurs internally and is more difficult to identify and quantify.

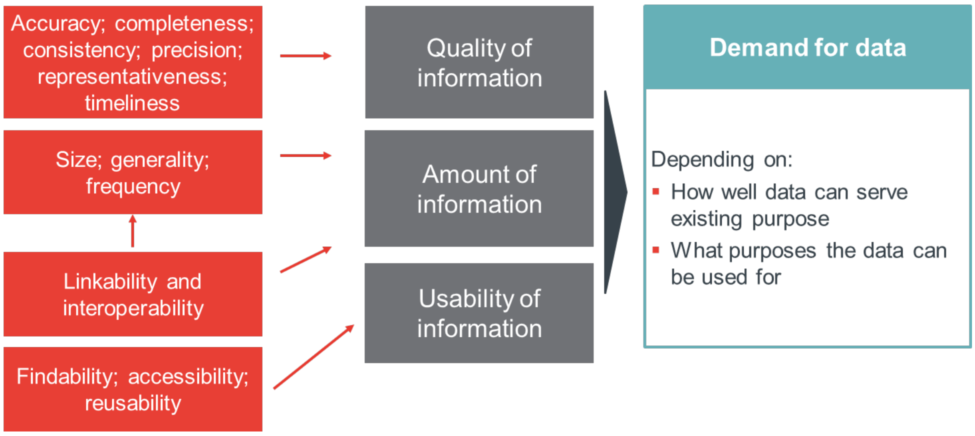
These changes are expected to shift the marginal benefits/revenues curve in Figures 5 and 6 outwards (if the change increases the productivity/demand for the data asset) or inwards (if it reduces it).

Relevant characteristics here include:

- The completeness, validity, accuracy, consistency, precision, representativeness and generality of the data, which are all aspects of the quality of the information included in a data asset;
- The findability, accessibility and reusability of the data, which influence how easily the data can be used and re-used (within the same organisation or across different organisations); and
- The linkability and interoperability of the data, which influence whether and how a data asset can be combined with others to achieve greater quality of information (i.e. greater completeness, generality and other characteristics listed above) or to generate new insights.

When it comes to the effect of characteristics on demand, it is also worth noting that linkability and interoperability can have direct value directly, by making data more usable, or indirectly (e.g. the ability to link multiple datasets together can lead to larger data assets, which include more information and from which more general and meaningful conclusions can be drawn).

Figure 7: Impact of data characteristics on demand for data



Supply/availability-related changes

Changes which affect the supply and availability of the data asset determine how complex (and therefore costly) it is to generate, collect, replicate, maintain and give access to the asset. This **complexity and the associated costs** will have an impact on value.

As a result, these changes are expected to shift the marginal costs curve in Figure 5 outwards (if the change increases the supply/availability of the data asset) or inwards (if it reduces it).

For example, changes in the timeliness of a data asset (say, delayed by 15 minutes to real time) can make it more complex to generate and maintain the data asset and, as a result, affect its cost and availability on the market.

Regulation/sensitivity-related changes

Lastly, **changes which affect the sensitivity and confidentiality of the asset have an impact on the type of regulation it is subject to and the compliance processes that need to be followed.**

These changes affect the operational, legal and reputational risks associated with data use. Data about individuals' health or sexual orientation, for example, is valuable in the sense that individuals want to protect this information and attach a value to this protection. In this context, best-practice frameworks, ethical guidelines and other non-regulatory codes which inform data collection and processing are relevant points of reference for understanding the impact that changes in these characteristics may have on value.

While this dimension is less immediately relevant to the question of how organisations (could) value their data assets, it has a direct impact on them when it is reflected in regulation.

On the one hand, regulatory compliance (e.g. General Data Protection Regulation (GDPR)) generates costs which can decrease the value of the data asset from the perspective of the organisation. Furthermore, as focus box 6 shows, the operational, legal and reputational risk associated with non-compliance could generate a variety of costs which need to be reflected in the value of the data asset under analysis.

On the other hand, it can increase the scarcity of certain data assets (e.g. assets containing personal data, which are collected less often by certain organisations), making them more valuable on the market or for the organisations who own them, or for specific purposes (e.g. research or internal monitoring of diversity and inclusion). These two examples highlight that, depending on the context, changes which affect the sensitivity/confidentiality of a data asset can affect both demand and supply considerations and have an impact on the costs and the benefits generated by a data asset.

In this context, the "quantity" of sensitive elements of a data asset is not particularly relevant, as a series of processes needs to be followed to comply with regulation regardless of the amount of sensitive information in the asset. More specifically, once data is classified as personal data, access to and sharing of this data is predominantly governed by the applicable privacy regulatory framework. As a result, the costs associated with the generation and management of personal data are semi-fixed (i.e. once one piece of personal information is included, the costs are higher regardless of how many additional personal datapoints are added). This is true irrespective of the collection sector, processing and (re-)use of the data, even if different privacy regulatory frameworks apply across these sectors.[15]

Which organisations and data assets are affected?

Lastly, it is important to understand whether the change under analysis can be classified as:

[15] Some suggest that this dichotomy (personal vs non-personal data) is not reflective of the risks associated with different data assets and that a focus on the degree of identifiability would capture this more accurately (Nguyen & Paczos, 2020).

a. **Intrinsic to the organisation**: This is a change which affects the intrinsic value of the data asset and which does not depend on other data assets used or owned by other organisations (e.g. a data asset with more observations or more consistent); or
b. **Extrinsic to the organisation**: This is a change which affects the interaction between the data asset under analysis and other datasets and environments owned or used by other organisations (e.g. interoperability between firm-level business addresses and geospatial data).

A change in data completeness is intrinsic because the impact of this change on the value of the asset occurs within the data asset itself (i.e. the data is more complete and therefore the insights that can be derived are more accurate and the value is higher). Conversely, a change in data interoperability is extrinsic because the mechanisms through which this change has an impact on value are related to other data assets (i.e. the data is more interoperable, it can be linked with other data assets, and only then can it generate more accurate insights and therefore more value).

This distinction is useful for understanding whether extracting value from this change in the data asset requires interaction between different organisations, and where and how the change under analysis can unlock value (see section 3.6 for more details on the policy implications of this distinction).

3.4 Use cases

Our literature review and the engagement with sector experts indicated that the value of data assets and their characteristics are highly dependent on how the data is used. The same data may have no value in one context while being very valuable in other cases.[16]

Use cases can be defined as an application of data and analytics to improve business activities and performance (TDWI, 2015). While uses of a data asset refer to any way in which the data can be used (e.g. to measure the correlation between variable y and variable x), use cases refer to how the data is used in the context of the organisation's value chain (e.g. to understand whether rainy days slow down deliveries of a logistics firm).

To value a change in the characteristics of a data asset, it is necessary to understand how value is currently extracted from the data, and how that could change as a result of the change in characteristics.

How is value currently extracted from the data asset?

Our literature review identified several taxonomies which classify how data is used. In our view, the most useful for valuation purposes is provided by Nguyen and Paczos (2020).[17] This is because they focus on how value is extracted from the data and they identify four main types of organisational use of data which can generate value (i.e. excluding uses which do not generate any yield, such as regulatory compliance):

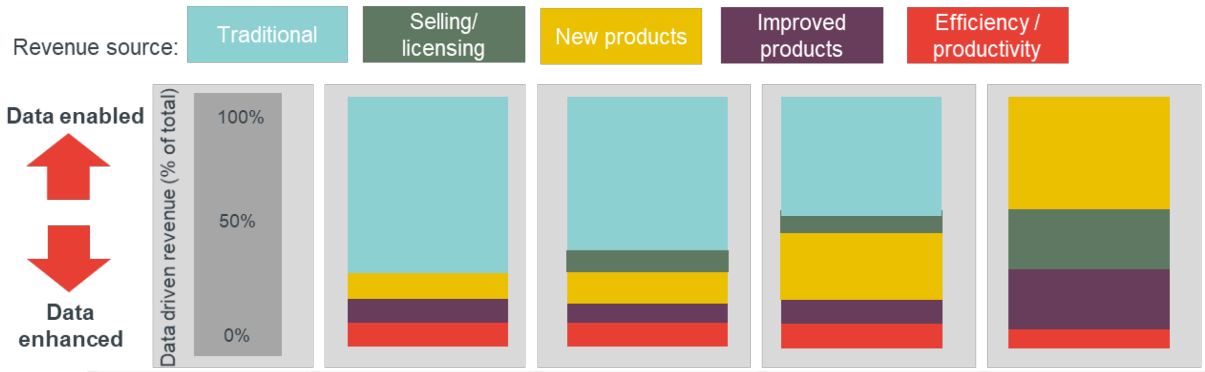
a. **To sell or license** data – for example, a vendor of financial data selling real-time data on commodity prices and volumes. In this case, the organisation sells the data itself without any analysis, processing or insights);

[16] Further details on this from our stakeholder engagement are provided in Annex B (focus box 7).

[17] Alternative taxonomies include PwC (2019) (which distinguishes between market-related use cases, competition-related use cases and internal uses) and Mydex CIC (2019) (which distinguishes the following uses of data: to measure and monitor; for administrative purposes; to analyse data sets and uncover patterns and trends; to make better decisions; to implement these decisions better; and to actually deliver services.)

b. To sell entirely **new data-related products**, **including selling insight based on data** (e.g. a predictive model to forecast under which conditions different components in an engine will fail. In this case the organisation is not selling the data itself but is selling a product which is fully based and reliant on the underlying data);
c. To **improve or enhance existing** products (e.g. a smartwatch providing a new feature to predict distance covered in future weeks); and
d. To **improve overall productive** capacities and efficiency (e.g. a supermarket chain using sales data to better manage its supply chain).

Figure 8: Data monetisation across business models and sectors from Bianchini et al. (2019)



Type of business	E.g. manufacturers	E.g. Utility providers	E.g. Banking or insurance	E.g. Online platforms
Type of internal data used	• Customer data • Internet of Things (machine sensors) • Production errors	• Energy usage patterns • Payment methods	• Credit rating data • Fraud data • Payment patterns	• User clickstreams, location, language, browser type, device, tastes.
Type of external data used	• Prices of raw materials, energy • Weather and climate • Demographic data	• Energy prices, exports and imports • Real-time grid reliability	• Interest rates • Macroeconomic variables (e.g. inflation, GDP)	• Consumer surveys • Internet penetration rates • Smartphone adoption rates
Example of data monetisation	• Real-time turbine monitoring	• Peak price surcharge	• Automated fraud detection software	• Targeted advertising

In principle, all organisations could use data in any of these four ways. Organisations with more data-enabled business models typically generate a larger proportion of their revenue through the sale of data or data-related products (right-hand side of Figure 8), while organisations with other business models (left-hand side of Figure 8) are less likely to sell data and typically generate a lower proportion of their revenues by using data.

This distinction is relevant because, unlike other types of “traditional” assets such as land, data assets are excludable but non-rival. As a result, they have an extremely wide variety of use cases that are the ultimate drivers of their value, which changes at a much more rapid pace in the data economy than in other sectors. In this context, there is no established or recognised framework for understanding data asset use cases.

It is important to note that this distinction is consistent but does not overlap with the grouping of data characteristics proposed in section 3.3, where some changes in data characteristics have an impact on the demand for the data asset (which is captured by the first three use cases listed above) or on the productivity of the asset (captured by use case (d) in the list above).

Does the change enable new ways to extract value?

In relation to use cases, it is also relevant to consider whether the change under analysis can be classified as:

a. **Incremental**: if the way in which value is extracted from the data remains constant (e.g. an increase in the size of a data asset already used to target online advertising enables more accurate targeting of users); and
b. **Innovative**: if value can now be extracted from the data in new ways as a result of the change (e.g. an increase in the consistency of a data asset used for administrative purposes in the health system enables it to be used for medicine discovery purposes).

For incremental changes in use cases, the costs and the benefits associated with the change will be more predictable and quantifiable and will be related to those which currently affect the data assets under analysis. For example, a regulatory change which enables businesses to link data on their customers from different sources and to improve the granularity of their online marketing activities will generate an increase in the value of the data asset.

In the simple microeconomic framework in Figure 5, incremental changes are likely to generate a parallel shift of the same curve. In other words, the use cases are the same but they can generate more or fewer revenues and more or fewer costs.

Conversely, for innovative changes in use cases, the costs and the benefits associated with the change are much less predictable and quantifiable and mirror the challenges of valuation exercises in the domain of innovation. In these instances, as focus box 8 (Annex B) shows, the marginal benefits and cost curves are likely to change shape and position as the use cases are different to those which existed before the intervention. This again highlights the complexity of assessing innovative changes in this framework (see section 3.6 for more details on the policy implications of this distinction).

3.5 Organisational characteristics

Lastly, it is important to consider **what are the characteristics of organisations which use (or could use) the data.**

In particular:[18]

a. **Size of the firm:** a key determinant of the capacity to extract value from data assets (this is directly related to the notions of economies of scale and scope discussed in section 1);
b. **Level of productivity in the firm**: cause and effect of high-value data assets and of the relevant skills needed to extract value from data assets; and
c. **Organisational culture:** an enabler and a multiplier of the value of a data asset and of the skills needed to extract value.

Assessing these aspects is helpful because the same change affecting the same data characteristics and the same use cases could have a materially different impact on different types of organisations.

[18] This project initially focused on characteristics of data only; following initial research, it was decided to broaden the scope to include, at least partially, literature on the characteristics of organisations which use data. While the review of the literature on data characteristics and use cases has been particularly comprehensive and systematically covered in all the fields and aspects that are relevant to this subject, the literature review on organisational characteristics was conducted at a later stage in the project and is not characterised by the same degree of completeness.

It is important to note that one of the key determinants of the value of data assets is the availability of the right skills and competencies (PwC, 2019). In this framework, we consider skills to be part of productivity and organisational culture. However, in contexts where skills are particularly relevant (e.g. a firm's investment in training software or a policy funding training programmes for staff of small and medium-sized enterprises (SMEs)), this aspect may require standalone analysis.[19]

Size: a key determinant of the capacity to extract value from data assets

Organisational size is relevant from two different perspectives:

- From a cost perspective, the cost of collecting and managing data assets and complying with relevant regulation has a fixed component which disproportionately affects smaller businesses (FSB, 2018).
- From a use-case perspective, large businesses operating in a variety of different markets and segments can unlock significant economies of scope (e.g. use data collected from mobile users to inform strategies in the credit card market) which smaller businesses cannot unlock. This creates material differences in the value that different organisations can extract from the same data asset (Crémer et al., 2019).

Level of productivity in the firm: cause and effect of high-value data assets

A growing body of literature suggests that, despite a marked slowdown in aggregate productivity after the Great Recession, productivity growth at the global frontier (defined as the most productive firms in each two-digit industry) has remained robust over the 2000s. At the same time, the rising productivity gap between the frontier and other firms raises key questions about why seemingly non-rival technologies do not diffuse to all firms (Andrews et al., 2019).[20]

To explain this productivity gap, a growing body of literature has recently focused on the use of intangible capital. It finds that the productivity of a company is positively related to its intangible capital (Rico & Cabrer-Borrás 2019), the barriers to productivity growth are more pronounced for intangible intensive sectors (Arquié et al., 2019) and firms' productivity levels can be partially explained by differences in investments in intangible assets and innovative capital (Crass & Peters 2014). In addition, Bakhshi et al. (2014) find that there was a greater impact of data use on productivity in firms which already had comparatively higher levels of productivity.

As shown in more detail in section 3.6 below, the idea that differences in productivity can be caused by a difference in the ability to use and extract value from intangible assets is particularly relevant to the purpose of this report, as data assets can be classified in this category.

---

[19] This aspect is discussed in more detail in focus box 10 (Annex B).

[20] This pattern is particularly evident in the UK, where, over the ten years to 2014, companies in the top 1% or 0.1% of the productivity distribution experienced annual productivity growth of 8% and 12%, respectively. However, the growth in average productivity among the bottom 99% of the productivity distribution averaged 1%. This suggests that, among a significant share of UK companies, levels of productivity must have flat-lined or fallen, with the bottom 25% of UK companies showing levels of productivity around 80% or less than the UK median (Haldane, 2018).

A more recent and granular update of these statistics showed that, between 1998 and 2018, services, but not manufacturing, increased in dispersion of productivity and that, on average, foreign-owned firms were more productive than equivalent domestically owned businesses (ONS, 2020).

Organisational culture

The value that different organisations can extract from the same data asset will also depend on the dominant organisational culture in relation to data.[21] For example, an organisation where the importance of using data to inform day-to-day operational decisions is embedded in the culture and in the way of thinking of every member of the organisation (both operational and decision-makers) will be able to extract more value from a given data asset than an organisation where decisions are traditionally based on qualitative evidence or on the "gut-feeling" of decision-makers.

Recent literature on the topic has shown that data culture is a kind of organisational culture and that a special form of data culture is the so-called "data-driven" culture: defined as a culture which focuses on the commitment to data-based decision-making and an ever-improving data analytics process (Kremser & Brunauer, 2019). This was confirmed by most of the organisations which we engaged with as part of this project. Annex B (focus box 9) gives further detail on how organisational culture, as relevant to the use of data, can be conceptualised and assessed qualitatively.

One of the key determinants of the value of data assets is the availability of the right skills and competencies. In this framework, we consider skills to be part of productivity and organisational culture. However, in contexts where skills are particularly relevant (e.g. a firm's investment in training software or a policy funding training programmes for SME staff), this aspect may require standalone analysis.

3.6 Policy implications

We have presented a framework that can be applied to every change in data characteristics which could be caused by a variety of factors, including investment decisions by firms, a changing regulatory framework or policy interventions by the government. In this final section of section 3, we consider matters related to the latter case: changes in data characteristics caused by policy interventions.

Market failure

First, it is important to understand where and when government intervention is required (i.e. where a market failure has or is likely to emerge).

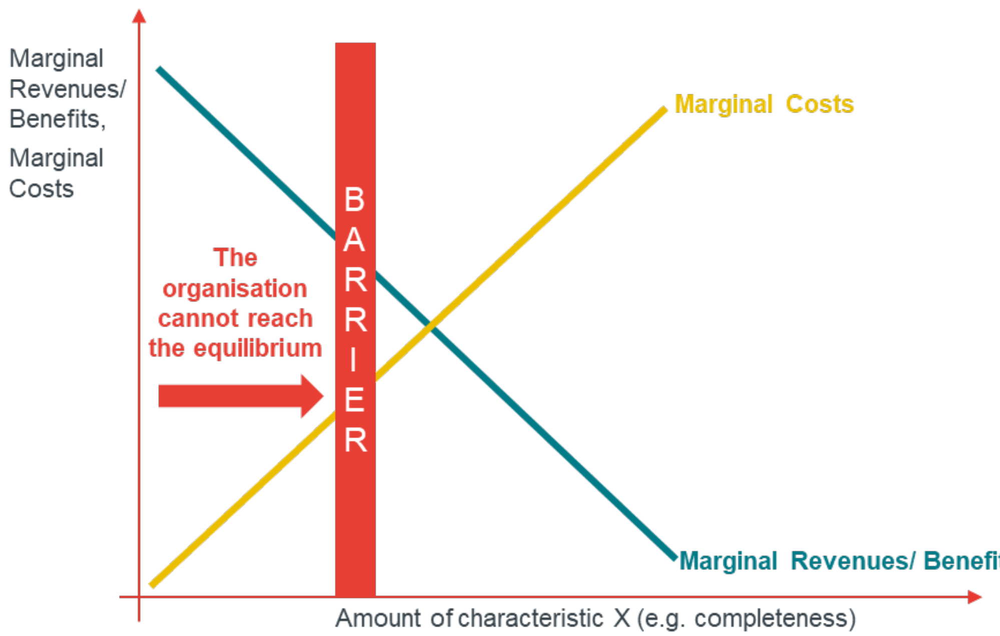
If the change in data characteristics is **intrinsic** (e.g. an increase in data precision), the decision about whether the benefits of generating more precise data assets outweigh the costs is internalised within the organisation and, in the absence of market failures, policy interventions are not necessary. However, if organisations face barriers and obstacles in making optimal investment decisions in relation to certain characteristics such as precision, policy interventions may be needed to remove these obstacles and induce organisations to optimise their investment in certain characteristics (as Figure 9 shows).

For example, a small enterprise which produces solar panels may not be able to collect data characterised by an optimal level of precision because of the high fixed costs of more precise measuring instruments or its lack of awareness of more precise products on the market.

[21] Organisational culture can be defined as a set of shared assumptions which guide behaviours in an organisation (Ravasi and Schultz, 2006): a corporate personality consisting of the values, beliefs and norms which influence the behaviour of people as members of an organization (Flamholtz and Randle, 2011).

In this context, policy interventions may be needed to remove these obstacles (e.g. by offering financing to cover the fixed costs of the instrument) to enable the organisation to reach equilibrium in this simplified microeconomic framework.

Figure 9: Organisations facing barriers in reaching optimum level of one or more characteristics

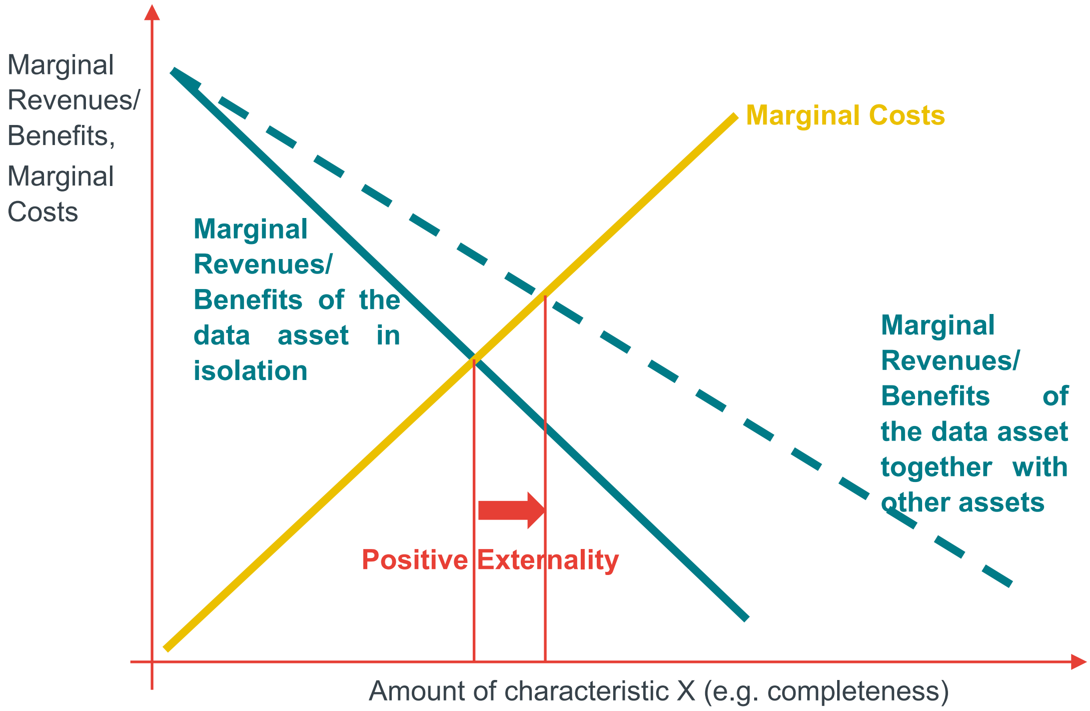


Alternatively, if the change in data characteristics can be classified as **extrinsic** (e.g. an increase in interoperability), policy interventions can add/unlock value, making it possible for the data asset to generate positive externalities within the data ecosystem or across the whole economy.

For example, as Figure 10 shows, if the marginal benefits/revenues curve shifts upwards when certain data assets (owned and used by different organisations for different purposes) are combined, the optimal amount of a specific data characteristic will be higher than the amount reached by the individual organisations which own or use the individual data assets. In this context, governments can intervene to facilitate the combination of these two data assets and unlock the positive externalities generated by this combination.

There may be instances where an extrinsic change in data characteristics increases an organisation's private return and the social return (generating a positive externality) at the same time. For example, if a pharmaceutical firm's data on the effectiveness of a vaccine becomes larger, more granular and complete thanks to a data-sharing partnership with a public health authority, this external change in these data characteristics will generate both an increase in the private value of the asset and a positive externality.

Figure 10: Positive externalities generate by external changes to data characteristics



Innovation and uncertainty

As mentioned in section 3.4, for incremental changes in use cases, the costs and the benefits associated with the change are relatively predictable and quantifiable. Conversely, for innovative changes in use cases, the costs and the benefits associated with the change are much more uncertain, as there will be no or limited historical evidence on them.

In these instances, government intervention may be necessary to take some or all the risks associated with this uncertainty and trigger the innovative uses of data assets hindered by this uncertainty. This role is very similar to the one that the public sector has conventionally played in innovation policy.

Unintended consequences

When intervening on certain data characteristics, it is important to fully understand the mechanisms through which these characteristics are expected to affect the value chain of the organisations involved. As focus box 12 in Annex B shows, particular attention should be paid to the unintended consequences/costs of interventions aimed at increasing the value which organisations can extract from data assets and which ultimately have an impact on the level of competitiveness of the market where these organisations operate.

Size

As mentioned above and highlighted in focus box 13 in Annex B, size is one of the most relevant organisational characteristics identified by the organisations we engaged with in our proof-of-concept exercise. In fact, the case study below highlights that the obstacles and the barriers that small, medium and large firms face when collecting, using and trading data assets are materially different. It is therefore particularly important to take account of these differences when designing, appraising and implementing data policies.

Productivity

In terms of productivity, from a policy-makers' perspective, when setting the objective of a policy, it is important to be clear about the target of the policy and the goal of the intervention. Is the goal to support highly productive firms in shifting the frontier of productivity even further or is it to help less-productive firms catch up with the best performers in the industry or in the economy as a whole?

Furthermore, when appraising the impact of a policy in this space, it is important to acknowledge and analyse the different impact it will have on different types of firms operating at different levels of productivity, as the same intervention affecting the same data characteristic will have a different effect on firms characterised by different levels of productivity.

Trade-offs

Lastly, data characteristics can often generate trade-offs, where one characteristic can be improved only "at the expense" of another. For example, as focus box 14 in Annex B shows, there are often trade-offs between timeliness and accuracy (ECB, 2001): where quicker production of data assets (e.g. statistics on the state of the economy in the previous quarter) is typically associated with a lower degree of accuracy. In these contexts, it is important to identify the optimal balance between two or more characteristics, depending on the context.

4. APPLICATION OF THE FRAMEWORK TO HYPOTHETICAL POLICY EXAMPLES

4.1 Introduction

In this section, we provide an initial application of the framework defined in section 3. We consider two hypothetical government interventions which could influence the value of data assets and illustrate how the framework could be used to assess the impact of the interventions.[22] There are a wide range of potential interventions which could influence the value of data assets. [23] We chose two interventions which affect different characteristics of data to better "test" the framework:

- Establishing a data trust or other type of intermediary. In this example the trust aggregates data provided by private sector organisations on a voluntary basis or for a fee. Recent such initiatives in the private and non-profit sectors include the sharing of safety and accident data in the maritime sector (HiLo Maritime Risk Management) and the creation of a hub to share, access and analyse data on animal health (Data Innovation Hub for Animal Health).[24]
- An intervention to support the digitalisation of documents held in analogue format.[25] This intervention could be targeted by sector or document type (e.g. legal documents).

This guidance is not intended to provide a prescriptive approach to the appraisal and evaluation of data policies and it does not reflect official government policy. Applying the framework presented in section 3 means answering the following questions:

- In relation to data characteristics:
  - Step 1: What characteristics are affected?
  - Step 2: Does extracting value require interaction between organisations?
- I relation to use cases:
  - Step 3: How is value currently extracted?
  - Step 4: Are there going to be new ways to extract value?
- In relation to organisational characteristics:

[22] Because of the timelines of this study, this section focuses on the characteristics which would affect the productivity of data assets (and which, in most cases, would therefore also affect the supply of the data asset) and on valuing the potential economic benefits of policy interventions. As described earlier, a full appraisal or evaluation of policy interventions should also consider the cost of carrying out the intervention and should use cost-based methods to assess the impact of the intervention on the cost of collecting, maintaining and analysing data assets.

[23] This is an area in development which could include a wide range of policies. For example, our previous study for DCMS identified six levers that could be used by government to increase the availability of data in the private and third sectors: improving knowledge and understanding; improving incentives; reducing cost; addressing risk; mandating data sharing; reducing perceived regulatory burden.

[24] Maddison and D'Addario (2020).

[25] Stored on paper.

- Step 5: What are the characteristics of the organisations which use the asset?

4.2 Data trusts for creating larger data assets

We assume that the intervention involves creating a data trust which receives data from a number of founding organisations in the relevant sector. The trust is charged with gathering data from founding organisations and combining them into one aggregated dataset. Researchers within the trust analyse the data and disseminate insights to the founding organisations. Examples of existing initiatives of this type include:

- Aggregating and using data from shipping companies to prevent accidents in the industry (HiLo Maritime Risk Management);[26]
- Aggregating and using data on animal health (e.g. from commercial poultry farms) to predict regional needs for vaccines and for treatment.[27]

Figure 11 provides a high-level theory of change for this data trust.

Figure 11 High-level theory of change for data trust intervention

Inputs and activities	Rationale for intervention	Outputs	Short- term outcomes	Long-term outcomes	Impacts
Providing guidance on appropriate governance structures for the trust. Creating a specific regulatory framework that provides clarity on how the trust can operate. Funding entirely or partially the operations of the trust.	Information failures (a lack of awareness that data can help with market-wide challenge). Coordination failures (a lack of existing industry fora or of existing governance arrangements that could support data sharing). Reputational risk and legal regulatory risk given concerns around data sharing between competitors.	1. Individual organisations' datasets are interoperable and combined into an integrated dataset.	Data trust analyses a general, complete, representative dataset.	Insights from aggregated data are provided by data trust to contributing organisations (e.g. predictions on infection in poultry farms).	Economic and social outcomes (e.g. infection spread is curtailed, with reduced costs for poultry industry, and improve animal health).
		2. Clear process to access integrated dataset.	Greater findability and accessibility of data: individual organisations and external researchers can access integrated dataset.	Individual organisations and external researchers produce additional insight from data.	Dependent on specific analysis conducted: improving productive efficiency/ improving products.

Applying the framework to data trust

Step 1: How could this change affect the value of the data asset?

Outputs 1 and 2 are changes which affect the demand for/productivity of the asset. This is because larger, more general and more representative data assets will be demanded more

26 Described in a recent case study by the Open Data Institute (ODI, n.d.)

27 DIHAD (n.d.).

and potentially more productive. At the same time, it could affect the supply/availability of the asset, in light of the delay needed to aggregate the individual components of the asset.

Step 2: Which organisations and data assets are affected?

This intervention generates an extrinsic change in characteristics by affecting the interaction between different data assets. In other words, this change generates value only across data assets and not within individual assets.

The intervention may also include developing processes under which individual founding organisations or external researchers (e.g. academics) can access the integrated dataset (or perhaps an aggregated or redacted version of the dataset which omits some of the more commercially sensitive detail).

Therefore, both outputs would lead to changes in the demand for the data asset and in the cost of providing the data.

Step 3: How is value currently extracted from this asset?

The data is used to improve productive efficiency by the data contributors (e.g. shipping companies, poultry farms). With the intervention:

Output 1 could lead to the improvement of productive efficiency – shipping companies or poultry farms can take more effective and timely action to prevent maritime accidents and to prevent the spread of disease, respectively.

Output 2 (wider accessibility and findability of data) could also lead to new uses for the data – for example, the data could be used by organisations outside the shipping or farming industries, in combination with other data sources, to generate and sell new insight.

Step 4: Does the change enable new ways to extract value?

There may be both incremental changes in value (further improvement of productive efficiency compared to a counterfactual where data is smaller, less general, complete, representative), and innovative changes (new insight being sold as a result of greater findability, accessibility and interoperability of data).

Across all outputs, cost-based methods should be used to assess the cost of achieving these outputs. Such methods are not appropriate for assessing their benefits as the intervention aims to increase the availability of data and returns from using existing data, now combined into an integrated dataset. Market-based methods are also not appropriate, as this use case does not involve selling data or insights based on the data. However, it may be possible to rely on market-based evidence from other sectors if there are examples of sales of data that are comparable to the use cases considered under this intervention.

This leaves us with use-based methods. Through output 1, the intervention leads to increasing size, generality, completeness and representativeness of data used in the relevant sector (e.g. shipping or farming). However, as described in section 3, there is no evidence on the specific impact of each characteristic in isolation, and it may not be feasible to generate such evidence. Therefore, a use-based approach to assessing this intervention could involve two types of analysis:

1. Assessing the effectiveness of a predictive model based on the integrated data generated through the intervention; and
2. Assessing how much organisations which may be interested in accessing the data maybe willing to pay for this data.

As shown in Figure 12, for both types of analysis, it would be useful to assess whether there is relevant evidence from other contexts (e.g., evidence on the effectiveness of prediction in preventing maritime accidents could provide a useful benchmark for the type of improvements that may be seen in other contexts).[28]

Figure 12 Assessment methods for data trust interventions

Output	Outcomes and impacts	Assessment methods
1. Individual organisations' datasets are interoperable and combined into an integrated dataset, resulting in increasing size, generality, completeness, and representativeness of data.	Insights from aggregated data are provided by data trust to contributing organisations. This leads to positive economic and social outcomes (reduced accidents, lower spread of animal disease).	Use-based: 1. Assess with experts how accurate the predictions generated through the data could be (e.g. the % of accidents which could be predicted correctly) and to what extent the predictions could inform action (the % of predicted accidents which could be avoided using insight from data on when and where they might occur). 2. Investigate organisations' willingness to pay: how much would they be willing to pay for the insights generated by these data if they were sold?
2. Clear process to access integrated dataset.	Individual organisations and external research produce additional insight from data.	Use-based: Investigate to what extent do organisations expect to use the integrated dataset? Would this replace or add to their current use of data? If the latter, what is the impact of additional data use on firm productivity? Is there any evidence specific to the sector relevant to this intervention?

Each of these use-based options has advantages and disadvantages. In both cases, the assessment involves a degree of judgement by stakeholders. The first method involves more detailed assessments on the specifics of predictive models. This could make the assessment relatively robust, with appropriate information (e.g. on the frequency and characteristics of shipping accidents) but tied to a specific use of the data. The second method offers greater flexibility but requires respondents to have an idea of appropriate prices for data, which is challenging as in many markets data is not often sold. However this limitation could be mitigated by providing respondents with appropriate comparators (comparing purchase of this data to other inputs) as best practice in contingent valuation. If sufficient information is available, the first method may be preferable – if not, the second could provide a valid alternative.

Step 5: What are the characteristics of organisations which use the asset?

The final step in the framework described in section 3 involves assessing the characteristics of organisations whose use of data may be affected by the intervention. The size, organisational culture and productivity of organisations are all likely to have a significant role: the impact of investing in data assets will be significantly higher when the organisations involved have greater capacity to extract value from these assets.

[28] ODI (2020).

Assessing organisational culture and organisational processes may not always be possible or feasible within the available resources for an impact assessment/evaluation. However, in this case, our framework suggests that a robust assessment should at least consider these factors qualitatively and justify why it may be appropriate or necessary to not consider them in detail.

4.3 Financial support for digitising data assets

As a second hypothetical example, we consider financial support to businesses for digitising documents and information currently stored in an analogue format.

Applying the framework to digitalisation

Step 1: How could this change affect the value of the data asset?

The main change introduced by this hypothetical policy intervention relates to the productivity of the data assets involved. Indeed, a digitised asset could be used in a more economic, efficient and effective way and could improve the overall productivity of the organisations involved.

For example, a small law firm could advise its clients on the main features which characterised past contracts previously stored in analogue format in a more timely and comprehensive way. Similarly, by digitising all the contact information of past, current and potential clients currently held in a paper format, the commercial department of a manufacturing firm could target its business development activities in a holistic and efficient way.

In some circumstances, this change could also affect demand for the asset, not just its productivity. For example, a sailmaker which has digitised information previously held in analogue format (e.g. all the episodes where a sail was damaged and sent by the customer to be repaired or replaced) may be able to sell that data to other sailmakers, to boat-building firms (who could use it to identify potential issues with boat design) or to customers themselves who could use that data to avoid and prevent the situations most likely to lead to damage in the sail.

Furthermore, this change would not have an impact only on the support of the data asset (analogue vs digital). It would also have an impact on accessibility (the data can now be accessed in different ways), linkability (it can now be potentially linked with other data sources) and location (digital documents can be stored in a different jurisdiction). It would also have an indirect impact on almost all the dimensions of data quality (e.g. accuracy, consistency, targetability, generality), in the sense that these characteristics would not change as a result of the digitisation, but it would be easier to assess them in a digitised asset as opposed to an analogue asset.

Step 2: Which organisations and data assets are affected?

Lastly, this change in data characteristics can be classified as intrinsic, as the change does not depend on other data assets or organisations.

As in the earlier example of a data trust, cost-based methods could be used to assess the cost of digitisation. However, other methods should be employed to assess the potential value of digitisation. In cases where the digitisation process will generate a data asset that is currently traded on the market (e.g. information on contracts), a market-based methodology would be appropriate. In all other cases, a use-based approach may be better suited.

Figure 13 High-level theory of change for financial support for digitising data assets

Inputs and activities	Rationale for intervention	Outputs	Short- term outcomes	Long-term outcomes	Impacts
Grants/loans or other forms of financial support for digitising assets. Training and skills development to support the digitalisation process. Sharing best practice from similar successful organisations.	In a well-functioning market, each firm chooses an optimal proportion of digitised assets, where marginal costs equal marginal benefits. However, some organisations may face obstacles and barriers in reaching that optimal amount of investment.[29]	1. Individual organisations' datasets are digitised but also more accessible, linkable and transparent in terms of quality.	Digitised assets are available in a more timely accurate and comprehensive way.	Insights from digitised assets can be derived more efficiently, economically and effectively (e.g. predictive models on causes of damage to boat equipment).	Increased productivity in the organisations involved.
		2. Individual organisations' datasets are digitised.	Digitised assets can be input into Customer Relations Management systems.	Resource management and business development activities are managed in a more economic, efficient and effective way.	Increased productivity (and potentially revenue) in the organisations involved.
		3. Individual organisations' datasets are digitised, but also more accessible, linkable and transparent in terms of quality.	Digitised assets can be traded and sold, either in isolation or as part of new or upgraded products.	Individual organisations can expand their product offering in width (different types of products) or breadth (more complete products).	Increased revenues and profitability of organisations involved.

Step 3: How is value currently extracted from this asset?

The relevance of use-based approaches in the case at hand highlights the importance of understanding the use cases affected by this hypothetical policy intervention.

In the four-category framework proposed by Nguyen and Paczos (2020), the main use case affected by this hypothetical policy intervention is in line with the fourth category: using data to improve overall productive capacities and efficiency.

Indeed, as outlined in the hypothetical examples above, in relation to a small law firm and to the commercial department of a manufacturing company, the digitisation of data assets would

[29] For example, they may not have access to credit to fund the digitisation process because of uncertainty surrounding the value of more digitised data assets. Alternatively, the organisation could be too small to sustain the investment needed to digitise its assets and be in a situation of competitive disadvantage vis-à-vis larger competitors.

mainly have an impact on internal use cases, which would make the existing activities of the firm more economical, efficient and effective.

Step 4: Does the change enable new ways to extract value?

However, in other instances like the hypothetical sailmaker example above, the newly digitised data assets would be used to improve and enhance existing products or even to sell or license the data directly to customers and other organisations.

These two alternative changes to use cases are conceptualised in a high-level theory of change framework in As a second hypothetical example, we consider financial support to businesses for digitising documents and information currently stored in an analogue format.

above.

In this context, appraisals should consider, at least qualitatively, the potential for the policy intervention to foster the creation of new products and services. If the theories of change for the intervention show that new product creation is a key intended outcome of the policy, existing evidence could help to measure the impact of an increase in firm-level investment in digitalisation on different proxies for innovation.[30]

Step 5: What are the characteristics of the organisations that use the asset?

Lastly, as discussed in section 3.5, it is important to understand the organisational characteristics which could cause a different impact of the intervention on different organisations.

With regard to size, as mentioned in the considerations on market failures, it is likely that smaller organisations would face more significant obstacles and barriers in accessing credit and opportunities to digitise their data assets. However, large firms may have much larger data assets in analogue format, the size of which could be a barrier to implementing a large-scale digitisation process.

In relation to productivity, one would expect firms characterised by comparatively higher levels of productivity to have already engaged in extensive digitisation exercises.

Lastly, in terms of organisational culture, the assessment would need to be carried out on a case-by-case basis, focusing on the history, the beliefs and the modus operandi of different sectors and organisations. For example, in the small law firm example above, one could look at the extent to which data and digital tools are already used in the organisation as a proxy for the strength of their data culture.

Similarly, in sectors which are conventionally characterised by more traditional values and ways of operating and communicating, the impact of more digitised data assets can be expected to be lower than in more open and innovation-prone sectors.

[30] For example, Mohnen et al. (2018) found that investments in ICT earn an expected average rate of return of 9.7%, followed by 6% to 7% on organizational innovation and a modest 1.4% to 1.8% on R&D in services and manufacturing respectively.

5. CONCLUSIONS

This report provides an overview of existing evidence on methods for valuing data and its characteristics. It also provides a framework which DCMS and other organisations could use to assess the economic impact of changes to the characteristics of data assets. This is a novel area with relatively few examples of empirical work applying the methods discussed in this report. We next outline suggestions for future research which could help to fill the gaps in the existing evidence base.

Recommendations for future research

Using a combination of methods to assess the value of data across the UK economy

Data on the labour employed by organisations in collecting, maintaining and analysing data has been used in the UK (Haskel & Goodridge, 2015) and elsewhere (e.g. Statistics Canada 2019) to assess the overall value of data in an economy. The approach in Haskel & Goodridge (2015) could be updated to provide more recent estimates. This approach could also be improved by using data on online job postings or new primary research to assess more precisely the average proportion of time spent collecting, maintaining and analysing data by workers across a range of relevant occupations.

Market-based approaches could provide annual estimates of the economic output generated in the UK through the sale and licensing of data. An existing study estimates this at €89 billion as of 2019 (Cattaneo et al., 2020). Generating similar estimates over time could help with assessing how this output is changing over time; it would also be helpful to understand how this value is generated: for example what types of data are sold, who are the buyers in the relevant markets? Similarly, a market-based approach could also be employed to assess the economic output generated in the UK by selling data-related products. However, this would require further research (e.g. a feasibility study) to define more precisely what should count as “data-related”.

Similarly, there is limited evidence which employs use-based methods to generate widely relevant estimates of the value of data. A possible avenue for future research could build on existing evidence in Bakhshi et al. (2014), which, to the best of our knowledge, is the only econometric study that estimates the impact of additional data use on the productivity of UK firms. A challenge for this approach is accurately measuring the extent to which businesses use data: because data is a pervasive and broad concept, organisations will struggle to answer questions such as “how much data does your organisation use?”. This challenge could be overcome in a number of, not necessarily mutually exclusive, ways, for example by:

- Asking about the use of related digital technologies, including cloud computing or machine learning (this is the approach of the United States’ Annual Business Survey, for example);[31]
- Focusing on the use of data in a particular industry (e.g. manufacturing, as in Brynjolfsson & McElheran 2016) or in a particular business function (e.g. marketing, operations); and

[31] Zolas et al. (2020).

- Measuring the use of data through information on labour costs, surveys or online data (e.g. LinkedIn, data on job postings).

Assessing the impact of data characteristics on the value of data

Econometric approaches similar to those employed by Bakhshi et al. (2014) or Brynjolfsson & McElheran (2016) could also be used to assess whether organisations which hold data of greater quality or usability perform better than other organisations. Greater quality and usability can be defined as data which has the relevant characteristics for increasing demand, as listed in section 3 of this report (e.g. generality, accuracy, completeness for quality and findability, accessibility, interoperability, reusability for usability). We would recommend focusing on overall changes in the quality and usability of data rather than trying to disentangle the impact of individual characteristics. The measurement challenge described in chapter 3 would also apply here, and to a greater extent: it would be difficult to assess the quality and usability of data employed by an organisation as a whole. Focusing on a specific type of data and on specific uses would help overcome that challenge. However, this would need to be traded off against the external validity of the study: the impact of data quality estimated in a specific context may not be representative of the impact in other contexts.

When asked about the most valuable data characteristic from their perspective, some of the businesses we engaged with mentioned linkability: specifically, the ability to link firm-level data from various sources using unique identifiers. There are many data products on the market which make it possible to match and integrate firm-level data from multiple sources. For example, Dun & Bradstreet's D-U-N-S number is used by lenders and potential business partners to help predict the reliability and/or financial stability of the companies they deal with. Similarly, the global reference data set SEDOL produced by the London Stock Exchange Group is used by investors around the world for security identification in their business activities (e.g. to assist with portfolio valuation, trade execution, processing price feeds and price validation). Analysing price developments in these markets could shed light on the value users attach to the possibility of linking and interoperating different data assets generated from different sources. While this analysis would not provide information on the value of certain characteristics in the financial sector, it could shed light on the value of interoperability and linkability in other industries.

As described in the framework put forward in this report, when valuing data assets and their characteristics, it is useful to define the economic purposes for which the data is used. However, relatively little is known about the size and features of economic activities which involve the sale of data and of data-driven products. In particular, the concept of data-driven products is frequently used in the economic literature on the value of data, but, to the best of our knowledge, there is no operational definition of "data-driven product" that could be used in applying our framework. Further research could aim to provide this definition and help to distinguish cases where a data asset is essential to the provision of a product from cases where data helps to improve the quality or reduce the cost of the product. The distinction is important because, if the data is essential, all or at least a significant part of the revenues from the sale of the product can be considered as economic benefits generated by the data. If, instead, the data is not essential, the benefits generated by the data amount to the quality improvement or cost reduction that can be achieved from using the data.

As focus box 6 in Annex B shows, data-location requirements are often associated with higher costs in terms of the collection, management and storage of data assets, which often have a negative impact on the value of the data asset. In this context, investments in research aimed at developing innovative ways to minimise these costs could be a policy option to consider in order to maximise the value of data.

Similarly, as focus box 9 in Annex B highlights, acknowledging that an accurate quantitative measurement will be impossible in most cases, even the most abstract and intangible aspects affecting the value of data assets can be conceptualised and assessed qualitatively. Depending on the industry and the context, investments in research aimed at investigating the relationship between organisational culture and the value of data assets could be a policy option to consider in order to maximise the value of data.

Furthermore, focus box 11 in Annex B indicates that some data characteristics, such as interoperability, are particularly complex to measure and assess from both a conceptual and technical perspective. Depending on the industry and the context, investments in research aimed at unpacking these complexities and at providing organisations with tools to measure these characteristics are a policy option to consider in order to maximise the value of data.

Lastly, as focus box 14 in Annex B shows, trade-offs between different characteristics often exist but can be minimised. Depending on the industry and the context, investments in research aimed at developing innovative ways to minimise these trade-offs could be a policy option to consider in order to maximise the value of data.

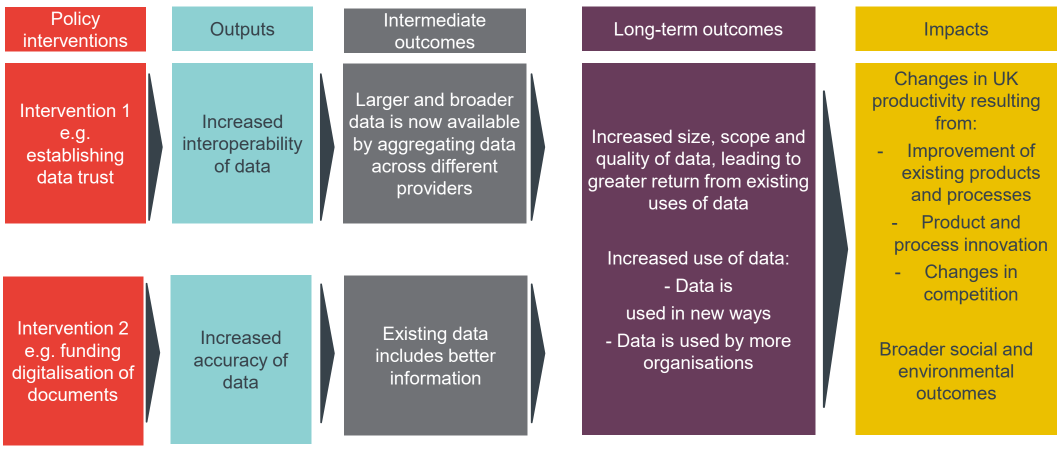
ANNEX A CHALLENGES IN ASSESSING THE IMPACT OF POLICY INTERVENTIONS

Evaluating the impact of policy interventions on specific characteristics of data, such as its timeliness or generality, can be challenging for two reasons:

- Understanding the costs and benefits of investing in specific data characteristics is likely to require detailed information which may not always be available. Moreover, these costs and benefits may vary significantly across different contexts; and
- It may be challenging to assess exactly which data characteristics will be affected by a particular intervention, and how that would differ for other interventions.

This can be illustrated through the example theories of change in Figure 14. In the simplest case, different policy interventions only affect one specific data characteristic. For example, intervention 1 could be a data trust which receives data from several providers (e.g. data on shipping accidents from separate shipping companies and ports) and makes it interoperable so that it can be analysed in the aggregate, while intervention 2 (which provides incentives for digitalisation of legal documents) improves the accuracy of information used, for example in residential conveyancing transactions.

Figure 14 Example theory of change for illustrative policy interventions

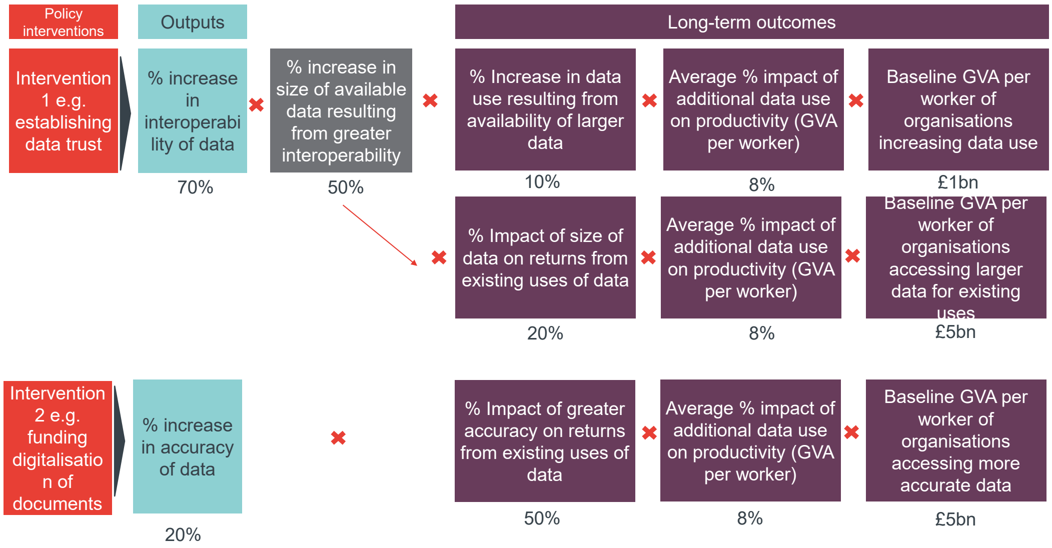


Assessing the potential impact of these two interventions would require calculations such as those described in Figure 15. In this example, the expected impact of intervention 1 on UK gross value added (GVA) is £30.8 million,[32] and the expected impact of intervention 2 is £40 million.[33] This assessment requires 10 parameters altogether if the calculation uses the same average impact of data use on productivity (8% in the example above) or 13 parameters if the impact of data use needs to be different for each of the three rows below. This illustrates the first challenge above – even in the simplest case, a quantitative assessment requires 10-13 parameters.

[32] (0.7*0.5*0.1*0.08*£1bn)+(0.7*0.5*0.2*0.08*£5bn)=£2.8m+£28m=£30.8m

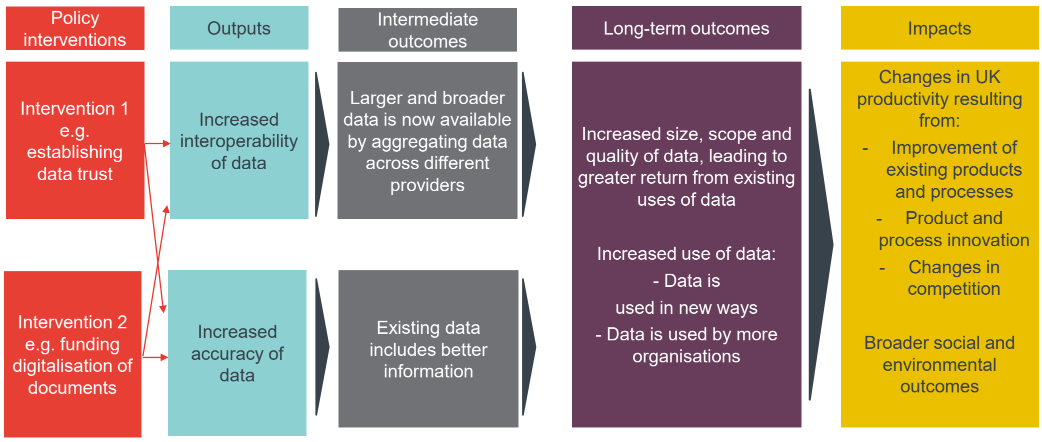
[33] 0.2*0.5*0.08*£5bn=£40m

Figure 15 Example calculation for impact of policy interventions on interoperability and accuracy of data



This is a simplified example as it is assumed that the effect of specific characteristics (size, accuracy of data) can be summarised as an average spanning all uses of data relevant to the policy intervention being evaluated. However, as discussed in section 3, the effect of these characteristics can be highly context-specific. Average estimates, even where they can be assessed robustly, can hide a significant amount of variation. Our interviews suggest that data assets tend to become more or less valuable as a result of changes in groups of characteristics (e.g. accuracy, generality and linkability taken as a whole) rather than changes in individual characteristics. Moreover, in reality, it may not be possible to anticipate precisely which data characteristics will be affected by a given intervention. As well as making data more interoperable, the data trust could give each data provider feedback which helps them make their data more accurate. Also, digitalising information could make it easier to match data across different sources, thereby increasing interoperability. This is illustrated in Figure 16.

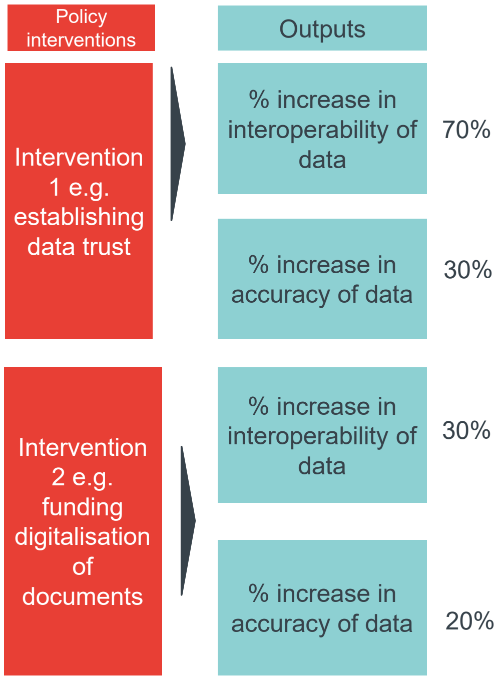
Figure 16 Example theory of change for illustrative policy interventions – alternative



In this more realistic example, the number of parameters required to assess the impact of the two interventions increases compared to the earlier simpler case. At a minimum, it is necessary to work out how much of an increase in interoperability and accuracy could result from each intervention, as shown in

Figure 17. Ideally, one would also want to adapt other parameters to each intervention, for example the % impact of greater accuracy on the returns to data use in intervention 2 (shown as 10% in the simpler example in Figure 16) could be different for intervention 1.

Figure 17 Effect of interventions on multiple characteristics of data



A key issue illustrated by this more realistic example is that the size of the expected net benefits from these interventions is likely to be very sensitive to the choice of parameters shown in Figure 16. In other words, intervention 2 could generate larger benefits than intervention 1 with the parameters shown in Figure 17, but this could change if the effect of intervention 2 on interoperability was 20% rather than 30%. Generating precise estimates of these parameters is likely to be challenging.

ANNEX B FOCUS BOXES AND CASE STUDIES

Focus on/1: using job postings data to estimate the value of data

Organisation A operates in the data and analytics industry. They suggested two different ways through which job postings data could be used to calculate the value of data assets.

The first is by estimating the value of data from salaries of data-intensive occupations. The variation of the salaries of data-intensive occupations by sector, geography or type of firm could be used as a proxy for the value of the data assets associated with these jobs. This approach could be particularly informative for policy-making purposes, as it would provide sufficiently granular estimates to target policy interventions at specific industries, areas or even firms.

The second is by identifying which occupations use data most intensively. Job postings include information on the skills required across occupations, sectors and firms, which could in turn be used to infer more precisely the extent to which different data assets are used by different industries and different types of organisations.

There are two main advantages of using data on job postings. First, it can provide a more granular breakdown of labour market information compared to conventional labour market data. Second, the data is real-time, therefore it can be used to gain immediate insights into the value of data or data use across different sectors and firm types.

KEY MESSAGE: Information on job postings (e.g. salary, competencies, number of positions) could be used to measure the value of the data assets which underpin different occupations.

Focus on/2: DORC methods for valuing water company assets (ACCC and PwC, 2008)

Asset valuation using cost-based methods is common in more traditional regulated sectors such as energy, water and others. A regulatory expert from the water sector confirmed that the rationale for these methods is that market prices for water company assets do not exist as they typically belong to state-owned monopolies and are very rarely traded between different entities.

In this context, a sensible approach to determining the value of an asset is to determine how much it would cost to produce it from scratch, and to use this estimate as a proxy of the price at which the asset would trade in a competitive market.

In water regulation, the standard cost-based method used to value assets is the depreciated optimised replacement cost, or "DORC". This method is designed to measure the current cost of replacing an asset with its modern equivalent version, with adjustments made for depreciation, optimisation and obsolescence.

These adjustments are necessary because in a competitive market a potential buyer will only bid an amount for the asset which corresponds to the value that could be accrued from it. This value will be lower if an asset has deteriorated, has a shorter remaining life, reduced functionality or inefficiency.

KEY MESSAGE: Cost-based methods have been traditionally used in regulated sectors where the presence of a state-owned monopoly prevents the existence of a market for regulated assets. A similar rationale could justify the use of cost-based methods for data assets that cannot be traded (e.g. those covered by restrictions in terms of national security), although there are limitations in the case of data, described in the rest of this section.

Focus on/3: the Sainsbury's/Nectar acquisition

In 2018, the supermarket chain Sainsbury's purchased the Nectar loyalty card business for £60 million, boosting the supermarket group's control of customer data. Sainsbury's, which has been part of the Nectar scheme since it *launched in 2002, acquired all the assets, staff, systems and licences required for the independent operation of the Nectar loyalty programme in Britain from Aimia, a data, marketing and analytics company (Vandevelde, 2018).*

As the main assets of Nectar are related to data, the value of this transaction can be considered a good proxy of the value of the data assets owned by Nectar. However, the price paid for this acquisition may also reflect other factors, such as the potential synergies between the companies, Sainsbury's valuation of the broader technology used by Nectar or of the organisation's capabilities to innovate.

KEY MESSAGE: Company valuations from mergers and acquisitions can provide an upper bound for the value of the data assets owned by the target company. The valuation is less likely to overestimate the value of data if the target business model is strictly or solely related to its data assets and if there are limited synergies beyond those related to data between the acquirer and the target.

Focus on/4: the importance of stakeholders in valuing data assets

During the proof-of-concept (PoC) exercise, we interviewed an organisation which supports businesses and government to understand the value of their data and to prioritise investments across the data assets they own and use.

The specific methodology used to achieve these objectives is, of course, very commercially sensitive. However, there is one high-level principle that is worth highlighting as it is particularly insightful from a policy-making perspective.

One of the main challenges in valuing data assets is understanding how they are used in different organisations and what weight to give to different use cases.

To address this challenge and understand how the data assets under analysis generate value in the organisation, the starting point of the valuation methodology discussed in the PoC exercise is to involve all the individuals and the organisations that are affected by the data asset, whether directly or indirectly, in positive or negative ways, intentionally or unintentionally.

The inputs from these stakeholders are then used to conceptualise, in a comprehensive and structured way, the ways in which the asset is used in the organisation and the mechanisms through which they generate relevant outcomes. Once the value chain of the data asset has been broken down in this very granular and structured way, a weight and a monetary value are attached to each outcome and to each mechanism in order to estimate the final value of the asset. This exercise is repeated regularly over time to capture changes in the value chain and in stakeholders' perspectives.

This stakeholder involvement process appears to be similar to the first step of the social return on investment (SROI) methodology (SROI Network, 2012) and seems to be affected by the same advantages (through and focused on relevant outcomes) as well as by the same limitations (time and resource-consuming and case-specific).

KEY MESSAGE: Because of the vast variety of use cases and contexts in which the same data asset can be used, including all stakeholders affected by the data asset in the valuation process is a starting point to understand which elements are more relevant for extracting value from the asset.

Focus on/5: amortisation of data assets (PGS, 2016; CGG, 2016)

Petroleum Geo-Services (PGS)

PGS is a firm which provides seismic images and 3D data that describe the subsurface beneath the ocean floor. Its business is data-driven and it uses the latest technologies to provide mapping libraries, seismic acquisition, *processing and imaging services to support its clients in the exploration and production of oil and gas reserves worldwide.*

Due to the relevance of data assets in its business model, PGS regularly measures the value of its data assets and they are a material component of its accounting activities. For example, in January 2016, after an extensive process involving industry participants, PGS decided to introduce an amortisation policy, based on two steps:

First, during the work-in-progress (WIP) phase (i.e. when the data asset is being generated), amortisation is based on total costs versus the forecast total revenues of the project.

Then, after a project is completed, a straight-line amortisation is applied. The straight-line amortisation is assigned over a remaining useful life, which for most projects is expected to be four years. The straight-line amortisation is distributed evenly through the financial year independently of sales during the quarters.

Compagnie Générale de Géophysique (CGG)

One of PGS's competitors, CGG, also recently introduced a new amortisation policy. More specifically, based on how many units have been sold to date, each survey is amortised over a five-year period in a manner that reflects the pattern of consumption of its economic benefits during both prefunding and after-sales periods.

For certain large sales, the amortisation rate is adjusted to reflect the commercial effects of price elements: for example, if a special rebate is granted to a customer buying a large volume of data, it could then trigger a higher amortisation rate.

KEY MESSAGE: As with more conventional asset classes, the approach taken to amortise data assets over time is an important element when assessing its value. This is important from both an accounting and an economic perspective. The period over which economic value can be extracted from a data asset will depend on the characteristics of the asset, on existing and future use cases and on the organisations which will use the data.

Focus on/6: accounting for data-handling requirements when storing data on the cloud (Henze et al., 2020)

Accounting for compliance with data-handling requirements and offering control over where and how data is stored in the cloud is becoming increasingly important due to legislative and organisational demands. Despite these incentives, practical solutions for addressing this in existing cloud storage systems are scarce. In a recent publication, Henze et al., 2020 propose a new framework called PRADA: a practical approach to account for compliance with data-handling requirements in key value-based cloud storage systems.

More specifically, PRADA introduces a transparent data-handling layer which empowers clients to request specific data-handling and data-location requirements and enables operators of cloud storage systems to comply with them in an automated way.

In their paper, Henze et al. (2020) implement PRADA on top of an existing database and show that complying with data-handling requirements in cloud storage systems using the PRADA approach is practical in real-world cloud deployments used for microblogging, data sharing in the Internet of Things and distributed email storage. In all these contexts, this innovative system allows clients to specify a comprehensive set of fine-grained data-handling requirements and data-location constraints and enables cloud storage operators to enforce them automatically.

According to the authors, by offering the enforcement of arbitrary data-handling requirements when storing data in cloud storage systems, PRADA enables the use of cloud storage systems for a wide range of clients who previously had to refrain from outsourcing storage, for example due to compliance with applicable data protection legislation. As a result, it can increase the value of the data asset stored on the cloud as it reduces the risk of non-compliance with regulations associated with data location.

KEY MESSAGE: Data-location requirements are often associated with higher costs in terms of the collection, management and storage of data assets, which can often have a negative impact on the value of the data

asset. In this context, investments in research aimed at developing innovative ways to minimise these costs could be a policy option to consider in order to maximise the value of data.

Focus on/7: changing preferences in terms of timeliness and accuracy in economic indicators (Haldane & Chowla, 2020; McKinsey Global Institute, 2016)

Businesses and investors have historically attached significant value to the accuracy of national account statistics, particularly gross domestic product (GDP) estimates, and have placed less weight on the timeliness and speed of the first publications of these indicators (McKinsey Global Institute, 2016). For example, after a public consultation published in July 2017, the Office for National Statistics (ONS) introduced a new publishing model for GDP in July 2018 (ONS, 2018), structured around two (rather than three) publication dates for quarterly GDP. This decision was justified by the desire to improve data accuracy, accepting the associated losses in terms of timeliness. The ONS press release explicitly stated: "although there will be some loss in timeliness and data content, the improved quality of our quarterly estimates will make the data more reliable and hence lead to greater confidence in these estimates".

Conversely, during the COVID-19 pandemic, in light of the unprecedented nature of the shock it generated in the economy and of the significant volatility it introduced in financial market, investors started to demand more timely indicators of economic activity, accepting the associated costs in terms of accuracy and precision.

As a result of this demand, economists and policy-makers have developed a raft of new "fast indicators" when tracking the economic response to the COVID-19 crisis. These are fast in the sense that they are often available on a daily basis with little, if any, lag. As such, they offer a close to real-time read on how the economy is performing but are inevitably characterised by a lower level of accuracy and precision (Haldane & Chowla, 2020).

KEY MESSAGE: The trade-offs between characteristics also depend on the use case of the data asset under analysis and can change in response to evolving contexts and use cases. It is important to take all these aspects into account when designing policy interventions aimed at minimising this trade-offs.

Focus on/8: unforeseen uses of data assets (Norwegian Petroleum Directorate, 2017)

The Norwegian Petroleum Directorate (NPD) has recently used both new and older datasets to map eastern parts of the North Barents Sea for oil exploration purposes. In fact, it combined 70,000 kilometres of 2D seismic lines in the North Barents Sea, purchased between 1973 and 1996, with 32,600 kilometres of new 2D seismic data in the area, purchased between 2012 and 2014.

The quality of many older datasets was comparatively poor and they were not expected to generate any value after their first use for oil exploration purposes in the few years that followed their purchase. However, older datasets have proved particularly helpful in recent years when combined with newer datasets, for two reasons.

First, while the quality of the new datasets is considerably better than older ones, they have been shown to have similar limitations in relation to the deeper parts of the sedimentary basins. In these contexts, older datasets are equally informative and can save a significant amount of money, which can be redirected to areas where newer technologies can add value to the process.

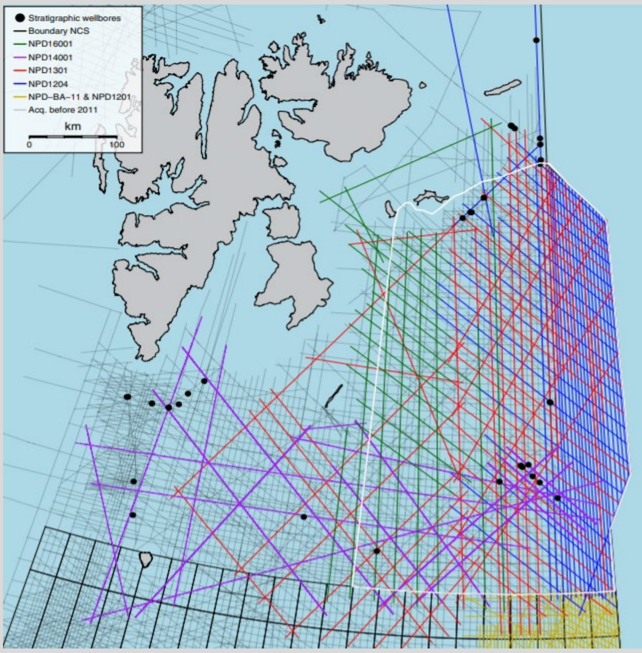
Second, where other economic activities have emerged in recent years (e.g. wind farms, fishing farms, etc.), new surveys with newer and more accurate technologies are more complicated or simply difficult as these activities impede the deployment of new measurement tools on the seabed.

In this context, older seabed data that was expected to have little or no value, because of the higher accuracy of more recent and accurate technologies, has gained value due to circumstances and developments which were unforeseeable when the original data was purchased in the 1970s. Figure 18, taken from a recently published report, shows the relative relevance of the older datasets (mapped in grey) alongside those purchased in the 2010s (mapped with colours).

KEY MESSAGE: The value of a data asset is mainly driven by the existing and potential use cases associated with it. However, in the fast-paced changing data economy, it is not conceptually or empirically feasible to account for all the potential uses cases that could be associated with a data asset in the future. This is particularly true in light of the rapid development of artificial intelligence technologies which could

automatically identify new use cases for old data assets. When designing and implementing data policy interventions, it is important to recognise and properly account for this element of uncertainty

Figure 18: Data coverage in the North Barents Sea, highlighting the relevance of older seabed data (Norwegian Petroleum Directorate, 2017



Focus on/9: Why data culture matters (McKinsey, 2018)

In 2018, McKinsey conducted semi-structured interviews with business leaders and executives on what they had done to progress the data culture in their organisation and to strengthen their analytics enterprise.

These interviews highlighted eight main factors and principles which underpin a healthy organisational culture around data, focused on different roles in the organisation:

- Business leaders lead analytics transformation across organisation;
- Delivery managers deliver data and analytics-driven insights and interface with end users;
- Workflow integrators build interactive decision-support tools and implement solutions;
- Visualisation analysts visualise data and build reports and dashboards;
- Data engineers collect, structure and analyse data;
- Data architects ensure the quality and consistency of present and future data flows;
- Analytics translators ensure that analytics solve critical business problems; and
- Data scientists develop statistical models and algorithms.

While these principles may appear overly generic and abstract, they offer a comprehensive starting point for organisations to assess their strengths and weaknesses in relation to their organisational culture around data and to intervene in the areas for improvement which offer the most promising potential.

KEY MESSAGE: Acknowledging that, in most cases, an accurate quantitative measurement will be impossible, even the most abstract and intangible aspects affecting the value of data assets can be conceptualised and assessed qualitatively. Depending on the industry and the context, investments in research aimed at investigating the relationship between organisational culture and the value of data assets could be a policy option to consider in order to maximise the value of data.

Focus on/10: the interaction between skills and culture (NIC Data, 2020)

AkzoNobel is a global expert in paint and coatings for decorating homes and businesses, protecting pipelines and turbines and coating aircraft, automotive vehicles and marine vessels.

In its marine business, the company sells anti-corrosive coatings to operators of fleets of marine vessels travelling the world. These coatings help to protect hulls from everything from saltwater damage to barnacles and can last for as long as 25 to 30 years. However, to assess how vessels are performing, ships still need to check back into dry dock every five years or so. Vessel operators want to keep the time in dry dock brief so that they can save time, money and fuel.

AkzoNobel therefore needed a solution to allow customers to predict how their hulls were holding up – one which did not involve having to wait for a docking or dive inspection.

Over decades of activity, AkzoNobel had accumulated extensive insight into the factors that caused corrosion and had access to a range of datasets which it could draw on, including the positioning of 80,000 vessels, taken every 15 minutes for a nine-year period.

The organisation initially worked with consultants on projects which pooled all this data into "predictive" models which would help the company to make smarter decisions in relation to its product offering and to the maintenance services it provided to its clients.

However, that arrangement meant that the company was reliant on external skills. It therefore decided to learn some of the skills that would enable it to do more of this work in house and to engage with NIC Data: an organisation funded by the UK government and Newcastle University with the goals of addressing the shortage of data skills in the UK, transferring practical data skills into the workforce of private and public sector organisation and empowering organisations to gain insights from their data.

The objective of this partnership was to train a core team to be able to replicate in house the same model produced by the external consultants. However, as the project progressed, more and more staff started to become involved and to sit with the core team to absorb some of the language and knowledge associated with this predictive model.

Furthermore, the team learned more about the data that was available and how they could use it. They looked at the formatting of data and how they could share it securely and seamlessly, and ended up almost completely re-building the model from the ground up, creating a product that was even better than the model they started from.

KEY MESSAGE: The availability of the right skills and competencies needed to extract value from data assets is an important aspect to consider when designing policies in this space. These skills can be analysed as part of the organisational culture or as a standalone item, depending on the policy intervention under analysis.

Focus on/11: a comprehensive approach to interoperability assessments (Leal et al., 2019)

Interoperability allows enterprises to exchange information and use it to achieve their shared goals. Interoperability Assessment (INAS) has the objective of determining the strengths and weakness of an enterprise in terms of interoperability. The literature has proposed many surveys and reviews for analysing existing INAS approaches. However, most of these reviews focus on specific properties rather than a general view of an INAS.

Leal et al. (2019) developed a systematic literature review of 38 INAS approaches to compare their properties (e.g. type of assessment, the measurement mechanism used, and the interoperability barriers addressed). Twenty-two of these approaches were presented in the context of real case studies or illustrative examples, making it possible to contextualise these theoretical frameworks.

The paper concludes that there is a gap in the literature with regard to approaches which provide best practices for improving interoperability based on the INAS results. Such guidelines are essential as they can help stakeholders (e.g. system engineers and enterprise architects) to design and implement interoperable systems and, according to the authors, should be the focus of further research in this field.

To fill this gap in the literature, the authors propose a new comprehensive assessment framework to improve the existing INAS approaches, using a computer-mediated tool for facilitating the overall INAS process.

More specifically, this tool automatically calculates the interoperability level of an enterprise or between two enterprises, it identifies potential interoperability barriers and impacts on different layers and concerns of *interoperability, and, most importantly, it provides best practices for reducing the detrimental effects of the identified barriers.*

KEY MESSAGE: Some data characteristics, like interoperability, are particularly complex to measure and assess from both a conceptual and technical perspective. Depending on the industry and the context, investments in research aimed at unpacking these complexities and at providing organisations with tools to measure these characteristics are a policy option to consider in order to maximise the value of data.

Focus on/12: unintended costs of open data (Drieger 2015; Debussche et al., 2019)

There can be unintended consequences and costs associated with open data regulations. For example, the first versions of the EU Directive 2003/98/EC on the re-use of public sector information forced public undertakings to make high-value datasets available for free. As a result, the French rail operator SNCF stated that obliging public transport operators to share their data with private competitors for free would distort competition and would put the public undertaking in a position of competitive disadvantage compared to the private sector competitors in the rail and road transport markets (European Parliament, 2019).

These unintended consequences were believed to hinder ongoing innovation in public service undertakings by increasing the risk of investing in their own datasets and collaborating with start-ups, thus taking away the incentive for public undertakings to carry out such activities. As a result, the directive was amended to expressly exclude the requirement to make such high-value datasets available for free in case this would lead to a distortion of competition in the relevant market (Debussche et al., 2019).

KEY MESSAGE: Data sharing and openness generate a variety of benefits in terms of economies of scale, scope and network effects, which often translate into increases in productivity and innovation and, as a result, into higher value of the data assets involved. However, depending on the context and the industry, these characteristics can also be associated with risks and unintended consequences which could distort competition and hinder innovation in certain markets.

Focus on/13: barriers to diffusion of data analytics and cloud computing and storage in SMEs (Bianchini et al., 2019; Eurostat, 2015)

A recent study by Bianchini et al. (2019), on behalf of the OECD provides useful insights about the main forces at play in the diffusion of advanced techniques such as data analytics, whose uptake outside of the ICT sector is still limited among SMEs, which are held back by a number of internal and external barriers.

Internal barriers include lack of knowledge and awareness, mistrust in digital solutions, inability to address digital security challenges and lack of skilled human capital. External barriers include limited access to finance and digital networks, limitations in the availability of data and regulatory constraints.

This list is based on a series of surveys and analysis conducted amongst SMEs in various OECD countries. For example, according to a 2014 survey, among 1,000 German SMEs, 70% of enterprises with annual revenue below €500 million do not consider the digitalisation of processes to be relevant (OECD, 2017).

Similarly, according to a 2015 survey of the manufacturing sector in Japan, the main obstacles to data use are related to lack of human resources and planning (Motohashi, 2017).

Furthermore, in 2015, only 30% of SMEs in Europe had a formal security policy, against almost 70% among large enterprises, with the share ranging from almost 50% in Sweden and Portugal to close to 10% in Poland and Hungary. Only 14% of European SMEs handle digital security and data protection using internal staff, against 64% of large enterprises (Eurostat, 2019).

These are some selected statistics which highlight the different barriers which small and large businesses face in extracting value from data assets that need to be reflected in the valuation of the asset(s) under analysis, as well as in data policies that are expected to have a differential impact on firms of different size.

KEY MESSAGE: The ways in which small and large firms use data assets are materially different and should be accounted for when designing data policy interventions. Most importantly, the obstacles and the barriers they face when collecting and using data assets are different. This is particularly important when designing policies aimed at stimulating the adoption of data analytics in the economy.

Focus on/14: the trade-off between timeliness and accuracy for low-voltage distribution system grid monitoring (Kemal et al., 2020)

The generation of power required to feed the loads of the entire electrical grid has traditionally been centralised in large power plants placed at the high-voltage level, and the low-voltage grids have merely been seen as loads. However, this operational philosophy is changing due to the appearance of decentralised or distributed generation based on renewable energy sources such as photo voltaic panels or small-scale wind turbines.

In this context, data collected by smart meters on the amount of energy produced by various low-voltage grids in the system is used to manage the performance of the grid and to ensure its operativity at any point in time. Due to the limited bandwidth and considerable delays in accessing smart meter measurements, it is typically not possible to access measurements from the complete set of smart meters in a low-voltage grid area for distribution grid monitoring.

As a result, distribution system state estimation can be performed based on measurements of a subset of selected smart meters. In this environment, a clear trade-off emerges: increasing the number of selected smart meters will, on the one hand, increase the accuracy of distribution system state estimation, while, on the other hand, it will degrade timeliness of the monitoring data.

In a recent paper, Kemal et al. (2020) developed an innovative methodology to quantitatively analyse this trade-off. This methodology was used to come to an operational recommendation for electrical grid managers to enable them to maximise the value they can extract from the data assets generated by the smart meters in the grid. More specifically, every smart meter has an "idle time" during which it does not input any data into the system. If properly calibrated, this time window can be used to introduce into the system other inputs from other smart meters that are not in their "idle phase". This increases the quantity of smart meters included in the distribution system state estimation process, with a positive impact on data accuracy and minimal costs in terms of timeliness.

KEY MESSAGE: Trade-offs between different characteristics often exist but can be minimised. Depending on the industry and the context, investments in research aimed at developing innovative ways to minimise these trade-offs could be a policy option to consider in order to maximise the value of data.

ANNEX C GLOSSARY AND KEY TERMS

- **Asset:** economic resource which is expected to provide a benefit over a period of time (EC et al., 2008, 10.32)
- **Data:** characteristics or information, usually numerical, which are collected through observation. Data is the physical representation of information in a manner suitable for communication, interpretation or processing by human beings or by automatic means (OECD, 2006).
- **Data asset:** in this report, we are particularly interested in data as an asset: that is, cases where data can provide an economic benefit over a period of time, as opposed to cases where data only provides an immediate benefit (data as a good). However, it is challenging to distinguish a priori where data is best characterised as an asset rather than a good. Therefore, we generally use "data" and "data asset" interchangeably throughout this report.
- **Information asset:** a body of information, defined and managed as a single unit so that it can be understood, shared, protected and exploited efficiently. Information assets have recognisable and manageable value, risk, content and lifecycles (National Archives, 2017).
- **Innovation**: the implementation of a new or significantly improved product (good or service) or process, a new marketing method or a new organisational method in business practices, workplace organisation or external relations (OECD, 2005)
- **Investment:** what happens when a producer either acquires a fixed asset or spends resources (money, effort, raw materials) to improve it (EC et al., 2008, 617).
- **Market failure:** in microeconomic theory, a market failure is a situation in which the allocation of goods and services by a free market is not efficient, often leading to a net loss of economic value and the need for policy intervention (Vernengo et al., 2008). Annex D presents three "traditional" types of market failures applied to the data economy.

ANNEX D MARKET FAILURES IN THE DATA ECONOMY

In microeconomic theory, a market failure is a situation in which the allocation of goods and services by a free market is not efficient, often leading to a net loss of economic value (Vernengo et al., 2008). The concept of market failures is typically used to identify situations where government intervention is required to achieve more desirable outcomes than those delivered by an unregulated market.

We present three "traditional" types of market failures in the following paragraphs. These have been analysed for many decades by the theoretical and empirical economics literature and have informed the design and the implementation of many policy interventions in different sectors. However, their application to the data economy has not been analysed as extensively due to the recent and fast-paced nature of this sector.

These paragraphs do not aim to provide a complete list of areas where policy intervention is required in the data economy. They only represent some examples, with the objective of:

- Contextualising the abstract concept of market failures to the real-world context of data assets; and
- Highlighting the importance of a thorough understanding of data characteristics to understand the market failures that might occur in the data economy.

Economies of scale and scope

In traditional markets, economies of scale are one of the most relevant economic principles in analysing fixed assets. They represent a situation where the average cost of production, and hence the unit cost of a good or a service, decreases when output is increased. Economies of scale typically occur when better use is made of the factors of production and by using the increased output to pay for a higher proportion of the costs of marketing, financing and development (Law, 2009) – for example, when the cost of a steel furnace can be spread across more units sold by a steel plant.

Similarly, economies of scope are the benefits which arise from engaging in related activities. While with economies of scale, cost savings arise from carrying out more of the same activity, with economies of scope, cost savings arise from producing and selling separate but similar goods or services. They typically occur when specialised labour, equipment and ideas used in one activity can be used in related activities (Hashimzade et al., 2017) – for example, when the patent for a chemical compound can be used to produce various drugs used for different clinical purposes.

In the context of data, both concepts are particularly relevant. Indeed, the new technologies of information show very strong "returns to scale" as the cost of producing data assets grows at a much slower pace than the number of customers who can be served with that asset. Once created, information can be transmitted to many people at very low cost. For example, once a search engine or mapping service has been developed to generate, store and analyse data, it can usually serve hundreds of thousands of users fairly cheaply (Crémer et al., 2019).

Similarly, data is an excludable, non-rival good, in the sense that individuals or organisations can be excluded from accessing it but, once they are included, the use of data by one person or organisation does not reduce the amount available for others (Nolin, 2019). As a result, the costs to an organisation of collecting, managing and analysing data can be reduced by operating simultaneously across multiple markets (Duch-Brown et al., 2017). These strong economies of scope are one of the reasons why the same small number of large digital companies have successfully built ecosystems across several adjacent markets (Digital Competition Expert Panel, 2019).

Negative externalities

Negative externalities occur when the effect of production or consumption of goods and services generates costs on other individuals and organisations which are not reflected in the prices charged for the goods and services sold on the market (OECD, 2003). Pollution is the typical example of a negative externality: chemical waste disposed of by an industrial plant into a river may affect the quantity and the quality of fish and plant life and, in turn, generate a cost for fishermen and farmers nearby which is not reflected in the prices the chemical firm charges for its products.

In the data economy, negative externalities typically occur in terms of privacy losses and, at their extreme, data breaches. For example, in certain circumstances, some business models in the data economy give rise to an excessive collection of personal information, which leads to a loss of privacy compared to the social optimum (Choi et al., 2019). In some cases, it is the social dimension of the individual data, whereby an individual’s data is predictive of the behaviour of others, which generates a negative data externality (Acemoglu et al., 2019; Bergemann et al., 2020).

Information asymmetry

Information asymmetry refers to a situation where sellers have more information than buyers, or vice versa, about some aspect of product quality. This situation is problematic when asymmetric information is exploited by one party.

In the context of data, information asymmetry is particularly relevant because, compared to other types of asset, data assets are characterised by an intrinsic lack of visibility in terms of data uses and data quality. For example, there is a high level of transparency in the market for corporate bonds, as the quality of these assets is certified by a variety of institutions and organisations and it is clear to all market operators how and when these assets can be used (e.g. sell them on the secondary market, use them as collateral for other financial operations, etc.).

Conversely, buyers of data products (e.g. advertisers on online platforms) may not have full visibility of the quality of the data used to target ads (Kim et al., 2018). Similarly, online advertising platforms may not have visibility of how the data they shared with their customers is used (Isaak & Hanna, 2018).

ANNEX E FURTHER DETAIL ON MEASUREMENT OF DATA CHARACTERISTICS

Figure 19: list of data characteristics identified in the literature

Characteristic	Description	Source
Support	Whether the data is in digital, analogue or mixed format.	DCMS (2020)
Source	From where it has been collected, purchased, obtained or transformed.	USGS (n.d.)
Size	How big the dataset is in terms of storage volume (terabytes (TBs)) or number of observations.	KPMG (2019)
Rationale for collection	Why the data has been collected (e.g. legal requirement, service provision, deriving insights).	DCMS (2020)
Data content/ subject matter	What the data asset refers to (e.g. geospatial data, business data, personal data).	DCMS (2020), Coyle et al. (2020)
Variety	Suitability for cross-sectional/panel analysis and to control for more factors. Key driver of the descriptive, analytical and predictive power of the data.	Hogan et al. (2016)
Findability	As defined in FAIR principles – for example, are (meta)data assigned a globally unique and persistent identifier?	Gofer (n.d.)
Functional form of returns/scalability	The extent to which an increase by 10% in the amount of data generates an increase in returns above, below or at 10%.	Coyle et al. (2020); Haskel and Westlake (2018)
Timeliness	Whether the data asset is real-time, delayed or historic.	Coyle et al. (2020)
Completeness	Proportion of missing values.	PwC (n.d.)
Validity	Proportion of invalid data points (e.g. temperature recorded as "abc").	GSMA (2018)
Consistency	Proportion of data points recorded in the same way.	PwC (n.d.)
Arithmetic precision	How precise are the data points in the asset. For example, number of decimals or detail on geospatial data.	Ginsburg & Phillips (2018)
Accuracy	Proportion of correct data points (e.g. thermometer being 2°C above or below real value).	Deloitte (2020), PwC (n.d.), Coyle et al. (2020)
Targetability	Extent to which a specific group can be singled out.	Deloitte (2020)
Generality	Size of the group/population to which the data refers (e.g. geospatial data is applicable to the entire population living/active in the area covered).	Hogan et al. (2016)

Representativeness	Regarding the population under analysis: randomised sample, semi-randomised, self-selected.	Coyle et al. (2020)
Interoperability	Technical standards that allow use of data across systems/platforms.	Coyle et al. (2020), GoFair (n.d.)
Linkability	Number and type of data assets with which it can be linked.	Coyle et al. (2020), PwC (n.d.), Haskel and Westlake (2018)
Collection method	The process through which the data has been collected (e.g. scraped online, phone interviews, observed in nature).	DCMS (2020)
Accessibility/exclude-ability	How data can be accessed: internal, named, group-based, public, open.	Coyle et al. (2020)
Use restrictions	Type, duration and source of restriction to the use of the asset.	PwC (n.d.)
Ownership	Who owns the data from a legal perspective?	Van Asbroeck et al. (2019).
Location (of storage and use)	Where is the data stored and to what jurisdiction it is subject?	Deloitte (2020)
Liabilities and risks	Type, duration and source of liability and risks generated by the asset.	PwC (n.d.)
Uniqueness and exclusiveness/ scarcity	Number and type of similar data assets available.	Deloitte (2020), Nguyen & Paczos (2020)
Functional form of costs/ sunkenness	Relationship and nature of fixed costs, semi-fixed costs and variable costs associated with the asset.	Coyle et al. (2020), Haskel and Westlake (2018)
Complementary assets	Investment in other intangibles, such as copyright, patents, market research required to make use of and extract value from the asset.	Haskel and Westlake (2018)
User of data (by industry)	In which industry the data is used. This characteristic could be merged with ownership but is presented separately here to reflect how it is discussed in the literature.	Nguyen & Paczos (2020)
User of data (by function)	By which function the data is used. This characteristic could be merged with ownership but is presented separately here to reflect how it is discussed in the literature.	Coyle et al. (2020)
Stage in the data value chain	The phase where the data asset is in its lifecycle. The OECD proposes five stages: collection, aggregation, analysis, use, monetisation. Coyle et al. (2020) propose seven stages: raw data, processed data, integrated data, analysis, actionable insights, action, (potential) value.	Nguyen & Paczos (2020), Coyle et al. (2020)
User of data (by business model)	How is the data used in the organisation(s)? This characteristic could be merged with ownership but is presented separately here to reflect how it is discussed in the literature.	Nguyen & Paczos (2020), PwC (n.d.), Deloitte (2020)

Figure 20: considerations on measurement and impact on value of characteristics identified in the literature

Characteristic	Considerations on measurement	Impact on value
Size	Can be easily measured quantitatively in gigabytes (GB), TB, etc,)?	Larger data includes more information. Marginal benefits from additional information, all else equal, are expected to be decreasing (e.g. the value of the first GB is higher than the value of the 100th GB) or characterised by a step-change (e.g. once the dataset reaches 100GB or 1,000 observations, it becomes usable for new and valuable use cases and any increase above that threshold does not have much additional value).
Findability	Can be measured using process-based definitions according to FAIR principles (e.g. assessing whether “data are assigned a unique and persistence identifier”).[34] Alternatively, outcome-based measures could be considered, which, for example, ask users how long it took them to find the data.	Increases in findability, accessibility and reusability generate value by allowing additional uses of the data. Assessing their impact therefore requires assessing the extent to which data use does increase, relative to a counterfactual when findability, accessibility and reusability increase. Changes in these characteristics may also affect how value generated by the data is distributed. For company A using a data asset, an increase in accessibility which leads its competitors to start using the data may make this data less valuable as a source of competitive advantage.
Accessibility	Can be measured using process-based definitions according to FAIR principles (e.g. assessing whether “data are retrievable by their identifier using a standardised communications protocol”). Alternatively, outcome-based measures could be considered, which, for example, ask users how easy or hard it was to access the data.	
Reusability	Can be measured using process-based definitions according to FAIR principles (e.g. assessing whether “ data are richly described with a plurality of accurate and relevant attributes”). Alternatively, outcome-based measures could be considered.	
Interoperability, linkability	Can be measured using process-based definitions according to FAIR principles (e.g. “data use a formal, accessible, shared, and broadly applicable language for knowledge representation”). Interoperability may be somewhat more challenging to assess in practice compared to findability and accessibility because a given data asset A may be interoperable relative to some data assets (B, C, D) but not others (E, F).[35]	As above for findability, accessibility and reusability. Moreover, interoperability and linkability allow different data to be integrated, which can lead to changes in other data characteristics (e.g. size, completeness, generality).
Timeliness	Timeliness can be measured using a categorical variable: real time, delayed and historic.	More timely data will generally be more valuable because it allows analysis and decisions based on the data to occur faster.

[34] GoFair (n.d.).

[35] Focus box 11 in Annex B provides a practical example.

		However, some analyses will be less time critical than others.
Completeness	Changes to validity and completeness can be measured in a relatively easy way using, respectively, the proportion of invalid values (e.g. a date in a field supposed to record a temperature) and the proportion of missing values (zero, NAs or empty entries). A qualitative assessment could be added to this measurement to give different weights to different missing and invalid values if necessary or helpful.	Greater completeness, validity, consistency, precision, accuracy, generality and representativeness all lead to better quality of information included in the data. These characteristics can interact. For example, if an organisation is using a database of 1 million customers with their addresses to deliver products, 5-10% missing addresses could be a huge issue. If the same asset is used for statistical analysis, then 5-10% missing values (if randomly distributed) is not particularly problematic on a dataset with 1 million observations.
Validity		
Consistency	Consistency, linkability and interoperability can be assessed from a qualitative perspective. However, this assessment requires more technical skills in relation to the technical standards and formats which characterise the data assets under analysis.	
Arithmetic precision	Relatively easy to measure using the units used to collect the data (e.g. mm, cm, km).	As above.
Statistical precision	Extent of variation between different measurements of the same data point.	
Accuracy	Accuracy is one of the hardest characteristics to measure. While conceptually easy (e.g. what proportion of data points are "correct"/ "true"), the key challenge is to know whether a data point is correct. The easiest way may be to ask individuals/organisations who use the data how confident they are that the information is correct. Alternatively, more specific questions on internal processes could be asked: for example, whether the data has been/is frequently validated against alternative data sources.	
Targetability	Targetability can be assessed using a categorical variable indicating the extent to which specific groups or individuals can be identified in the dataset.	
Variety	Changes in other quality-related characteristics such as representativeness, generality, accuracy/trustworthiness and variety, are more challenging to quantify and to assess qualitatively. A comparison with other data assets (e.g. national statistics) could help in this assessment and serve as a benchmark against which all the data assets under analysis are compared. However, in many instances, only a general high-level	
Accuracy		
Generality		
Representativeness		

	qualitative assessment may be feasible in relation to these characteristics	
Support	The support of the data asset can be measured using a categorical variable: analogue, digital or mixed.	The marginal costs of digitising a data asset can typically be expected to be decreasing (i.e. digitising the first 10% of a data asset will cost more than digitising the last 10%).
Source	Data source typically refers to the organisation responsible for curating it (e.g. ONS, OECD).	Completely depends on the context, use case and organisational characteristics.
Rationale for collection	The wide variety of rationales makes the measurement of this characteristic very data asset-specific. A potential way to address this challenge is to link rationales for collection to use cases, as indicated in section 3.5 above.	Completely depends on the context, use case and organisational characteristics.
Data content/ subject matter	Same as above.	
Structure	This can be measured using a categorical variable: structured or unstructured.	Structured data can be expected, keeping everything else constant, to be more valuable than unstructured assets.
Collection method	Challenging when the data is manipulated multiple times and is transferred across different organisations.	Changes in collection method are related to sensitivity because of the regulatory framework built around the notion of consent. For example, a questionnaire where the respondent has signed a section at the end of the survey authorising its use for certain purposes may have a higher value compared to a series of information scraped from the internet without the consent of the subject with which the data is associated. In this context, it is important to assess whether the legal and reputational risks associated with the use of the second type of data outweigh the costs generated by collecting and managing the data asset in a "compliant" way.
Accessibility/ excludability		As focus box 12 on open data shows, data sharing and openness can generate a variety of benefits in terms of economies of scale, scope and network effects, which often translate into increases in productivity and innovation and, as a result, into higher value of the data assets involved. However, depending on the context and the industry, these characteristics can also be associated with risks and unintended consequences, which could distort competition and hinder innovation in certain markets.
Use restrictions		Similarly, changes to use restrictions, ownership, and location are typically subject to regulation, generating direct and indirect compliance costs as well as potential operational, legal and reputational risks associated with non-compliance.
Ownership	A distinction between legal ownership (e.g. the individual resident owns their own address) and economic ownership (the e-commerce platform owns my purchase records, which also include the address to which the purchases were shipped) may be relevant in some contexts.	

Location	Changes in this characteristic are relatively easy to measure, but material challenges may arise in situations where cloud storage can be allocated to different servers located in different jurisdictions by algorithms that do not take account of location.	
Liabilities and risks	Assessing and measuring liabilities and risks requires more complicated assessments for which specialist legal and operational expertise is needed. Reputational risks are also particularly challenging to quantify. Looking at previous cases of data breaches or non-compliance could be informative and provide a benchmark for measurement.	On the one hand, regulatory compliance (e.g. GDPR) generates costs which can decrease the value of the data asset from the perspective of the organisation. Furthermore, the operational, legal and reputational risk associated with non-compliance can generate a variety of costs that need to be reflected in the value of the data asset under analysis. On the other hand, it can increase the scarcity of certain data assets (e.g. assets containing personal data), making them more valuable in the market or for the organisations which own them. These two examples highlight that, depending on the context, changes affecting the sensitivity/confidentiality of a data asset can affect both demand- and supply-side considerations and have an impact on the costs and the benefits generated by a data asset.
Uniqueness and exclusiveness/ scarcity	This can be measured qualitatively, by listing the comparable/similar data assets available to one organisation or in the market. A quantitative element can be added to identify the proportion of overlapping information between similar (but not identical) assets.	Completely depends on the context, use case and organisational characteristics.
Complementary assets	This can be measured qualitatively by listing the different types of complementary assets needed to extract value from an asset. A quantitative element can be added to identify the costs generated by these.	Keeping everything else constant, the need for a complementary asset will decrease the value of the asset (as organisations need to incur additional costs to use the asset). However, once an organisation has already incurred these costs, it can make the asset even more valuable to that specific organisation.

ANNEX F DETAILS ON LITERATURE REVIEW AND PROOF OF CONCEPT

Literature review

We undertook a non-systematic review of evidence from academic literature and “grey” literature (including think-tanks, international organisations, consultancies, government publications) on the characteristics of data and data valuation. Because of the timeframe of this project, the fact that a number of reviews of evidence were undertaken (see section 5) and because the evidence base is likely to include tens rather than hundreds of directly relevant material, this review falls between the “literature review” and “quick scoping review” categories defined in Table 2 of Collins et al. (2015). This involved targeted searches for relevant terms on Google and Google Scholar (e.g. “data asset valuation”; “value of data”; “valuation of intangible assets”), using existing reviews (e.g. Coyle et al., 2020) and references included in those, and identifying material citing publications identified by the DCMS team and through Frontier’s previous work on this topic.

Stakeholder engagement

This report reflects input from the following experts and stakeholders:

Academics and sector experts:

- Jeni Tennison from the Open Data Institute (ODI),
- Richard Heys from the Office for National Statistics (ONS),
- David Nguyen from the National Institute of Economic and Social Research (NIESR),
- Daniel Ker from the Organisation for Economic Co-operation and Development (OECD),
- Prof Diane Coyle from the Bennett Institute for Public Policy at the University of Cambridge,
- Martin Duckworth, Mike Huggins and Rob Francis from Frontier Economics.

Businesses and stakeholders:

- Common Data Access Limited (CDAL),
- The Association of the British Pharmaceutical Industry (ABPI),
- UK Cloud,
- Mydex CIC,
- Dunn & Bradstreet,
- Burning Glass,
- Anmut,
- National Innovation Centre for Data,
- AzkoNobel,
- Belmana,
- Ernst and Young.

ANNEX G CONCRETE APPLICATION OF DIFFERENT VALUATION METHODS TO DATA

Cost-based methods

Existing studies which use cost-based methods to assess the value of data typically rely on information on labour costs.

To the best of our knowledge, the only study of this type on the UK is Goodridge & Haskel (2015), which estimates that in 2010, "big data" contributed £5.7 billion to UK GDP.[36] This study uses data from the Annual Survey of Hours and Earnings (ASHE) to compute total labour costs in "big data" occupations. Engagement with industry experts is used to assess the proportion of these costs used to build data assets in each occupation. More recently, a similar approach was used in Statistics Canada (2019).

Tambe et al. (2020) combine data on labour costs with data on market and asset values for large US businesses. They use econometric techniques to assess the value of the businesses' "digital capital".[37] They find that digital capital accounted for at least 25% of firms' assets by 2016.

As discussed in more detail in Annex B (focus box 1), using information on job postings (e.g. salary, competencies, number of positions) represents a growing opportunity to measure the value of the data assets which underpin different occupations using a cost-based method.

Market-based methods

Market-based methods have only been used in a handful of existing studies. This likely reflects the fact that data are only relatively rarely exchanged in market transactions.

The most recent study to use market prices of data is an EU-level study by the International Data Corporation (Cattaneo et al., 2020). This study estimated the revenues of "data suppliers" in the UK plus the UK at €83 billion in 2019.[38] This estimate was obtained through a combination of proprietary data sources (e.g. IDC Core IT Spending Guide, IDC Worldwide Black Book), desk research and publicly available statistics from Eurostat and the International Monetary Fund.

Methods using the market value of companies have been used in a handful of studies on US businesses. These studies have estimated the proportion of a business' market value that can be attributed to information technology (IT) assets.[39] In the future, similar methods could be used to assess more specifically the value of data-related assets.

---

36 This is around 25% of the total UK gross domestic R&D expenditure in 2010, £26.2 billion according to ONS (2013, Figure 1).

37 Defined as "factors of production 1) that are complementary to recorded investments in IT assets (such as hardware and software), but 2) that are not otherwise recorded on a firm's balance sheets. Examples include employee training which is related to new information technologies, firm-specific human capital related to technology systems, and the development and implementation of business processes and other forms of organizational transformation required to support or use new information technologies".

38 Note: the IDC study does not provide a formal definition of "data suppliers".

39 For example, using market-based methods, Brynjolfsson et al. (2002) estimate that for every $1 asset value of IT hardware included in businesses' P&L accounts, there are additional $9 in value of related assets that are not recorded. Saunders and Brynjolfsson (2016) extend this finding for a sample of 127 firms in 2003-06. They found that the IT services which were not

Use-based methods

In general, the empirical literature on the benefits of data to firms, consumers and the wider economy is relatively nascent. However, some useful studies have been published which could serve as a starting point for building out a framework on the benefits of data.

There is some literature on the **performance of data-intensive firms**. For example, Brynjolfsson et al. (2011) found that US firms which adopt data-driven innovation have output and productivity that is 5-6% higher than what would be expected given their other investments and information technology usage.[40]

To measure the **specific benefits of data use**, AlphaBeta (2017) estimatedd that digital maps reduced travel time by 12% on average, that consumers valued digital maps at up to $105 per user and that geospatial data saved consumers more than 21 billion hours in 2016, with related reductions in congestion and pollution.[41]

Some studies have also used **contingent analysis**. Coyle & Nguyen (2020) applied contingent valuation techniques to estimate how much consumers value different ranges of digital services, based on survey data which elicits willingness to accept the loss of 30 goods or services before and during the COVID-19 pandemic. This approach was a useful option for measuring the value of digital goods and services when consumers often pay no upfront cost for these services. The study found that for many digital services, consumers' willingness to pay is higher than estimated revenues per user, suggesting that a substantial part of the value of digital services is not measured in official statistics.

As this study is trying to value digital services (which are data-intensive) rather than valuing data assets per se, using a similar approach to value data assets could create the same challenges of using company valuations mentioned in the previous section.[42]

accounted for on balance sheets were being priced into the market value of firms. Of this value shortfall, there was a 45% to 76% premium in market value for firms with the highest IT capabilities, compared to firms with the lowest IT capabilities.

40 Similarly, Bahkshi et al. (2014) found that UK firms in the top quartile of online data use were, other things being equal, 13% more productive than those in the bottom quartile. Following a more bottom-up approach, Hogan et al. (2020) estimated that from 2015 to 2020, the benefit to the UK economy of big data analytics was £40 billion on average per year, equivalent to 2.0% of annual GDP.

41 Similarly, Sadlier et al. (2018) estimated the benefits of satellite Earth Observation data across a number of uses for the UK government and forecast the total value for government applications across nine civil use cases to reach £1.2 billion per year by 2020.

42 In other words, variations in contingent valuations of services that are more/less data-intensive can be exploited and used as a proxy for the value of the underlying assets, but it will be challenging to disentangle whether changes in the valuation are really due to differences in data use rather than other factors which may not be measurable.

ANNEX H MEASURING DATA CHARACTERISTICS

As mentioned in section 3.2 above, our research indicates that, in general, it would not be feasible or particularly informative to quantitatively measure the impact that each characteristic has on the value of a data asset. This is because:

- Many characteristics drive value together as a “bundle”;
- Value will depend on many other relevant factors, such as organisational characteristics and use cases; and
- Data sources which could be used to measure the impact of different characteristics on the values of a wide range of data assets do not currently exist.

This is illustrated in a worked example in Annex A assessing public policy interventions which affect the characteristics of data assets.

In other words, as many of the organisations involved in our proof-of-concept exercise mentioned, data characteristics such as consistency and size are similar to the ingredients of a cake: it is difficult and, in many instances, not particularly meaningful to disentangle the impact of the flour from the impact of the eggs in determining the final quality of the cake.

Against this backdrop, valuing individual characteristics could be useful for understanding how different users value specific aspects of specific datasets and, in the case of changes, which affect the sensitivity and confidentiality of the asset. This Annex includes some limited examples of where this exercise could be performed.

Conjoint valuation of specific characteristics

In some very specific circumstances, conjoint analysis could be used to disentangle the value of different data characteristics. For example, the ONS recently used this approach to give a monetary value to a series of attributes of earnings data. These were: source (official or non-official), frequency (monthly or annually), geography (regional or national) and price, which was used to derive value (free, £1,500 or £5,000).

From our perspective, this approach cannot be widely applied as it relies on a series of strict assumptions to be genuinely indicative of the value of a specific characteristic as opposed to others. First, the list which the respondents are shown needs to be exhaustive and very concise (to avoid asking about dozens of different combinations). Second, the sample of interviewees needs to be representative of the entire sample of data users and of use cases. Third, respondents need to have perfect information on how the data assets are used and how value is generated from them (ONS, 2021).

Changes which affect the sensitivity/confidentiality of the asset

Similarly, for changes which affect the sensitivity/confidentiality of the asset, the impact that previous cases of non-compliance have had on the market value of similar organisations could be used as a proxy for the impact that specific risks and liabilities have on the value of the data asset under analysis. In light of the discussion in section 2 of this report, this approach could be classified as a market-based methodology (IBM, 2021).

“Data-centric and liquid” markets

Over the last decade, data has become the most important asset in financial markets, with 85% of the banks, investors and capital markets service providers planning to increase their spending on data management technology in the next three to five years (Refinitiv and Greenwich Associates, 2019). In this context, most market participants expect large financial market data aggregators (such as Bloomberg and Refinitiv) to continue to be their primary data source. In industries (such as the financial sector) which are characterised by a high number of data transactions, relatively standardised products, the presence of a manageable number of data aggregators and highly sophisticated and well-informed customers, it may be possible to analyse the price and the characteristics of different data assets in order to isolate the relative value of a specific set of characteristics (such as timeliness or granularity).

ANNEX I REFERENCE LIST

ACCC and PwC (2008). Review of ARTC DORC Valuation Assessment of the Reasonableness of the Proposed Valuation for the 10 Year Access Undertaking.

Acemoglu, D., A. Makhdoumi, A. Malekian & A. Ozdaglar (2019). Too Much Data: Prices and Inefficiencies in Data Markets. NBER Working Paper 26296. National Bureau of Economic Research.

AlphaBeta (2017). The Economic Impact of Geospatial Services: How Consumers, Businesses and Society Benefit from Location-based Information.

Andrews, D., C. Criscuolo & P.N. Gal (2019). The Best versus the Rest: Divergence across Firms during the Global Productivity Slowdown. CEP Discussion Paper No 1645.

Arquié, A., L. Demmou, G. Franco & I. Stefanescu (2019). Productivity and Finance: The Intangible Assets Channel.

Bakhshi, H., A. Bravo-Biosca & J. Mateos-Garcia (2014). The Analytical Firm: Estimating the Effect of Data and Online Analytics on Firm Performance. Nesta Working Paper No. 14/05.

Bartz-Zuccala, W., P. Mohnen & H. Schweiger (2018). The Role of Innovation and Management Practices in Determining Firm Productivity. Journal of Economics, 60: 502-530.

Bergemann, D., A. Bonatti & T. Gan (2020). The Economics of Social Data. Cowles Foundation Discussion Paper No. 2203r2. Yale University.

Bianchini, M. & V. Michalkova (2019). Data Analytics in SMEs: Trends and Policies. OECD SME and Entrepreneurship Papers, No. 15, OECD Publishing, Paris.

Brynjolfsson, E., L.M. Hitt & H.H. Kim (2011). Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? SSRN.

Brynjolfsson, E., L.M. Hitt & S. Yang (2002). Intangible Assets: Computers and Organizational Capital. Paper 138. Center for eBusiness @ MIT.

Brynjolfsson & McElheran (2016). The Rapid Adoption of Data-Driven Decision-Making. American Economic Review, 106(5): 133-139.

Cattaneo, G., G. Micheletti, M. Glennon, C. La Croce & C. Mitta (2020). The European Data Market Monitoring Tool Key Facts & Figures, First Policy Conclusions, Data Landscape and Quantified Stories. European Commission.

CGG (2016). Impact of IAS 38 Amendment on Multi-Client Library.

Choi, J. P., D.-S. Jeon & B.-C. Kim (2019). Privacy and Personal Data Collection with Information Externalities. Journal of Public Economics, 173: 113-124.

Collins, A., D. Coughlin, J. Miller & S. Kirk (2015). The Production of Quick Scoping Reviews and Rapid Evidence Assessments: A How to Guide. DEFRA and NERC.

Coyle, D., S. Diepeveen, J. Wdowin, J. Tennison & L. Kay (2020). The Value of Data: Policy Implications. Cambridge: Bennett Institute for Public Policy and the Open Data Institute.

Coyle, D. & D. Nguyen (2020). Valuing Goods Online and Offline: The Impact of Covid-19. Economic Statistics Centre of Excellence Discussion Papers 2020-10.

Crass, D. & B. Peters (2014). Intangible Assets and Firm-level Productivity. Discussion Paper No. 14-120. Centre for European Economic Research.

Crémer, J., Y.-A. de Montjoye & H. Schweitzer (2019). Competition Policy for the Digital Era. European Commission.

DCMS (2020). National Data Strategy. Policy Paper. Department for Digital, Culture, Media and Sport.

Debussche, J., J. César & I. De Moortel (2019). Big Data & Issues & Opportunities: Open Data. Bird & Bird.

Deloitte (2020). Data: Could It Be Your Most Valuable Asset?

Digital Competition Expert Panel (2019). Unlocking Digital Competition: Report of the Digital Competition Expert Panel.

DIHAH (n.d.). Poultry Dashboard: Distribution and Epidemiological Patterns for IBV. Data Innovation Hub for Animal Health.

Drieger (2015). All Aboard with Infrastructure 4.0 – Splunk Wins Deutsche Bahn Internet of Thinkgs Hackathon.

Duch-Brown, N., B. Martens & F. Mueller-Langer (2017). The Economics of Ownership, Access and Trade in Digital Data. JRC Digital Economy Working Paper 2017-01. European Commission.

EC, IMF, OECD, UN & World Bank (2008). System of National Accounts 2008.

ECB (2001). Trade-off Between Timeliness and Accuracy ECB Requirements for General Economic Statistics. European Central Bank.

European Parliament (2019). Re-use of Public Sector Information. Briefing – EU Legislation in Progress.

Eurostat (2015). File: Factors limiting enterprises from using cloud computing services, by enterprise size class, EU-28, 2014 (% of enterprises using the cloud) YB15-de.png.

Eurostat (2019). European Statistical Recovery Dashboard.

Flamholtz, E.G. & Y. Randle (2011). Corporate Culture: The Ultimate Strategic Asset. Stanford University Press.

Frontier Economics (2020). Geospatial Market Data. Report for the Geospatial Commission.

Frontier Economics (2021). Increasing Access to Data across the Economy. A report prepared for the Department for Digital, Culture, Media and Sport.

FSB (2018). Data Ready: Mitigating the Impact of Data Protection Regulation on Small Businesses.

Ginsburg, G.S. & K.A. Phillips (2018). Precision Medicine: From Science to Value. Health Affairs (Millwood), 37(5): 694-701.

Global Partnership for Sustainable Development Data (2018). What Do We Know About the Value of Data?

GoFair (n.d.). FAIR Principles.

Goodridge, P. & J. Haskel (2015). How Much is UK Business Investing in Big Data? Imperial College London Business School.

Government Data Quality Hub (2020). Personal Information Charter.

GSMA (2018). The Data Value Chain.

Haldane, A. (2018). The UK's Productivity Problem: Hub No Spokes. Academy of Social Sciences Annual Lecture, London, 28 June.

Haldane, A., & S. Chowla (2020). Fast Economic Indicators. Nature Reviews Physics 3(68-69).

Hashimzade, N., G. Myles & J. Black (2017). A Dictionary of Economics (5 ed.). Oxford, Oxford University Press.

Haskel, J. & S. Westlake (2018). Capitalism without Capital: The Rise of the Intangible Economy. Cato Journal.

Henze, M., R. Matzutt, J. Hiller, E. Muhmer, J.H. Ziegeldorf, J. van der Giet & K. Wehrle (2020). Complying with Data Handling Requirements in Cloud Storage Systems.

HM Treasury (2020). The Green Book. Central Government Guidance on Appraisal and Evaluation. HM Treasury.

Hogan, O, L. Holdgate & R. Jayasuriya (2016). The Value of Big Data and the Internet of Things to the UK Economy. Report for SAS.

IBM (2021). Cost of A Data Breach Report.

Isaak, J. & M. J. Hanna (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. Computer, 51(8): 56-59.

Kemal, M., R. Sanchez, R. Olsen, F. Lov & H.-P. Schwefel (2020). On the Trade-off Between Timeliness and Accuracy for Low Voltage Distribution System Grid Monitoring Utilizing Smart Meter Data. International Journal of Electrical Power & Energy Systems, 12.

Kim, T., K. Barasz & L. K. John (2018). Why Am I Seeing This Ad? The Effect of Ad Transparency on Ad Effectiveness. Journal of Consumer Research, 45(5): 906–932.

KPMG (2019). Data as an Asset: Initiate Your Journey to Unlock Data's Full Potential.

Kremser, W. & R. Brunauer (2019). Do We Have a Data Culture? In P. Haber, T. Lampoltshammer and M. Mayr (eds), Data Science – Analytics and Applications, pp. 83-87.

Law, J. (2009). A Dictionary of Business and Management (5 ed.). Oxford, Oxford University Press.

Leal, G., W. Guédria & H. Panetto (2019). Interoperability Assessment: A Systematic Literature Review. Computers in Industry, 106: 111-132.

Li, W.C.Y., M. Nirei & K. Yamana (2019). Value of Data: There's No Such Thing as a Free Lunch in the Digital Economy. VOX EU CEPR.

Lindsay, M. (2015). Finding Use Cases for Big Data. TDWI.

Locatelli, G., M. Mancini & G. Lotti (2020). A Simple-to-Implement Real Options Method for the Energy Sector. Energy, 197.

Maddison, J. & J. D'Addario (2020). Case Study: The Value of Sharing Data for Benchmarking and Insights. Open Data Institute.

McKinsey (2018). Alejandro Díaz, Kayvaun Rowshankish & Tamim Saleh. Why Data Culture Matters.

McKinsey Global Institute (2016). Digital Globalization: The New Era of Global Flows.

Motohashi, K. (2017). Survey of Big Data Use and Innovation in Japanese Manufacturing Firms. RIETI Policy Discussion Paper FY2016.

Mitchell, A. (2019). A Practical Vision for Personal Data Empowerment. Mydex CIC.

National Archives (2017). Information Asset Factsheet.

Nguyen, D. & M. Paczos (2020). Measuring the Economic Value of Data and Cross-border Data Flows: A Business Perspective. OECD Digital Economy Papers, No. 297, OECD Publishing.

NIC Data (2020). AKZONOBEL: How a Global Company Learned to Harness Its Data Science Skills and Create Products for the Future. Case Study.

Nolin, J.M. (2019). Data As Oil, Infrastructure or Asset? Three Metaphors of Data As Economic Value. Journal of Information, Communication and Ethics in Society. https://www.emerald.com/insight/content/doi/10.1108/JICES-04-2019-0044/full/html

Norwegian Petorleum Directorate (2017). The Barents Seas North: Geological Assessment of Petroleum Resources in Eastern Parts of Barents Sea North.

ODI (n.d.). The Value of Sharing Data for Benchmarking and Insights: A Case Study in the Shipping Industry. Open Data Institute.

OECD (2003). Glossary of Statistical Terms. Organisation for Economic Co-operation and Development.

OECD (2005). Glossary of Statistical Terms. Organisation for Economic Co-operation and Development.

OECD (2006). Glossary of Statistical Terms. Organisation for Economic Co-operation and Development.

OECD (2017). OECD Digital Economy Outlook 2017. Organisation for Economic Co-operation and Development..

ONS (2013). Gross Domestic Expenditure on Research and Development, UK: 2011. Office for National Statistics.

ONS (2018). Introducing a New Publication Model for GDP. Office for National Statistics.

ONS (2020). Firm-level Labour Productivity Measures from the Annual Business Survey, Great Britain: 1998 to 2018. Office for National Statistics.

ONS (2021). Valuing Official Statistics with Conjoint Analysis. Office for National Statistics.

PGS (2016). PGS Changes MultiClient Library Amortization from 1 January 2016.

PwC (2019). 22nd Annual Global CEO Survey: CEOs' Curbed Confidence Spells Caution 2.

PwC (n.d.). Putting a Value on Data.

Ravasi, D. & M. Schultz (2006). Responding to Organizational Identity Threats: Exploring the Role of Organizational Culture. Academy of Management Journal, 49(3).

Refinitive & Greenwich Associates (2019). The Future of Trading: Redefining Data.

Rico, P. & B. Cabrer-Borrás (2019). Intangible Capital and Business Productivity. Economic Research, 33(1): 3034-3048.

Sadlier, G., R. Flytkjaer, F. Sabri & N. Robin (2018). Value of Satellite-derived Earth Observation Capabilities to the UK Government Today and by 2020: Evidence from Nine Domestic Civil Use Cases. London Economics.

Saunders, A. & E. Brynjolfsson (2016). Valuing Information Technology Related Intangible Assets. MIS Quarterly, 40(1): 83-110.

SROI Network (2012). A Guide to Social Return on Investment.

Statistics Canada (2019). The Value of Data in Canada: Experimental Estimates. Latest Developments in the Canadian Economic Accounts.

Tambe, P., L. Hitt, D. Rock & E. Brynjolfsson (2020). Digital Capital and Superstar Firms. NBER Working paper 28285. National Bureau of Economic Research.

USGS (n.d.). USGS Geospatial Data Sources.

Van Asbroeck, B., J. Debusshe & J. César (2019). Big Data & Issues & Opportunities: Data Ownership. Bird & Bird.

Vandevelde, M. (2018). Sainsbury's Buys Nectar Reward Scheme from Aimia for £60m. Financial Times.

Vernengo, M., E.P. Caldentey & B.J. Rosser Jr. (2008). The New Palgrave Dictionary of Economics. Palgrave Macmillan, London. https://link.springer.com/referencework/10.1057/978-1-349-95121-5

Zolas, N., Z. Kroff, E. Brynjolfsson, K. McElheran, D.N. Beede, C. Buffington, N. Goldschlag, L. Foster & E. Dinlersoz (2020). Advanced Technologies Adoption and Use by U.S. Firms: Evidence from the Annual Business Survey. NBER Working Paper 28290. National Bureau of Economic Research.

frontier economics