

INCREASING ACCESS TO DATA ACROSS THE ECONOMY

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

March 2021



CONTENTS

Executive Summary	4
1 Introduction	13
2 Rationales for Government intervention	15
2.1 Setting the market framework	15
2.2 Addressing market failures	15
2.3 Adjusting market outcomes	18
2.4 Selected characteristics of data	19
3 Evidence of barriers to access to data	21
3.1 The data sharing journey	21
3.2 Issues that may prevent optimal data sharing	22
4 Determining the set of levers for Government intervention	34
4.1 Defining levers for government intervention	34
4.2 Mapping data-sharing issues to levers	35
4.3 Logic models	39
5 Our approach to assess levers for intervention	46
5.1 Overview of our approach	46
5.2 Market sizing	47
5.3 High-level assessment of the levers	49
5.4 Assessing how lever effectiveness may vary in different ecosystems	51
6 Initial application of the framework	56
6.1 Initial market sizing	56
6.2 Applying the assessment criteria	60
7 Conclusions	67
7.1 Summary of our findings	67
7.2 Opportunities for future research	67
8 References	70
Annex A Further detail on evidence of data sharing issues	74
Annex B Further detail on case studies of data sharing ecosystems	81

Executive Summary

The Department for Digital, Culture, Media and Sport (DCMS) recognises that the full value of data held by organisations in the private and third sectors may not be unlocked without government intervention. Increasing access to and use of data has the potential to generate significant economic and social benefits, from increased productivity to supporting research and the delivery of public services.¹

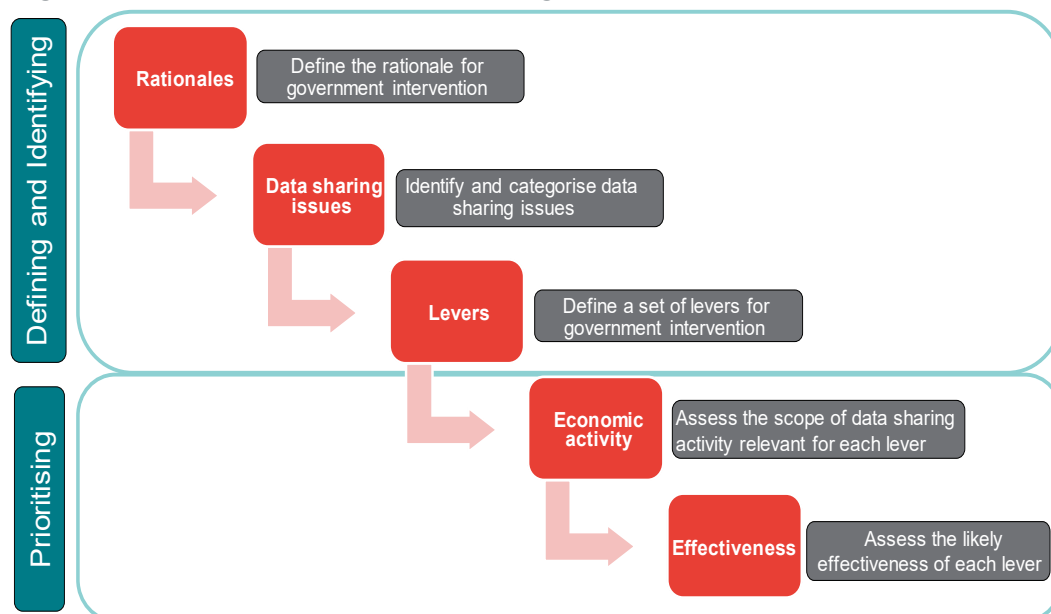
DCMS wants to ensure that decisions about appropriate policy interventions in this arena draw on the best available knowledge and evidence in order to focus on those areas where the returns from government action are expected to be highest.

As a first stage in this process, DCMS commissioned Frontier Economics to create an evidence-based framework which will help the department prioritise possible interventions related to increasing the availability of data (henceforth “access to data”).

A framework to assess possible interventions

The framework defined through this study consists of five steps, as shown below.

Figure 1 A framework for assessing possible interventions



Within this report, we have completed steps 1 to 3 and provide an initial application of the approach for steps 4 and 5. We also make a set of recommendations as to how each of the steps in this framework could be improved in future work.

The analysis in this report indicates that a package of effective interventions could incorporate all of the following: improving knowledge and understanding of data sharing; improving or demonstrating incentives; supporting ways to address risk;

¹ DCMS (2020), OECD (2015).

reducing the cost of sharing through data foundations (e.g. improved findability and interoperability of data); reducing the (perceived) regulatory burden; and mandating data sharing.

Addressing risk and reducing the cost of sharing are the levers that, based on our initial assessment, are likely to have a direct effect on the largest amount of economic activity. However, improving incentives and increasing knowledge of data sharing are particularly likely to have wide-ranging, indirect effects. Research on the effectiveness of existing interventions is very limited, but there is some evidence for the effectiveness of demonstration activities and for mandating data sharing where there is a clear case that this could lead to the development of additional services (e.g. current account comparison services in the case of Open Banking) or to increasing choice and competition.

However, given the breadth of economic activity where additional data sharing may be beneficial, the range of issues that may prevent sharing and the sparsity of the evidence base, these conclusions should be interpreted with caution. This study provides a starting point for further development of public policy in this field rather than a set of firm conclusions that could, in isolation, inform the design of evidence-based policy interventions.

Rationales for government intervention

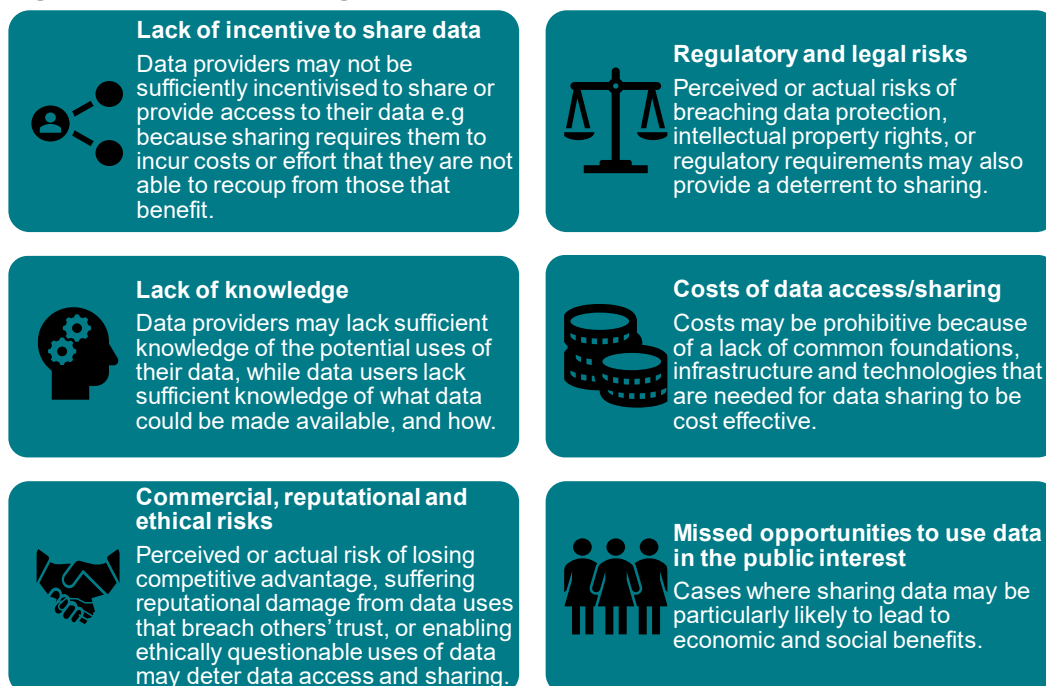
The starting point for this study is a rapid review of the literature to identify the key reasons why government might be well placed to increase access to data across the economy. Setting out these reasons allows us, in the following phase of the study, to focus on barriers or opportunities that would not be unlocked, or would be unlocked more slowly, in a counterfactual of no government intervention. Otherwise, government action will simply displace private sector activity.

We broke down rationales for government intervention into three categories:

- Setting the market framework: for example, considering whether the intellectual property protection regime strikes an appropriate balance between protecting the rights of organisations that have invested in data, versus allowing others to access the data;
- Addressing market failures: for example, incentivising organisations to share data where the social gain of doing so is larger than the benefit of keeping it private;
- To adjust market outcomes: for example, protecting consumers or introducing competition where market outcomes lead to monopolies, and the threat of entry into the market from other companies is not sufficient to promote desired outcomes for consumers.

Identifying and categorising data sharing issues

Following a rapid review of the literature, we have identified six data sharing issues where there is a rationale for government intervention. These include both barriers to greater data access and opportunities that could be unlocked from wider access.

Figure 2 Data sharing issues

The levers available for government to use

The six data sharing issues translate into six potential levers for government action, described in Figure 3 below.² We think of a lever as something that government could manipulate, at least in principle, and would have an impact on one of the “access to data” issues set out above.

² In line with the National Data Strategy, we define investment in data foundations as making data “fit for purpose, recorded in standardised formats on modern, future-proof systems and held in a condition that means it is findable, accessible, interoperable and reusable”. (DCMS, 2020).

Figure 3 Levers for intervention**1 Improve knowledge and understanding of data sharing**

For example, by establishing convening organisations, disseminating information, delivering data sharing projects or pilots

2 Reduce costs of data sharing through better data foundations

For example, by supporting innovation in more efficient data sharing solutions; developing standards and encouraging or mandating the adoption of standards; supporting data stewarding initiatives

3 Support (new) ways of addressing the risks of data sharing

For example, by supporting innovation in secure data sharing solutions or investing in data sharing platforms

4 Improve or demonstrate incentives for data sharing

For example, through testbeds and trials demonstrating the benefits of data sharing; financial or reputational incentives for data sharing and access

5 Reduce (perceived) regulatory burden associated with data sharing

For example, disseminating information and providing advice on regulatory compliance; creating or expanding regulatory/competition sandboxes

6 Mandate data sharing in the public interest

For example, by identifying datasets of national importance or public interest datasets

Each lever has the primary objective of addressing a specific type of data sharing issue. However, all levers may also have secondary effects on other types. Many aspects of data sharing are complex and straddle several of the issues described above. For example, we have categorised the perception of some organisations that data sharing poses a high risk that is best tackled by a lever focused on reducing the associated dangers. But it could also be considered a “knowledge” issue, if organisations actually don’t fully understand the risks involved. Therefore, the lever focused on improving understanding of data sharing would primarily address “lack of knowledge”, but it may also have a secondary effect on “commercial, reputational and ethical risks”. Figure 4 maps the levers to each type of data sharing issue: the primary effects are reported in blue and the secondary effects in yellow.

Figure 4 Mapping of levers to data sharing issues

Lever	Types of data sharing issues					
	Lack of incentives	Lack of knowledge	Commercial, reputational and ethical risks	Legal and regulatory risks	Costs of data access/sharing	Using data in the public interest
Improve knowledge	✓	✓	✓		✓	
Reduce cost	✓	✓			✓	
Address risks			✓	✓		✓
Improve / demonstrate incentives	✓	✓				✓
Reduce regulatory burden				✓		
Mandate data sharing			✓			✓

Note: ✓ indicates primary effects of levers; ✓ indicates secondary effects.

The economic activity in scope of each lever

Using the available evidence on data sharing issues in different industries and ecosystems, we have generated an initial assessment of the economic activity that may be positively affected by each lever. This provides an initial long list of ecosystem and lever combinations with the greatest potential to unlock significant value by addressing data sharing concerns. This list needs to be evaluated further to ensure the levers are well matched and appropriate for the ecosystems or industries that appear to have the highest potential value.

This initial assessment draws on current data but has several limitations.

- The existing evidence of data sharing issues is not comprehensive so, while we have undertaken extensive searches, there is a reasonable chance that we have missed some relevant ecosystems. This could be addressed by additional primary data collection.
- Our estimates of the economic impact reflect the size of current activity in those ecosystems where each lever could be applicable, rather than the potential increase in economic activity that might result from using each lever. This is likely to be more difficult to address in the future. We have used evidence on both Gross Value Added (GVA) and the size of the data market in each ecosystem to provide two views of the extent of economic activity.

Lever effectiveness

Estimating the economic activity potentially in scope allows us to rank the levers according to their potential impact. It is a high-level estimate of the “size of the

prize” that could be unlocked if each lever were used. The exercise is driven by both the prevalence of various data sharing issues across industries or ecosystems as well as the economic value of the ecosystems identified.

It does not take account of the extent to which a given lever would be effective at addressing the relevant data sharing issues relative to a counterfactual of no policy intervention. This requires assessing the lever against a number of further criteria. We divide these criteria into those for which an evaluation can be made at a lever level (general criteria) and those which require assessment at an ecosystem level (ecosystem criteria). We recognise that it is unlikely to be feasible for DCMS to review all possible lever/ecosystem combinations when prioritising policy action. We therefore suggest that a pragmatic compromise would be to focus the ecosystem-level assessment on the top two or three ecosystems that are pertinent for each lever.

Figure 5 Assessment criteria

General assessment criteria		
Criterion	Description	Assessed in this report?
Strength of evidence	a. Strength of data policy-specific evidence: is there any evidence that interventions relevant to this lever have been effective, relative to a counterfactual of no intervention? How robust is this evidence? b. Strength of evidence in other areas: are there effective interventions under this lever in other policy areas (e.g. innovation policy, trade policy, ...)?	In part
Timing of impact	a. What stage of a hypothetical data sharing journey does this lever act on? (early-middle-late) b. Are there existing government interventions relevant under this lever that could be expanded/adapted?	Yes
Indirect effects of lever	Is the lever likely to have an effect on barriers other than the one it intends to address? For example, does improving incentives for data sharing have an indirect impact on the risks from data sharing?	Yes
Unintended consequences and trade-offs	What trade-offs should be considered when applying this lever? What is the risk of adverse effects?	Yes
Cost of intervention	How much would this intervention cost, considering both financial and in-kind inputs?	No – could be assessed in future work
Ecosystem-specific assessment criteria		
Gap analysis	Are there existing private-, third- or public-sector initiatives to increase access to data in this ecosystem? Might there be a role for government to support any of these initiatives?	No – could be assessed in future work
Lever interaction	Are there several types of data sharing issues in this ecosystem? If so, is the lever more likely to be effective if other levers are also used first or at the same time?	No – could be assessed in future work

General assessment criteria		
Criterion	Description	Assessed in this report?
Complementary conditions	Are the conditions in place for data sharing to have a positive impact on economic and social outcomes? For example, what is the overall level of digitalisation in this ecosystem? Is there sufficient access to the skills required to use data effectively?	No – could be assessed in future work
Drivers of effectiveness	Do the characteristics of this ecosystem indicate that this lever is particularly likely or unlikely to be effective?	No – could be assessed in future work

Assessment against general criteria

Figure 6 below summarises our initial assessment of the levers against the general criteria.

Figure 6 Summary of general lever assessment

Lever	Key advantages	Key disadvantages
Improve knowledge / understanding of data sharing	Some evidence of potential for positive impact on data sharing; possible large indirect effects	Relatively slow impact
Reduce costs of data sharing through better data foundations	Some evidence of positive impact on data sharing; possible indirect effects; relatively quick impact in improving the efficiency of existing data sharing	Relatively slow impact in leading to additional data sharing; mandating standards may hinder innovation
Support (new) ways to address the risks associated with data sharing	Relatively swift impact once ways to address risk are established	Impact may be relatively slow if developing new ways to address risk requires significant innovation
Improve / demonstrate incentives for data sharing	Possible large indirect effects	Risk of incentivising data sharing that may not ultimately lead to significant economic and social benefits, or that may reduce trust in and trustworthiness of data sharing
Reduce (perceived) regulatory burden associated with data sharing	Relatively quick impact on data sharing (but slow implementation of lever if requiring legislation)	Risk of reducing trust in and trustworthiness of data sharing, which may lead to lower sharing volumes
Mandate data sharing where this is in the public interest	Some evidence of positive impact on data sharing	Risk of disincentivising data collection; risk of unfairly putting incumbents at a competitive disadvantage

Assessment against ecosystem-specific criteria

The ecosystem criteria involve assessing whether the characteristics of any particular ecosystem mean that a lever is more or less likely to be effective. Ecosystems vary hugely, so in principle this could involve an extensive exercise of

identifying all possible characteristics of an ecosystem and considering the potential impact they might have on the effectiveness of a lever. This would not be tractable, so we explored six case studies of data sharing, reported by the Open Data Institute and the Royal Academy of Engineering,³ to identify a number of key characteristics. Figure 7 overleaf summarises our findings. We have applied economic theory to draw general lessons based on the six studies.

Figure 7 Summary of findings from existing case studies

Lever	Drivers of effectiveness
Improve /demonstrate incentives	<p>This lever may be more likely to lead to additional data sharing where:</p> <ul style="list-style-type: none"> the data is dispersed (there are many organisations potentially involved in the ecosystem); using the data requires integration to reach a sufficient scale; and the benefits from using the data are not immediately obvious/rely on innovation/take some time to materialise
Reduce costs of data sharing through better data foundations	<p>This lever may be more likely to lead to additional data sharing where trusted relationships between data users and providers already exist or can be developed/nurtured.</p> <p>Vice versa, the additionality of this lever may be lower where data is being shared by many owners with one large-scale intermediary or user who can monetise the data</p>
Improve /demonstrate incentives and Improve knowledge / understanding of data sharing	<p>These levers can be used to support the development of data ecosystems where data-driven products have not yet been created and/or customer take-up of those services is likely to be relatively slow in the absence of intervention.</p>
Addressing risk	<p>Where trust between data users and providers has not yet been established, using this lever could help interventions that reduce the cost of data sharing to achieve the objective of increasing access to data.</p>
Reduce (perceived) regulatory burden, mandate data sharing	<p>Limited evidence from existing ecosystem case studies. Could be explored in future case studies.</p>

Opportunities for future research

Opportunities for future research to support the development of policy interventions in this area include the following:

- Primary research to fill gaps in the existing evidence base on the issues that prevent data sharing, and in particular on where they arise: which industries, ecosystems, types of organisation and types of data are especially affected by each issue.
- A market and ecosystems review to build on our initial sizing analysis. This could involve:

³ Royal Academy of Engineering (2019) and ODI (2020).

- A more precise assessment of the current size of data ecosystems; and
- A more precise assessment of the importance of data sharing issues in the ecosystem and of the activities that may be unlocked by addressing those problems. Existing evidence mainly describes in very general terms the types of data sharing concerns that arise in an industry. Stakeholder engagement and/or primary data collection could strengthen the evidence base by investigating this in more detail.
- Case studies of data sharing ecosystems, to investigate how the characteristics of ecosystems may determine the effectiveness of levers for intervention.
- Quantitative analysis to look into the impact of data sharing on organisational performance.
- Developing frameworks to assess and categorise the types of benefits that may arise from data sharing.

For all of these research areas, it may be useful to investigate specifically which data sharing issues exist in the third sector, their role and the value that could be generated by addressing them.

1 Introduction

The objectives of this report

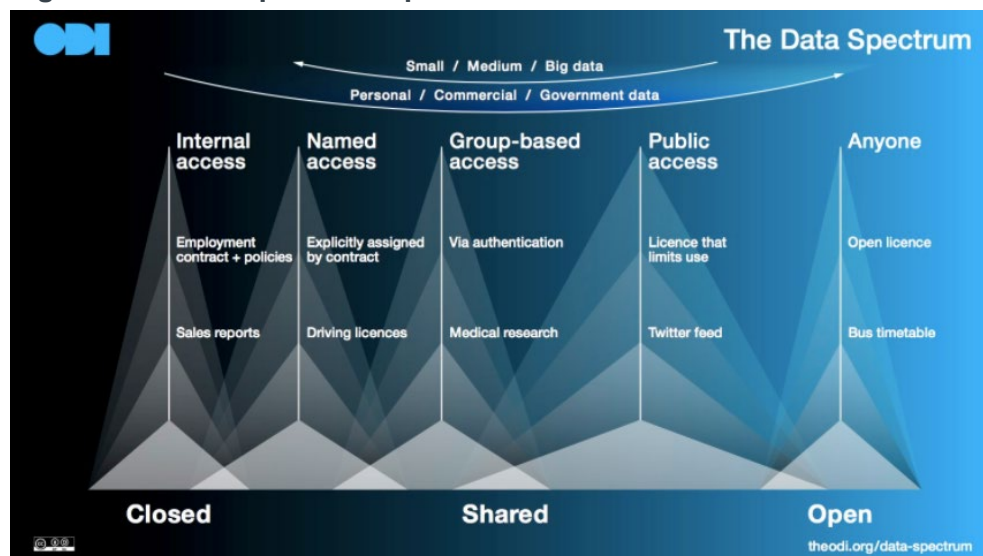
The Department for Digital, Culture, Media and Sport (DCMS) is looking to improve its knowledge and understanding of the evidence on how and when government intervention can best increase the availability of data generated outside the public sector, by both the private sector and the third sector, e.g. voluntary groups, non-profit organisations and social enterprises (referred to from here on in this report as "access to data"). Expanding access to data and making greater use of it has the potential to generate significant economic and social benefits, from increasing productivity to supporting research and delivering public services.⁴

To work towards this objective, DCMS has commissioned Frontier Economics to create an evidence-based framework which will help the department to prioritise possible interventions. This project focuses on how government could increase the availability of data that already exists at least in "raw" form and could be processed and analysed to generate insights. The emphasis, then, is on expanding access to existing data, or in other words how to increase data sharing. Throughout this report we refer interchangeably to "increasing access" and "greater data sharing" as objectives of government intervention.

Types of data access in scope

DCMS is interested in exploring different levels of data availability along the open data spectrum, from named access to fully open access, as described in Figure 8. Increased availability of data potentially includes providing more access to other private- or third-sector organisations in the UK, organisations outside the UK, central and local government, and individuals.

Figure 8 ODI open data spectrum



Source: Open Data Institute

⁴ DCMS (2020), OECD (2015).

Access to data is managed in many ways, from one-off licensing arrangements between a specialist data provider and a user to more complex data ecosystems that involve many organisations both providing and using data through a central platform. In principle, all of the possible technical means, business models and legal arrangements through which data is accessed fall within the scope of this study.

Throughout the report, we define the context in which data sharing is or may be happening as a “data ecosystem”. A data ecosystem includes “the people and organisations involved in either creating outputs using data, or benefiting from its use”.⁵ Data ecosystems may include many actors, at times linked by complex relationships. In this report, for simplicity, we generally refer to two categories of actors: “data providers” and “data users”. When describing some of the possible interventions to increase access to data, and when describing examples of existing ecosystems, we also refer to “data intermediaries”.⁶

Our approach and the structure of this report

Our approach consists of the following steps:

- An initial review of evidence on the rationales for government intervention in the area of access to data, including discussions with six experts and stakeholders in the public and third sectors. Findings from this review are reported in Section 2.
- A further review of evidence on what specific issues may prevent or limit access to data, and how they are related to rationales for government intervention. This is described in Section 3.
- Defining categories of government intervention that could address the issues described in Section 3. We define these categories as “levers” for intervention, and describe them in Section 4.
- Developing a framework for how DCMS might go about prioritising possible interventions. This framework is outlined in Section 5.
- Finally, in Section 6, we provide an initial assessment of the likely impact of each lever. The evaluation includes:
 - Reporting and analysing data on the Gross Value Added generated by industries where each lever could be applicable, and information on the value of data exchanges taking place in each industry (Section 6.1); and
 - Defining and applying high-level qualitative assessment criteria to compare the levers (Section 6.2).
- Section 7 concludes and provides indications for future research in this field.

⁵ ODI – *Mapping Data Ecosystems* (2018)

⁶ In this report, we define “data users” as organisations that generate insights, products or services using data; “data providers” as organisations that can provide access to data users; and “data intermediaries” as any organisation that acts as intermediary between data providers and users. Data intermediaries may, for example, access data from several providers, aggregate and anonymise the data, and provide it to users. These definitions simplify the reality of many ecosystems. For instance, in some ecosystems all participating organisations may be both providers and users.

2 Rationales for Government intervention

As a starting point for this study, we have reviewed available evidence on the rationales for government intervention to increase access to data generated by the private and third sectors. This helps us identify why, in theory, the availability of data across the economy may be below a socially optimal level, and if and how government intervention could make a difference.

Building on the analysis in OFT (2009), NAO (2014) and other literature included in our evidence review, we place rationales for government intervention that are pertinent to the scope of this project into the following categories:

1. Setting the market framework, including the relevant legislation and regulations
2. Addressing market failures
3. Adjusting market outcomes

Government also provides a range of public goods and services. Because this project focuses on sharing data from organisations outside the public sector, we consider the potential benefits of using this data for the provision of public goods and services as a sub-group of category 2 above.

We also list a number of characteristics of data (non-rivalry, economies of scale, economies of scope) which interact with the rationales described below in providing opportunities for government intervention to increase access to data.

2.1 Setting the market framework

This category of rationales includes government's role in setting the conditions under which the use and sharing of data take place. Broadly, this means setting up the necessary institutions for the operation of markets, including for example property rights and consumer and competition law frameworks. Three elements within this broad category are likely to be particularly relevant to data access:

- The intellectual property rights system.
- The legal and regulatory framework around the use and stewardship of data and digital technology.
- Industry-specific regulations.

Across these three categories, providing certainty wherever possible on the relevant legal and regulatory framework is likely to have a positive effect on investment in data and on data access and sharing.⁷

2.2 Addressing market failures

Market failures prevent markets from working efficiently in delivering goods and services that are demanded by consumers. When such failures occur, markets provide too much or too little of a good or service compared to what would be

⁷ The impact of regulatory uncertainty on investment decisions has been explored, for example, in the context of investment in renewable energy, e.g. Fabrizio (2013).

socially optimal. When it comes to data sharing, the causes of market failure include:

- Positive externalities.
- Negative externalities.
- Imperfect information.
- Heuristics and biases.
- Market power.

There are positive externalities from the use of data

The use of data can produce positive externalities. These are benefits for organisations and individuals beyond those that have generated and/or used the data. For instance, in healthcare, using data may improve the quality of diagnoses or treatments for all, including patients who did not provide data themselves.⁸

Where there are positive externalities, access to data is likely to be insufficient. The gains to society from increased access would be greater than the benefits to individual organisations in the data ecosystem – crucially to data holders. Government could therefore intervene in these instances to expand access to data and create additional benefits for society.

There are negative externalities from the use of data

The use of data can also generate negative externalities – harmful effects on individuals or organisations that have not been involved in the collection and analysis of the data. An example reported recently concerns the Strava fitness app. Strava enables users to share map depictions of their physical workouts with other users online. Because the map shows large, concentrated clusters of users in areas of dense activity, it allowed the detection of secret US military sites around the world.⁹ In this case, data being shared by individual runners had a potential harmful effect on public security.

Where there are negative externalities, the cause is likely to be either excessive collection of and/or access to data, or inadequate management of the risks of data sharing. Therefore, negative externalities create a rationale for government to intervene to restrict access to data and data collection, or to ensure the adoption of appropriate risk management practices. These interventions may increase the trustworthiness of data use, which may in turn enable additional data sharing. Interventions to mandate data sharing, described in following sections of this report, may reduce incentives to accumulate data and thus mitigate the associated negative externalities.

There is often imperfect information around data

Limited information can lead to less data sharing than might otherwise be optimal.

¹⁰ These “information failures” can happen as a result of:

⁸ Coyle et al. (2020)

⁹ BBC (2018)

¹⁰ London Economics (2019)

- Uncertainty over the availability of data and the costs and benefits of collecting, analysing, sharing and accessing data.
- Limited information on the most appropriate technical, legal and governance approaches to sharing data.
- Uncertainty about how the data shared will be used by those who access it. This includes the risk that data will be used in a way that damages or does not benefit the data provider (“incomplete contracts”).¹¹

Limited information provides a rationale for government to intervene through policies that would generate additional knowledge on data sharing, allow a better flow of information and/or help align incentives between actors in data ecosystems.

Heuristics and biases

Decisions about data sharing and access may be affected by heuristics and biases that lead to under- or over-sharing relative to fully informed, rational choices.¹² For example, where sharing or accessing data involves immediate costs but the benefits are expected to arise in the future, people making decisions about sharing may underestimate their future costs or benefits (“hyperbolic discounting”). Decisions may also depend on how choices have been framed, or they may be excessively influenced by recent events (“recency bias”). These biases have been discussed in the literature primarily in relation to consumer decisions about sharing their data, but they could also apply to choices made by individuals within data-providing or data-using organisations.¹³

Market power in data-intensive industries

In some markets, it is more efficient for only one or a few firms to produce a good or service. In the case of digital markets, market power may result from:

- Network effects¹⁴, which make services (in particular, online platforms) more valuable to users as the number of participants increases.
- Economies of scale and scope in the use of data, which make larger companies more efficient than smaller ones.

The presence of concentrated market power does not necessarily lead to poor outcomes for consumers, particularly if new competitors may enter the market in the future. However, where there is limited competition, government may regulate the market to either increase competition (for example, mandating access to resources necessary to provide services, as in the case of the telecommunications network), or to directly improve outcomes for consumers (in the case of utilities, for example, prices are regulated). The Competition and Markets Authority’s (CMA)

¹¹ Coyle et al. (2020)

¹² Heuristics are mental shortcuts that allow individuals to make decisions quickly. Heuristics are highly economical and usually effective, but they can lead to biases (systematic and predictable errors). For example, people may assess the probability of an event by the ease with which such an incident can be brought to mind. This “availability” heuristic may lead to overestimating the probability of recent events (Tversky & Kahneman, 1974).

¹³ Acquisti & Grossklags (2007)

¹⁴ Network effects occur when a good or service becomes more valuable as more people use it. For example, if only one person uses a social media platform, it is of no value. As more people sign up, the more valuable it becomes. This can mean that it is difficult for a new platform to get established, as early users will derive little value from it.

Online Platforms and Digital Advertising Report (2020), following on from the report of the Digital Competition Expert Panel (2019), has recommended that a new Digital Markets Unit could enforce a new regulatory regime to promote competition in digital markets, potentially including powers to increase interoperability and access to data.

Economies of scale and network effects may also mean that markets for data have only a few providers. For example, data on user location generated by mobile networks is available at scale only to a small number of telecommunications operators. This is one of the reasons why the market for mobile network data is likely to be “thin”, i.e. there are relatively few suppliers and buyers. Data sharing in a thin market can be limited, because it is difficult for suppliers to set prices if they cannot benchmark against similar service offerings, and because potential users lack information about a suitable price to pay.

Enhancing choice and competition across other industries

Expanding access to data could also provide opportunities to increase choice and competition in markets beyond those in which network effects and economies of scale lead to market concentration. Sharing data on consumer use of a particular service, for example payment accounts (Open Banking), energy usage (midata), and other services (Open Communications, Open Finance) can help consumers compare different offers and switch between suppliers more easily.¹⁵ This can be facilitated by firms that analyse the data shared by existing providers in the market to offer comparison and other analytics services. For instance, third-party providers are allowed to access current account data shared by the nine largest banks in the UK, following the CMA’s Retail Banking Market Investigation Order 2017.

2.3 Adjusting market outcomes

Fairness of market outcomes

Government can intervene to improve market outcomes for specific groups, e.g. vulnerable consumers. The National Audit Office (2017) has reported that initiatives to share information about vulnerable consumers could help companies in regulated industries to assist these customers. Government and regulators could incentivise and back such efforts.¹⁶ For example, Ofwat has challenged companies in the water and sewerage industry to use data more collaboratively to identify vulnerable consumers.¹⁷ More generally, government can and does request data from businesses to ensure transparency and support law enforcement.¹⁸

Moreover, the use of data to help make automated decisions is increasingly common. These decisions can affect individuals (e.g. credit or healthcare), organisations (e.g. placement) and society at large (e.g. through the provision of public services). It is therefore important to make sure that these automated

¹⁵ These initiatives may also have benefits beyond an increase in competition, for example by enabling the creation of services that help account management or simplify everyday tasks for consumers.

¹⁶ National Audit Office (2017)

¹⁷ BEIS (2018)

¹⁸ For example, under the Equality Act 2010 (Regulations 2017), all large organisations must report specific figures about their gender pay gap.

decisions are reached using algorithms which are fairly designed. To achieve fair market outcomes, government may wish to intervene to ensure appropriate scrutiny is possible, including through greater data sharing.¹⁹

Industrial policy

Industrial policy objectives attempt to improve the business environment for specific sectors or technologies that are considered deserving of targeted support. Ensuring that organisations operating in these sectors have access to data that would foster growth and innovation could be a useful form of support. This may be the case in particular where the industrial policy objective focuses on data-intensive activities, as for example with the Artificial Intelligence grand challenge set out in the government's Industrial Strategy White Paper.²⁰

2.4 Selected characteristics of data

Data is non-rival

Data is non-rival. This means that existing data can be used by a number of people simultaneously without reducing the 'amount' of data available to anyone else. This means that in principle there could be benefits from sharing data very widely, because each additional use would not prevent other uses or deplete the data. Conversely, it may be necessary to invest in excluding others from using the data in order to protect commercial advantage, intellectual property or privacy.

Academic literature points to both the benefits of non-rivalry and the underutilisation of non-rival data. Jones and Tonetti (2019) develop a growth model with data as input to a firm's production. Unlike the other inputs (labour, land etc), data is non-rival. The implication of their modelling is that it is desirable for data to be widely shared, and that the welfare costs arising from limits to using non-rival data can be large.²¹

The use of data and economies of scale

There are economies of scale in data. For example, when two datasets capture the same or similar information, more can be gained when different datasets are merged than when each is used individually. Relatively low sample sizes may not allow useful analysis to be performed, for instance due to a lack of statistical power. And smaller datasets may provide an incomplete picture of the underlying reality, for example if they include a non-representative sample of a broad population. Larger scale can increase the robustness of analysis, including for example prediction accuracy.²² Economies of scale may also result from high initial fixed costs (e.g. to set up collection and access to data) and low ongoing costs of using the data.

¹⁹ Rovatsos M., Mittelstadt, B. & Koene, A. (2020).

²⁰ HMG (2017)

²¹ Jones et al (2019)

²² European Commission (2020)

The presence of economies of scale has consequences for government intervention. Firstly, economies of scale are one of a number of factors that can lead to concentration in data-intensive markets, which provides a rationale for intervention. Conversely, in markets or ecosystems that involve many different actors, individual organisations may have access only to smaller datasets, and, as described in further detail in Section 3.2, transaction costs and other challenges may prevent aggregation of smaller datasets. Where this is the case, government could intervene by supporting data sharing so that economies of scale can be achieved.

There are often economies of scope in the use of data

In addition to economies of scale, data also exhibits economies of scope. This means that when datasets are complementary, more insights and economic value can be extracted from merging them than by keeping them in separate data silos.²³ Targeted advertising provides an example of economies of scope in practice. An advertising service is likely to generate more click-through rates and revenue when it has access to information about consumers gleaned from a variety of sources, such as web-browsing, financial transaction, mobility and social media data. Economies of scope may also help companies that have access to valuable data to expand into other markets.

As in the case of economies of scale, additional economic benefits may be stimulated by integrating different datasets held by different organisations. Government could help remove barriers to this integration.

²³ Ibid.

3 Evidence of barriers to access to data

3.1 The data sharing journey

Providing or achieving access to data requires a number of decisions to be made. The exact type of decisions and the order in which they are made will vary case by case, but it is possible to identify a number of steps that one would expect to see in most “data sharing journeys”. Adapting information and guidance from a range of sources (including the ODI’s Data Toolkit for Business²⁴, the Royal Academy of Engineering’s data sharing checklist²⁵ and the European Commission’s Staff Working Document giving guidance on sharing private-sector data²⁶), we have identified the following common steps for data providers:

1. Identify data that could be shared, and any associated skills required
2. Engage with potential data users
3. Share details about data with potential users
4. Identify and/or develop a business model to generate benefits from sharing
5. Identify the costs and risks from sharing
6. Define the specifics of the sharing model
7. Develop the legal and technical infrastructure
8. Develop governance arrangements (if sharing takes place through ongoing, complex relationships, e.g. data sharing platforms with many participants, rather than through bilateral, one-off exchanges)

At the fourth step above, the key choices to be made include the following:

- How “open” the access provided could be, on a scale from named access to fully open access.
- How providing access would give rise to benefits, and their likely size. Benefits could include compensation in monetary terms, accessing services in exchange for the data, or outcomes that are useful to the provider in other ways. For example, the provider may want to share data that helps address a common sectoral challenge (e.g. safety incidents in the maritime sector, explored in Section 5.4). Another type of indirect benefit may arise when allowing access to data helps the development of products and services that are complementary to those of the data provider. For example, an online platform may want to share data about its users with developers/third parties who provide services through the platform, thereby increasing the value of the platform.

In practice, many data sharing journeys may not be linear. For example, identifying the risk of the chosen business model at step 4 may require going back to step 3 to redefine that model before going any further.

Users would also follow a data access journey with broadly similar steps:

²⁴ ODI – *Data Toolkit* (2020)

²⁵ Royal Academy of Engineering (2019)

²⁶ European Commission (2018).

1. Identify a challenge or opportunity that may be addressed through the use of data.
2. Identify the data needed, how it might be used, and what skills would be required to use the data.
3. Find data providers.
4. Ascertain whether the data is fit for purpose.
5. Identify and/or develop a business model for the data sharing relationship, and ascertain the costs of accessing the data.
6. Identify risks from accessing the data and ways to mitigate them.
7. Where possible, work with the data provider to:
 - a. Develop the specifics of the data sharing model.
 - b. Develop the legal and technical infrastructure required.
 - c. Develop governance arrangements, if needed.

3.2 Issues that may prevent optimal data sharing

Issues that may limit data sharing and access may occur at any point of the data sharing journey described in the previous section. After a review of the existing evidence and discussions with DCMS, we identified six common issues related to data access:

- **Lack of incentive to share data:** data providers may not be sufficiently incentivised to share or grant access to their data; for example, sharing may incur costs that they are unable to recoup from the beneficiaries of expanded access.
- **Lack of knowledge:** data providers may lack enough knowledge of the potential uses of their data, while data users lack enough knowledge of what data could be made available. Both parties may have limited understanding of how data can be shared even where a potential use has been identified.
- **Commercial, reputational and ethical risks:** the perceived or actual risk of losing competitive advantage, of suffering reputational damage from data uses that may breach others' trust, or from data sharing exposing controversial business practices, may deter providers from sharing data.
- **Regulatory and legal risks:** the perceived or actual risk of breaching data protection, intellectual property rights or regulatory requirements may also be a deterrent to sharing.
- **Costs of data access/sharing:** costs may be prohibitive because of a lack of coordination, common foundations, infrastructure and technologies that are needed for data sharing to be cost effective.
- Moreover, there are cases where sharing data may be particularly likely to lead to economic and social gains beyond the benefits that accrue to the organisations providing and obtaining access to data. As discussed in Section 2, there are **missed opportunities to use data in the public interest** to improve transparency, tackle significant social problems, promote competition, and support policymaking and regulatory enforcement.

In the remainder of this section we explore each of these problems and the underlying rationales that may warrant government intervention to tackle them or seize the related opportunities.

Figure 9 Types of data-sharing issues




Source: Frontier Economics

3.2.1 Lack of incentive to share data

There are three data sharing issues associated with a lack of incentive to share data. These are summarised and discussed in turn below.

Figure 10 Data sharing issues associated with a lack of incentive to share



Data sharing issues	Rationales for intervention
1. Data providers unwilling to make data available due to uncertainty about benefits (relative to costs)	Information asymmetries; option value of data; thin markets; setting market framework; regulatory uncertainty
2. Data providers unwilling to make data available because the benefits of investment to them do not outweigh the costs	Positive externalities; heuristics and biases; setting market framework

Source: Frontier Economics

1. Providers unwilling to make data available due to uncertainty about benefits

Data providers may be unsure about the nature and scale of benefits to them from enabling access to data. This could be caused by three factors:

- Data providers may have limited information on how many potential users of the data there are and what their valuation of the data would be ('information asymmetries'). This, coupled with high investment costs, deters them providing access to data.
- It may be difficult for data providers and users to forecast how data could be used and what benefits it could generate.²⁷ Data providers may see value in

²⁷ Data can have the characteristics of "experience goods", the quality of which cannot be fully determined before they are purchased (Nelson, 1970). Food, for example, cannot be fully evaluated before consumption.

waiting to see if more lucrative uses for the data become available before sharing their data (“option value” of data).²⁸

- There may be relatively few data providers and/or users. This creates a ‘thin market’ where the number of transactions is relatively small, and prices may be volatile or difficult to predict.
- The lack of information on the benefits of providing access to data may also be exacerbated by broader uncertainties over future government policies or regulations that affect the potential for, and returns to, sharing data.

2. Data providers unwilling to make data available because the benefits to them do not outweigh the costs

Even where data providers have sufficient information on the benefits to them from providing access to data, these benefits may still be outweighed by the costs. This could be caused by the following factors:

- There may be high upfront investment costs associated with data sharing and access, for example to set up systems and processes for sharing. These costs are incurred by the data provider, whereas at least some of the benefits may be captured by data users and wider society (‘positive externalities’).
- The benefits from using the data may materialise only in the future, for example because they require complementary investments to be made. Some organisations may place limited value on future benefits relative to present costs (“hyperbolic discounting”).

An extreme case of this may be where organisations are not incentivised to share data because there is little or no direct benefit to them. The benefits may have a wider public value, such as furthering knowledge or enabling innovation, but result in no commercial return to the provider (“positive externalities”). Organisations may be unaware of this public value or insufficiently motivated by it to incur the costs of sharing data.²⁹

²⁸ Coyle et al, (2020)

²⁹ The BEIS Smart Data Impact Assessment found that while the private sector in principle could develop effective standards to share data with third parties, in key markets these had failed to materialise. This is thought to be due to concentrated costs and dispersed benefits. That is, although customers and innovative organisations could benefit from data sharing, the implementation costs fall on incumbents who have little to benefit from sharing. BEIS – *Impact Assessment* (2020)

Uncertainty on benefits from sharing data on food waste

The ODI piloted a data trust model in 2019 to explore whether it could increase data sharing to support global food waste reduction efforts. The aim was to improve the ability of stakeholders to track and measure food waste within supply chains.

As part of the pilot, the ODI interviewed food retailers and manufacturers who were actively trying to both measure and reduce food waste. They found that the most powerful incentive for tracking and sharing relevant data is the reduction in costs. A representative of a waste-measuring software provider told the ODI that data sharing practices are starting to improve where organisations can make a clear business case for measuring food waste. (ODI - Food Waste 2019).

3.2.2 Lack of knowledge

We have identified five issues associated with a lack of knowledge around data sharing, outlined in Figure 11 below. The fourth and fifth issues are associated both with lack of knowledge and with risks of data sharing.

Figure 11 Data sharing issues associated with a lack of knowledge



Data sharing issues	Market failures
1. Organisations sharing / accessing data do not know where to find users / providers	Information failure; thin markets
2. Organisations sharing / accessing do not know whether the data they are collecting / considering buying are fit for purpose	Information failure
3. Organisations accessing data lack knowledge of the benefits of the data	Information failure; thin markets
4. Organisations perceive high risks in data sharing	Information failure; negative externalities
5. Organisations have or perceive a lack of trust in data sharing	Information failure; heuristics and biases

Source: Frontier Economics

1. Organisations sharing/accessing data do not know where to find users/providers (coordination)

Potential data providers may not know where to find potential data users, and vice versa, especially if there is no existing relationship between the two groups. This could be the case, for example, if the data held in one industry (e.g. data on mobility patterns collected by telecommunications companies) could benefit a different industry (e.g. marketing and advertising providers). Coordination problems also arise in situations where the data could be used for novel applications, thereby necessarily limiting the number of users. As a result, it can take a lot of time and effort for an organisation holding data to determine whether anybody might be interested in using a dataset; likewise, potential users may have to go to great lengths to find out whether datasets exist that meet their needs.

2. Organisations sharing/accessing do not know whether the data they are collecting/considering buying is fit for purpose

Even if data users are able to identify datasets that potentially meet their needs, they may hold insufficient information before purchase to determine whether the data is fit for purpose ("information failure"). This could be due to a number of factors:

- There may be limited information that can be gained from accessing "test" versions of datasets. For example, it may not be possible to identify data quality problems with a dataset until it has been integrated with other sources of information.
- Prospective buyers may not have the time/resources to assess in detail whether the data is what they need.
- Would-be buyers may not trust sellers to provide suitable data, particularly if they have had no prior dealings with them.

3. Organisations accessing data lack knowledge of the benefits of the data

Potential users of data may not know enough about its benefits for at least two reasons:

- Organisations may lack the skills and knowledge to fully appreciate the benefits of data availability or the value of the data to them ('information failure'). This problem may be especially pronounced for new or innovative types of data and data uses.
- It may not be possible to assess the benefits of data before trying to use it for a particular purpose.
- If an economy's data culture is underdeveloped, few organisations will seek to share data and so there are relatively few potential buyers ('thin market').

4. Organisations perceive high risks in data sharing

Organisations may overestimate or underestimate the risks from information sharing (see Section 3.2.3 below). This may be due to poor understanding of the nature of these risks or how to mitigate them ("information failure"), or to behavioural barriers (in particular a high or low aversion to risk).

Underestimating the risks from data sharing has the potential to incur costs for wider society when those risks materialise, for example as a result of data privacy breaches ("negative externalities").

Some organisations may also be reluctant to share data that would expose their business practices to unwanted scrutiny, even where this does not involve divulging commercially sensitive information.

5. Organisations have or perceive a lack of trust in data sharing

There may also be misperceptions about the trustworthiness of data sharing in circumstances where that sharing would have been beneficial ("information failure"; heuristics and biases). For example, concerns over fraud or data theft may

undermine trust in legitimate data sharing practices.³⁰ Low trust may stem not only from concerns about potential illegal activities: only a minority of the general population trusts banks, utility providers, social media organisations, central and local government organisations to follow ethical practices in collecting personal data.³¹

Examples of Challenges in finding relevant data

Healthcare, Extractives

The ODI recently conducted a project to explore the options for making data available, *Increasing access to data while retaining trust: why is it important?* (ODI, 2018b). Part of the project involved interviewing businesses across a range of industries to pin down their main challenges in data sharing, with the aim of identifying sectors and problems which the ODI could focus on for maximum impact. It discovered that the healthcare and extractive industries found it difficult to source relevant data. This is particularly true when data is published in unstructured and inconsistent formats in different places.

Smart cities

In partnership with Hitachi Consulting, Copenhagen has been attempting to implement a smart-city model with the goal of generating improvements and growth for the city by 2025. One aspect involves creating a marketplace for data, the *City Data Exchange*. To this end, Hitachi conducted interviews and workshops with organisations in the city and found that data seekers did not know what kinds of data are useful for them. While they have a clear understanding of the problem to be tackled and the information that is missing, they generally lack the knowledge about which raw data is best suited to their purposes. Given this, Hitachi suggested that the City Data Exchange should, among other things, identify the sort of data that can answer organisations' questions (IDC, 2016).

3.2.3 Commercial, reputational and ethical risks associated with data sharing

Commercial, reputational and ethical risks linked to data sharing include three types of problems, outlined below.

³⁰ The Transport Systems Catapult conducted surveys and interviews with stakeholders in the transport industry. A number of them said the risk of data being used maliciously was the main reason for not sharing data. They expressed concern that data for which they were responsible could be compromised and that they would be liable for the consequent loss of privacy, reduced security and reduced safety. Transport Systems Catapult (2017)

³¹ ODI – *Building Trust* (2019)

Figure 12 Commercial, reputational and ethical risks associated with data sharing



Data-sharing issues	Market failures
Private companies may retain data because they are worried about losing a potential source of competitive advantage	Public good; market power
Organisations perceive high risks in data sharing	Information failures; heuristics and biases
Organisations have or perceive a lack of trust in data sharing	Information failures; heuristics and biases
Sharing and accessing data may lead to a loss of trust and to ethically questionable uses of data	Negative externalities from data use; trust as a public good

Source: *Frontier Economics*

1. Private companies may retain data for commercial advantage - with a detrimental effect on public goods, consumer welfare and competition

Some data holders may be deterred from providing data access due to a risk of losing a commercial edge as a result. This can occur where data is in the hands of a small number of potential providers that are in a position to retain the data for their own advantage (“market power”). This could lead to worse outcomes for consumers (e.g. higher prices) or to less competition and choice in the market.

The under-provision of data sharing may be especially prevalent in cases where, once shared, the data has the characteristics of a public good: it is non-rivalrous (it can be used at the same time by different users, unlike most goods provided by the market) and non-excludable (it is difficult to exclude specific users from accessing the data). This can lead to the data provider being unable to capture a payment or return on the investment from data users because they cannot prevent access for free. Therefore, they have no incentive to incur the costs associated with providing the data.

2. Sharing and using data may lead to a loss of trust

The risks associated with data sharing may erode trust that undermines future data sharing. For example, where instances of risks materialising have occurred this may negatively impact trust in the ability of other data providers and users to manage data appropriately in future. These instances may include security breaches but also data uses that may be seen as ethically questionable (for example, to monitor employees in ways that are legal but may not be acceptable to employees and to wider society).

Example of perceived risks of losing competitive advantage II construction

The Committee for European Construction Equipment (CECE) recently looked at the digitisation of the building sector. One area explored was data sharing and the value of data for construction. Via surveys and interviews, the CECE found that companies are often reluctant to share data for fear of handing their know-how to rivals. Some players (especially medium-sized firms) believe that IT companies could erode their market share if data is not adequately protected. (CECE, 2019).

3.2.4 Legal and regulatory risks associated with data sharing



Figure 13 Legal and regulatory risks linked to data sharing

Data sharing issues	Market failures
Organisations concerned that data sharing may lead to harm / breach of the law, regulations or intellectual property rights	Information failure; negative externalities

Source: *Frontier Economics*

1. Concerns that data sharing may lead to harm / breach of the law or regulations

Data providers may be uneasy about legal and regulatory risks associated with data sharing. This may arise as a result of legal and regulatory barriers; a lack of certainty on the legal and regulatory framework; or insufficient knowledge of the risks involved ('information failure'). Where material, these risks could lead to costs to wider society ('negative externalities'). Data providers may be concerned for a number of reasons:

- Sharing data could be a breach of competition law, for example if sharing data with another organisation risks being perceived as collusion to gain an unfair competitive advantage.
- There could be costly reputational damage if they (or users) are seen to breach privacy regulations by sharing data, with uncertainty as to who would be liable for any breaches.
- Data sharing may reveal a lack of compliance with other laws and regulations (e.g. Equalities law).
- There may be security risks from data sharing.

3.2.5 Costs of data access/sharing



Figure 14 Costs of data access/sharing

Data-sharing issues	Market failures
There is a lack of coordination in investment in data foundations	Positive externalities; market power
Organisations do not invest sufficiently in data foundations that would increase efficiency and lower costs (e.g. standards)	Positive externalities; information failure
Organisations lack access to technologies / infrastructure that support data sharing	Information failure; positive externalities; market power; thin markets

Source: *Frontier Economics*

1. There is a lack of coordination in investment in data foundations

Duplication of efforts to support better data foundations and governance may lead to higher costs that hamper the provision of access to data. This is for the following reasons:

- High fixed costs may prevent smaller/younger businesses from participating in data foundations and governance arrangements. This could lead to smaller firms being excluded from the benefits of such arrangements (“equity” failures).
- The benefits of coordination to individual organisations may be too small for them to feel the required effort is justified, even though such coordination would be valuable for the wider market (“positive externalities”).
- Organisations with the resources and scale to develop foundational resources may not be incentivised to share these if doing so risks eroding their commercial advantage (“market power”).

2. Organisations do not invest sufficiently in data foundations that would increase efficiency and lower costs

Data providers may balk at the investment required to make their data interoperable, or to maintain data quality, so that it would be of benefit to potential users. This could result from the following:

- There may be uncertainty over the benefits from investing in data foundations because the implications of maintaining data quality are unclear to the data provider (“information failures”).
- Investing in higher-quality data may lead to benefits for other organisations, i.e. the savings in analysis costs that data users make may be greater than the return to the organisation making the investment (“positive externalities”). This can lead to under-provision if there is no mechanism for the data provider to capture a sufficient return on the investment.³²

³² The Data Sharing in Infrastructure Case Study conducted by Deloitte found that stakeholders in the industry said they experienced several technical issues when sharing data or using shared data, including inadequate data formats and standards. Deloitte found this to be due, in part, to a lack of commercial incentive to make necessary investments to overcome technical glitches (2017)

3. Organisations lack access to technologies / infrastructure that support data sharing

Organisations may lack access to the technologies or infrastructure needed to support data sharing for the following reasons:

- Some may not invest in data infrastructure because the costs are prohibitive, they do not think the outlay is worth it or they do not have the technical expertise (“information failures”).
- There may be a lack of financing for providers of data sharing support because financiers do not fully understand the demand for such services and the returns they can generate (“information failures”, “thin markets”).
- Some larger organisations may invest in technologies and infrastructure but not have an incentive to open them up to other organisations because the technologies give them a commercial advantage. There is also a risk that when technologies are opened up to others, this may lead to other organisations being locked into the use of a proprietary technology (“market power”).
- The benefits from investment in technologies and infrastructure that support data sharing may be greater for the wider market (e.g. other data providers and users able to use the infrastructure themselves) than the commercial return to organisations making the investment (“positive externalities”). This can lead to underinvestment, as those developing technologies and infrastructure may not take these wider benefits into account when deciding whether, or how much, to invest.

Example of inefficiency and technical barriers to data sharing in smart devices

A team of researchers at University College London conducted a study on the right to data portability in the emerging Internet of Things (IoT). The research identified several inefficiency and technical barriers to data sharing. (Turner et al. (2020).

The right to data portability, as outlined in the European Union's General Data Protection Regulation (GDPR), allows individuals to obtain and reuse their personal data for their own purposes across different services. (Right to data portability).

The UCL project involved an experimental study with four ubiquitous consumer IoT devices: two wearable fitness trackers (operated by Garmin and Fitbit) and two home assistants (operated by Google and Amazon). Researchers used one of the wearable fitness trackers and home assistants for three weeks and then attempted to port their data to the other ones.

All device operators enabled the researchers to download a copy of their own data in commonly used, machine-readable file types. However, the data was not always helpfully structured within the files, nor did it come with adequate guidance. There were also differences in how the data was saved and the level of complexity of the content.

Across all four tested IoT devices, direct transmission to another device was not possible. The format in which data is collected does not lend itself to being transmitted and thus was not reusable by IoT providers of similar services.

3.2.6 Missed opportunities to use data in the public interest

Figure 15 Missed opportunities to use data in the public interest

Data sharing issues	Market failures
There are missed opportunities to use data to enhance competition, support regulatory enforcement, improve policymaking and support research, as well as for other purposes (but there is no direct benefit for the data holder)	Public good; market power; positive externalities
Organisations are reluctant to share data that may expose them to unwanted scrutiny, even when in the public interest	Positive externalities; public good

Source: *Frontier Economics*

1. Missed opportunities to use data to enhance competition, support regulatory enforcement, and for other purposes

There are a few reasons why data providers may retain data that would be in the public interest to share:

- There may be a commercial advantage in retaining sole access to the data. This is most likely to have detrimental impacts on welfare in concentrated sectors where there are few organisations, or only one, that hold certain types of data (“market power”).
- There may be insufficient incentives for the organisation to undertake the effort or shoulder the costs required to share data, even though there is a public interest in doing so. This can occur where there is no reward mechanism for the provider because the data, once shared, becomes freely available (“public good”).
- Public interest datasets are by definition relevant for key sectors, research areas and policy issues. The use of these datasets is likely to lead to benefits to wider society by contributing to progress in these fields, for example better-informed medical interventions as a result of using health datasets (“positive externalities”).
- Research that draws on public interest datasets may help promote knowledge and innovation in the wider economy or develop new tools for using data. “Knowledge spillovers” from such innovations can then be applied to other datasets or in different sectors.

2. Organisations are reluctant to share data that may expose them to unwanted scrutiny, even when it is in the public interest

Organisations may be concerned about the risk of coming under public scrutiny as a result of sharing data they hold, even if doing so would be in the public interest.³³ For example, sharing data on an organisation’s activities might be of no benefit to the organisation itself – indeed, it could incur costs – but could be beneficial to wider society by providing transparency and holding decision-makers to account (“positive externalities”).

³³ The ODI’s Illegal Wildlife Pilot surveyed organisations involved in combating illegal wildlife trade. Interviewees said they hesitated over sharing data that, once analysed, might portray data holders in a negative light (by highlighting operating mistakes or poor data quality, for example). ODI – *Lessons from three pilots* (2020)

4 Determining the set of levers for Government intervention

4.1 Defining levers for government intervention

The purpose of this project is to help DCMS assess possible options to increase access to data across the economy. To this end, we needed to define a relatively small number of suitable options that could be compared with each other in the following phase of our work.

There are many ways of categorising government interventions to expand access to data. Possible starting points include using:

- The intended long-term outcomes that would result from addressing an issue (e.g. a framework aiming to increase productivity, one for better policymaking...).
- The rationales for intervention (e.g. a framework to address information asymmetry, one for uncertainty over the benefits of sharing, one for regulatory complexity/uncertainty...).
- The different styles of government action (leading and influencing; procuring...).³⁴
- Levers, i.e. how a set of interventions could help address a data sharing issue (e.g. raising awareness, fostering trust...).

For the purposes of this work we decided that levers were the appropriate basis on which to define potential frameworks for intervention.³⁵ We think of a lever as something that government could change, at least in principle, and that would have an impact on a problem or an opportunity related to data sharing. We agreed with DCMS to group possible options for intervention into the following six categories of lever:

- Improve knowledge / understanding of data sharing (“Improve knowledge lever”).
- Reduce costs of data sharing through better data foundations (“Reduce cost lever”).
- Support (new) ways to address the risks associated with data sharing (“Address risk lever”).
- Improve / demonstrate incentives for data sharing (“Improve/demonstrate incentives lever”).

³⁴ As defined in Policy Lab’s “Government as a System” toolkit, available at: <https://openpolicy.blog.gov.uk/2020/03/06/introducing-a-government-as-a-system-toolkit/>

³⁵ We ruled out defining frameworks based on long-term intervention outcomes because that would result in a handful of extremely broad frameworks, as solving most problems would ultimately lead to increased productivity and economic growth. We also ruled out using different styles of government action because each style could be used, in principle, to address any of the issues listed earlier in this note. Therefore, a style of government action in itself does not define a framework for intervention that describes the three components mentioned above in this section (what problem or opportunity is addressed; what could be achieved by addressing it; why government could intervene effectively). Rationales for intervention, as defined in Section 2, are relatively broad and so defining frameworks as groups of interventions justified by the same rationale would mean being too vague about the problems the intervention is trying to solve.

- Reduce the (perceived) regulatory burden associated with data sharing (“Reduce regulatory burden lever”).
- Mandate data sharing where this is in the public interest (“Mandate sharing lever”).

4.2 Mapping data-sharing issues to levers

4.2.1 Primary levers

The table below shows how each of the intervention levers maps to the types of data sharing problems outlined in Section 3. Each lever intends to address **primarily** a specific type of issue (indicated with a blue tick in the table), but many of them can also have a secondary effect on other types (indicated with a yellow tick). **Secondary effects** arise because our classification of data sharing issues into six types (the columns in Figure 16) involves some simplification. In practice, some issues straddle several types. For example, “organisations perceive a high risk from data sharing” has been put in the “risk” category, but it could also be considered a “knowledge” issue since it may arise from either an accurate or an inaccurate perception of risk. Matters that straddle several types can therefore be addressed by several levers. It should be noted that there are also secondary barriers that some interventions address. For example, ‘reducing (perceived) regulatory burden associated with data sharing’ primarily tackles legal and regulatory risks, but it may also reduce data sharing costs by lightening the burden of regulatory compliance. We explain our classification of secondary effects in the next part of this report, section 4.2.2.

Figure 16 Mapping of interventions to each type of data sharing issue

Lever	Types of data sharing issues					
	Lack of incentives to share data	Lack of knowledge	Commercial, ethical and reputational risks	Legal and regulatory risks	Costs of data access/sharing	Using data in the public interest
Improve knowledge	✓	✓	✓		✓	
Reduce cost	✓	✓			✓	
Address risks			✓	✓		✓
Improve / demonstrate incentives	✓	✓				✓
Reduce (perceived) regulatory burden				✓		
Mandate data sharing			✓			✓

Source: Frontier Economics

Beyond the primary and secondary effects described so far, the use of some of the levers may have **indirect** effects on data sharing. For example, improving incentives to share data may also spur investment in data foundations and therefore partly answer the “costs of data access” question. Moreover, improving discoverability as part of investing in data foundations may help improve knowledge and understanding of data. Such indirect effects should be considered when prioritising levers. We discuss this in Section 6.2.3.

4.2.2 Mapping data sharing issues to secondary levers

This section describes which data sharing issues can be addressed by deploying more than just the primary levers described above. This mapping will be used later in the report to provide an initial sizing of the economic activity that could be affected by each lever. In most cases, we find that there are at most two levers that can address directly each data sharing issue. Therefore, to keep the framework relatively simple and the sizing exercise manageable, we restrict our mapping to “secondary” levers.

Lack of knowledge

Figure 17 below shows the secondary levers pertinent for issues linked to a lack of knowledge about data sharing. For issues 1 and 2, the lack of knowledge relates to the availability of data and whether it is fit for the purposes of the organisation accessing it. Interventions which focus on Improving knowledge (primary lever) or improving data foundations and thus reducing costs (secondary lever) can address this question. For issue 3, however, the lack of knowledge relates to the benefits from using the data. Therefore, it can clearly be tackled not only by providing knowledge but also by demonstrating benefits (incentives) for data sharing.



Figure 17 Data-sharing issues associated with a lack of knowledge

Issue	Secondary lever
1. Organisations sharing / accessing data do not know where to find users / providers	Reduce costs
2. Organisations sharing / accessing do not know whether the data they are collecting / considering buying is fit for purpose	Reduce costs
3. Organisations accessing data lack knowledge of the benefits of the data	Improving/demonstrating incentives

Lack of incentives

Figure 18 below shows the secondary levers relevant to the two issues to do with a lack of incentives to share data. Both concern the balance between the benefits and the costs of data sharing. Appropriate interventions may therefore focus on improving the incentives (primary lever) or on reducing the cost of doing so (secondary lever).

Figure 18 Data sharing issues associated with a lack of incentives to share

Issue	Secondary lever
1. Data providers unwilling to make data available due to uncertainty about benefits (relative to costs)	Reduce costs
2. Data providers unwilling to make data available because the benefits of investment to them do not outweigh the costs	Reduce costs

Commercial, ethical and reputational risks

Figure 19 below shows the secondary levers pertinent for issues related to a lack of incentives to share data. Where companies are concerned about losing competitive advantage (issue 1 below), supporting ways to mitigate that risk is the primary lever; but where sharing the data may have clear social benefits, mandating data sharing may be an effective secondary lever.

Issues 2 and 3 are related to organisations' perception of risk and trust in data sharing. Where this perception is accurate, the appropriate lever for intervention is to address the risk (primary lever); however, this may be a *misperception* – in that case, interventions to improve knowledge and understanding of data sharing may be required. Finally, improving knowledge and understanding of data sharing may prevent losses of trust, by helping ensure that data is used appropriately.

Figure 19 Data sharing issues associated with commercial, ethical and reputational risks

Issue	Secondary lever
1. Private companies may retain data because they are worried about losing a potential source of competitive advantage	Mandate data sharing
2. Organisations perceive high risks in data sharing	Improve knowledge / understanding of data sharing
3. Organisations have or perceive a lack of trust in data sharing	Improve knowledge / understanding of data sharing
4. Sharing and using data may lead to a loss of trust	Improve knowledge / understanding of data sharing

Legal and regulatory risks

Figure 20 below refers to legal and regulatory risks. They could be addressed by interventions that reduce the (perceived) regulatory burden, including steps to foster knowledge and understanding of the legal and regulatory framework – as distinct from interventions that improve knowledge of data sharing in general. Supporting ways to address the relevant risks may be an effective secondary lever.

Figure 20 Data sharing issues associated with legal and regulatory risks

Issue	Secondary lever
1. Organisations concerned that data sharing may lead to harm / breach of the law, regulations or intellectual property rights	Support (new) ways to address the risks associated with data sharing

Costs of data access and sharing

Issue 1 concerns instances where current data-sharing practices are not as efficient as they could be. For example, there may be duplication of efforts to support data sharing. Where efficiency is thwarted by the high costs involved, reducing those costs is the appropriate lever. However, where a lack of knowledge is responsible, interventions to improve understanding of efficient data sharing techniques may be required.

Figure 21 Data-sharing issues associated with high costs of sharing data

Issue	Secondary lever
1. There is a lot of inefficiency / duplication of efforts to support better data foundations and governance (coordination efficiency)	Improve knowledge / understanding of data sharing
2. Organisations do not invest sufficiently in data foundational issues that would increase efficiency and lower costs	N/A
3. Organisations lack access to technologies / infrastructure that support data sharing	N/A

Using data in the public interest

Figure 22 below reports the secondary levers relevant for issues related to data sharing in the public interest. Issue 1 concerns organisations deriving no direct benefit from sharing their data. In some circumstances, such as when the goal is to increase competition, mandating data sharing may be required. However, in some cases it may be possible to incentivise data holders to share data for the public interest (e.g. to support research).

Issue 2 is related to organisations that are reluctant to share data in case it exposes them to unwanted scrutiny, even when data sharing is in the public interest. Government intervention to mandate data sharing is the appropriate lever to overcome this objection. However, where organisations are concerned that their data may be misused, the proper lever would be to support mitigation of the risks associated with data sharing.

Figure 22 Data sharing issues associated with data sharing in the public interest

Issue	Secondary lever
1. There may be opportunities to use data to enhance competition, aid regulatory enforcement and support research (but there is no direct benefit for data holder)	Improving/demonstrating incentives
2. Organisations are reluctant to share data that may expose them to unwanted scrutiny, even when in the public interest.	Support (new) ways to address the risks associated with data sharing

4.3 Logic models

Having categorised the levers for intervention and the data sharing issues they might address as either “primary” or “secondary”, the next step in our framework involves defining the levers in more detail in order to buttress our initial assessment of their potential effects (Sections 5 and 6 of this report).

To this end, we developed an initial *logic model* or *theory of change* for each lever. A logic model sets out the expected causal pathway (the ‘logic’) of an intervention. For each lever, the model shows the categories of activity that would fit that the lever along with their outputs, outcomes and ultimate impact on data sharing.

Figure 23 Components of the logic model

Source: Frontier Economics

We define the components of the logic model as follows:

- **Inputs** are the resources used to deliver the activities. We expect the types of inputs to be similar for each lever (e.g. financial, human resources);
- **Activities** are the actions taken, using the resources described in the inputs, to direct the course of change. We have kept the activities broad and used the Government as a System Toolkit language where appropriate;
- **Outputs** are the direct products of the activities. We have listed a few examples of what the outputs might look like under each category of activity, but our list is non-exhaustive;
- **Short-term outcomes** are the short-run effects of the activities. They can reach beyond immediate target audiences and can include changes in knowledge, attitudes and behaviours;
- **Long-term outcomes** describe the lasting effects of the outputs. These describe changes to the overall data economy;
- The purpose of each lever is to ultimately increase useful data access and sharing. This is the **impact** for each lever and logic model.

We present initial logic models for each lever in Figures 24-29. In Sections 5 and 6, we use the models to support our assessment of the levers. Developing a high-

level logic model for each lever helps us, later on, to understand which part of the data sharing journey the lever acts on as well as its possible indirect effects and unintended consequences.

Figure 24 Summary logic model: Improve knowledge and understanding of data sharing

ACTIVITIES & OUTPUTS	OUTCOMES Short-term	OUTCOMES Long-term
Convening Convening bodies established (e.g. user groups, providers and policymakers) and events conducted	Coordination of data users and providers	Increased knowledge of available data for users Increased knowledge of quality of available data
Developing data access and analysis tools Data access (e.g. data pools, portals) and analysis tools (e.g. data labs) developed	Increased access through data tools Increased analysis through data tools	Increased knowledge of how data will be used by data providers Data users know where to find data and how to request data (in line with FAIR principles)
Disseminating information e.g. published reports, technical/regulatory guidelines, showcasing uses	Engagement with disseminated information	Increased knowledge of data legislation
Delivering pilot/projects Data sharing projects/pilots delivered	Engagement with pilots/projects, e.g. data sharing from participants	Data providers learn more efficient ways to store and share data

Source: Frontier Economics

Figure 25 Summary logic model: Reduce costs of data sharing through better data foundations

ACTIVITIES & OUTPUTS	OUTCOMES Short-term	OUTCOMES Long-term
Supporting innovation in data sharing solutions Support provided to organisations, e.g. incubator programmes, innovation grants for security/privacy/IP, tax incentives, regulatory sandboxes	Solutions developed for more efficient data sharing	Lower legal costs of sharing and accessing data Lower technical costs of sharing and accessing data More data is identified by potential users due to increased findability Less duplication of fixed costs to access data (e.g. investment in interoperability, developing contractual terms for data sharing agreements)
Developing shared resources and providing advice Resources developed and disseminated, e.g. template data sharing agreements, other contract guidelines	Use of shared resources by data users and providers	
Incentivising investment in data foundations Reputational rewards (e.g. accreditations), financial rewards (e.g. payments, vouchers, tax incentives)	Engagement with rewards, e.g. accredited businesses, data shared	
Data standards Data standards developed, e.g. for interoperability, instructions for reuse	Uptake of standards	
Data intermediaries to oversee the data sharing process Data institutions developed, e.g. data trusts, commercial intermediaries	Improved coordination and trust in data ecosystem	

Source: Frontier Economics

Figure 26 Summary logic model: Address (perceived) risks of data sharing

ACTIVITIES & OUTPUTS	OUTCOMES Short-term	OUTCOMES Long-term
Supporting innovation in data sharing solutions Support provided to organisations, e.g. innovation grants for security/privacy/IP	Solutions for more secure data sharing solutions developed	Less uncertainty about commercial risks of data sharing
Developing shared resources and providing advice Resources developed and disseminated, e.g. template data sharing agreements, other contract guidelines	Use of shared resources by data users and providers	Increased knowledge of how data will be used by data providers Increased confidence in avoiding reputational damage
Developing and incentivising adoption of accreditation Accreditation schemes developed, e.g. to demonstrate adoption of data foundations and security	Engagement with accreditation schemes and improved trust in data ecosystem	Increased confidence in avoiding uses of data that may lead to ethical risk
Data intermediaries to oversee the data sharing process Data institutions developed, e.g. data trusts, commercial intermediaries	Improved coordination and trust in data ecosystem	Increased comfort with data legislation compliance

Source: Frontier Economics

Figure 27 Summary logic model: improve/demonstrate incentives for data sharing

ACTIVITIES & OUTPUTS	OUTCOMES Short-term	OUTCOMES Long-term
Testbeds and trials Funding or support to research projects that demonstrate the feasibility of data sharing and its benefits	Testbeds and trials demonstrate feasibility and benefits of data sharing	Less uncertainty about feasibility and benefits of data sharing Greater private benefits of sharing and accessing data (as a result of incentives) Increased knowledge of benefits of data sharing for data users Increased knowledge of benefits of data sharing for data providers Better understanding of potential value of data and new data use cases developed Appreciation of value of data is embedded in organisational culture
Research on data valuation and data sharing business models Funding or support to develop tools to value and monetise data sharing	New/better ways of valuing data; more evidence on value of data developed; use cases	
Financial incentives for data sharing e.g. tax credits for investment in data sharing, vouchers, support for academic-private sector collaboration	Greater investment in data sharing; organisations taking up vouchers and other forms of support	
Reputational incentives for data sharing E.g. kitemarks, accreditation schemes	Organisations take up accreditation schemes	
Disseminating information Information disseminated on data-sharing benefits and how to value data, e.g. reports and guidelines	Information is used by potential data providers, users, intermediaries	

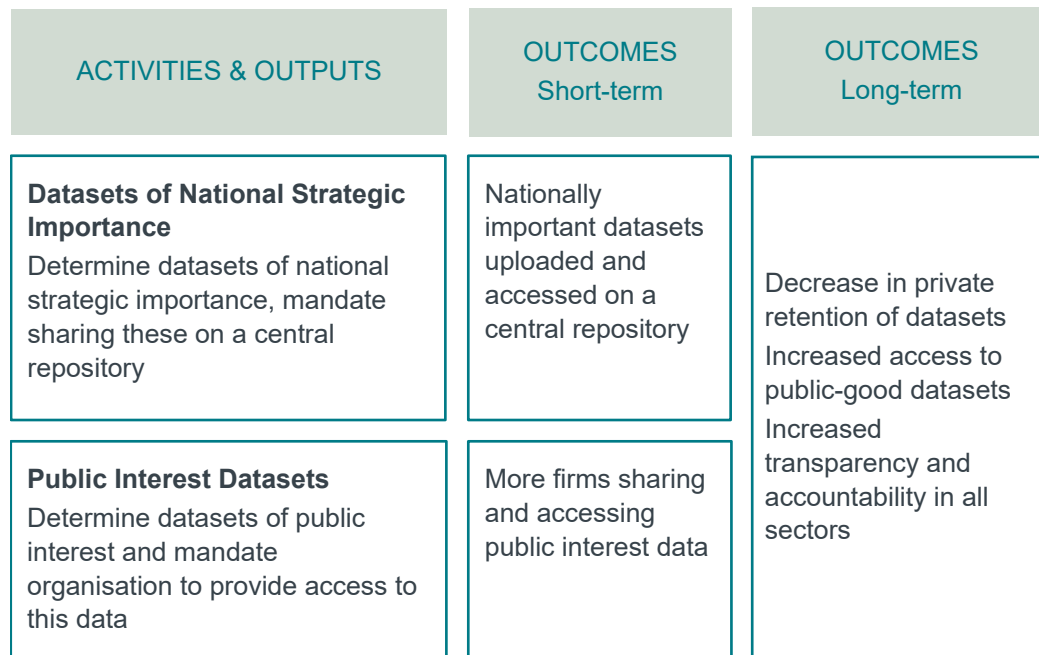
Source: Frontier Economics

Figure 28 Summary logic model: Reduce (perceived) regulatory burden associated with data sharing

ACTIVITIES & OUTPUTS	OUTCOMES Short-term	OUTCOMES Long-term
Supporting innovation in data-sharing solutions Support provided to businesses, e.g. incubator programmes, innovation grants for security/privacy/IP, tax incentives, regulatory/competition sandboxes	Better data sharing solutions developed	Lower legal costs of sharing data
Disseminating information Information disseminated, e.g. regulatory advice for data providers/users, self-assessment regulation tools, regulatory helpline, public communication campaigns on data regulation	Engagement with disseminated information by data users and providers	Lower regulatory costs of sharing data Increased knowledge of data legislation
Data intermediaries to oversee the data sharing process Data institutions developed, e.g. data trusts, commercial intermediaries	Data provision and access provided through intermediaries	Increased knowledge of liability rules Increased confidence in compliance with regulation
Revise legislation and regulation Updated legislation leading to (e.g.) increased transparency, better monitoring and enforcement	Greater compliance, familiarity and understanding of legislation	

Source: Frontier Economics

Figure 29 Summary logic model: Mandate data sharing where this is in the public interest



Source: Frontier Economics

5 Our approach to assess levers for intervention

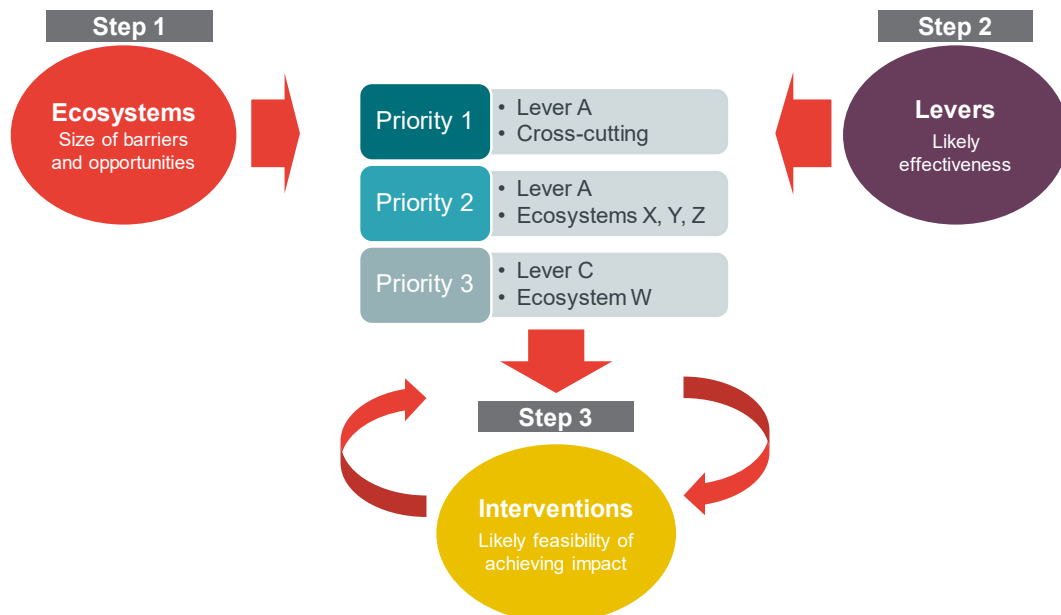
In Section 4, we described six levers government could use to increase access to data across the economy. In this section we set out a framework to support government in setting priorities. An initial application of the framework and some emerging findings are provided in the next section.

5.1 Overview of our approach

Government would like to select interventions with the greatest potential economic and social impact. In an ideal world, this would involve taking every possible intervention that could increase access to data in every possible ecosystem or industry and assessing them all on the basis of evidence to arrive at the best combination. As interventions in this policy sphere are in their infancy, evidence of their potential value is sparse and such an approach is not possible. The framework we set out in this chapter aims to provide a set of criteria that can be used to make an informed judgment about which interventions have the greatest potential. The framework draws on the available evidence wherever possible. However, in some areas there is no evidence, so the criteria also include theory-based assessments of the effectiveness of different interventions.

Our framework has three steps.

Figure 30 Overview of approach to assess the potential impact of each lever



The **first step** involves assessing the “addressable scope” for each lever, meaning how much economic activity in total might be positively affected, in principle, by using each lever. This is achieved by identifying the data-sharing ecosystems where there is evidence that the barriers each lever could address are present. It then estimates the size of these ecosystems.

In order to prioritise policy interventions, it is useful to judge how effective each lever would be in tackling the relevant data-sharing issues, relative to a counterfactual of no policy intervention. The **second step** involves a high-level assessment of the advantages and disadvantages of using each lever to achieve the objective of increasing access to data across the economy. This evaluation is undertaken lever by lever. However, the relative pros and cons of each lever may vary from one data-sharing ecosystem to another. Therefore, the **third step** in the framework involves developing criteria that could be used to assess whether levers are likely to be effective in a specific ecosystem.

The output of step 3 should be a set of feasible policy interventions that address existing data-sharing issues; can affect positively a lot of economic activity; and are likely to be effective given their general characteristics and the specific features of the key ecosystems in which they would be applied.

Reaching the end of step 3 may not be a linear process. The prioritised interventions that result from steps 1 and 2 may not bear the additional scrutiny of step 3; if so, they may ultimately be discarded and an option further down the initial list considered. This process will be iterative, until there is little to be gained from additional analysis.

Each step is described in detail in the following sections.

5.2 Market sizing

The levers considered in this report are defined as ways of addressing different types of issues related to data sharing. Therefore, we can assess how much economic activity might be positively affected by each lever (the size of the prize), by:

- Mapping each of the possible data-sharing issues (as defined in Section 3) to the levers that may tackle the issue.
- Mapping data-sharing issues to the different areas of the economy (industries and/or cross-industry ecosystems) where these arise.
- Measuring the value of economic activity affected by the issues.

5.2.1 Mapping data-sharing issues to ecosystems

To be able to estimate the economic value potentially associated with each of our levers, we need to understand what data-sharing issues are most prevalent across the economy and in which ecosystems. This mapping can then be used to identify the ecosystems where each lever would have maximum impact.

Ideally, the mapping exercise would draw on comprehensive evidence of the prevalence and severity of these issues in different ecosystems. This would allow the development of an evidence-based view of the ecosystems for which each lever is most relevant. To the best of our knowledge, there is no such source of data. For the purposes of this work we have therefore relied on the existing evidence base, including a number of studies of the data-sharing issues in specific industries. This evidence allows us to form an initial view of where data-sharing issues exist in the economy. We set out the results of our preliminary mapping exercise in section 6.1.

It is possible that the levers could impact additional markets/data ecosystems that are not included in our analysis because no specific research on the barriers to data sharing in those areas has yet been published. In the future, it would be possible to fill the gaps in the evidence base through primary research, including a survey of organisations in the private and third sectors and/or further stakeholder engagement in the industries for which there is currently limited evidence.

5.2.2 Measuring the total value of relevant industries by lever

Once we have mapped levers to ecosystems, the next step is to estimate how much economic activity in each industry is likely to be affected by the pertinent data-sharing issue in order to obtain an indication of the size of the prize for each lever.

In principle, one would want to assess the benefits from data sharing that could be unlocked by each lever. These may include not only increases in economic output and social benefits, but also greater resilience.³⁶ In practice, limitations in the evidence base significantly constrain our ability to measure how much data sharing is taking place and how valuable it is.

Given these limitations, we generally use industries as a proxy for data ecosystems and two metrics as proxies for the potential value of data sharing in each industry:

- the total Gross Value Added of industries relevant for each lever.
- the value of the existing “data market” in each industry as measured by a European Commission study.

We present the results of this sizing exercise in Section 6.1.

In the future, this analysis could be enhanced through a more precise definition of data ecosystems and the activities they encompass. This could involve:

- A more precise assessment of the size of the data ecosystem:
 - Measuring the activities of “core” data providers in the ecosystem, that is organisations for which the supply of data is a core part of their product or service offering. The metrics used could include revenues and/or funding raised by “core” organisations; and
 - Assessing the extent to which providing data is a significant part of the activities of other (non-core) organisations. This evaluation could be based on qualitative interviews or on surveys asking organisations the extent to which they engage in data sharing. This may allow part of the revenues or investment of non-core organisations to be included in the value of the data ecosystem.
- A more precise assessment of the importance of data-sharing issues in the ecosystem and of the activities that may be unlocked by addressing them. Existing evidence mainly describes in very general terms the types of data-sharing problems that arise in an industry. Stakeholder engagement and/or primary data collection could strengthen the evidence base by investigating in more detail how many organisations are affected by each issue and what difference it would make if it was resolved or mitigated.

³⁶ As, for example, in the case of data sharing in smart cities (McKinsey, 2019)

5.3 High-level assessment of the levers

Having defined, at least in broad terms, the size of the potential gains, it would be useful to weigh up the advantages and disadvantages of each lever against the objective of increasing access to data across the economy. We report our proposed criteria to evaluate the levers in Figure 31 below and Figure 32 overleaf. The criteria in Figure 31 are applicable to both the economy as a whole and to specific ecosystems. The criteria in Figure 32, by contrast, are ecosystem-specific: any assessment is likely to vary by ecosystem and an overall economy-wide evaluation is unlikely to be useful.

5.3.1 General (cross-ecosystem) assessment

Figure 31 lists five criteria that can be used to assess the likely advantages and disadvantages of different levers in addressing the pertinent data sharing issues. These criteria can be applied in general, across data ecosystems. We provide our initial assessment in Section 6.2. In the future, it could be expanded with additional evidence and refined by using the logic models presented in Section 4.3 to apply the criteria to each intervention category under each lever.

Figure 31 General lever assessment criteria

Criterion	Description
Strength of evidence	<p>a. Strength of data policy-specific evidence: is there any evidence that interventions relevant to this lever have been effective at increasing access to data, relative to a counterfactual? How robust is this evidence?</p> <p>b. Strength of evidence in other areas: are there effective interventions under this lever in other policy areas (e.g. innovation policy, trade policy...)?</p>
Timing of impact	<p>a. What stage of a hypothetical data sharing journey does this lever act on? (early-middle-late)</p> <p>b. Are there existing interventions relevant under this lever that could be expanded/adapted?</p>
Indirect effects of lever	Is the lever likely to have an effect on barriers other than the one it intends to address? For example, does improving incentives for data sharing have an indirect effect on the risks from data sharing?
Unintended consequences and trade-offs	What trade-offs should be considered when applying this lever? What is the risk of adverse effects?
Cost of intervention	How much would it cost to deliver this intervention, considering both financial and in-kind inputs?

The first criterion relates to **the strength of the relevant evidence base**. This would suggest prioritising levers which incorporate interventions that have been proved to be effective at increasing access to data. However, because this is a novel policy area, there are two caveats to the application of this criterion:

- First, the most effective interventions may not have been tested yet. Therefore, this criterion should be traded off against other elements of our prioritisation framework.

- Second, it could be useful to consider examples of interventions from other policy areas when assessing the likely effectiveness of each lever.

The second criterion concerns the **timing of possible impacts**. All else equal (for example, assuming different levers are expected to be equally effective), government may choose to prioritise interventions that produce a quicker impact or choose a mix of interventions that yield results at different times. Swifter impacts may be expected:

- When an intervention acts on later stages of the data sharing journey, as described in Section 3.1. For example, an intervention that helps define the specifics of the technical sharing infrastructure could have a relatively early impact if data providers/users have already been identified and business models to generate and distribute the associated benefits have been agreed.
- When an intervention addresses problems that affect developing or developed data ecosystems, rather than nascent ecosystems.
- When an intervention can be deployed relatively quickly, typically because it is an adaptation of an existing intervention, e.g. one that has already been piloted.

Within this study, we consider the first bullet above, i.e. how each lever maps to different stages of the sharing ecosystem. Assessing which intervention is more appropriate for nascent, developing or developed ecosystems would require a mapping of ecosystems to the data sharing issues described in Section 3. This would require primary research beyond the scope of this study.

Some levers may have greater potential than others to generate **indirect effects**, addressing data sharing issues other than those they are intended to affect. The overall impact on the relevant data ecosystems could be amplified as a result, warranting prioritisation of the levers in question.

As in any policy intervention, using all levers is likely to entail **trade-offs** against other policy objectives and potential **unintended consequences**. These need to be considered, along with the likely **cost** of intervention, in choosing what levers could deliver greater value for money.

Ecosystem-specific assessment

Some of the criteria that could be useful to evaluate the advantages and disadvantages of different levers are more likely to require an ecosystem-specific assessment. We list these criteria in Figure 32. They include a “gap analysis”, i.e. taking an overall view of the data sharing barriers in an ecosystem, and checking how these barriers interact. We believe that this analysis should be ecosystem-specific because, across the entire economy, all of the issues considered in Section 3 could be relevant.

Figure 32 Ecosystem-specific lever assessment criteria

Criterion	Description
Existing ecosystem-specific initiatives	Are there existing private-, third- or public-sector initiatives to increase access to data in this ecosystem? Might there be a role for government to support any of these initiatives?
Gap analysis	Is the lever more likely to be effective if other levers are also used first or at the same time?
Complementary conditions	For example: What is the overall level of digitalisation in this ecosystem? Is there sufficient access to the skills required to use data effectively in this ecosystem?
Ecosystem characteristics	Do the characteristics of this ecosystem indicate that this lever is particularly likely or unlikely to be effective?

The next section of the report explores in more detail the role of ecosystem characteristics.

5.4 Assessing how lever effectiveness may vary in different ecosystems

This criterion involves judging whether the characteristics of an ecosystem mean that a lever is more or less likely to be effective. Ecosystems display huge variety, so in principle this could require an extensive exercise to identify all possible characteristics of an ecosystem and to consider their potential impact on the effectiveness of a lever. This is not feasible, so we have prepared a preliminary list of key ecosystem characteristics which, from the evidence we have reviewed to date, are most likely to determine the potential effectiveness of a lever.

- The number of organisations in the ecosystem and the relationships between them. In more complex ecosystems it may be harder to understand what data is available, how it can be accessed and how the benefits and costs of using data could be distributed. There may be a greater need for coordination across different organisations to integrate the data and make it fit for purpose. And if more organisations have access to the data there may be greater confidentiality risks that have to be managed.
- The nature of the data to be shared and of the collection mechanism. Data about businesses, people, the natural world or the built environment is likely to be put to different uses. There are also likely to be different challenges for data protection, intellectual property protection and regulatory compliance, as well as for the trustworthiness of data use and trust between actors in data ecosystems.
- The stage of development of the data ecosystem. Nurturing a new ecosystem is not the same as helping to expand or improve an existing one. Different levers may have to be pulled.

The next section uses six case studies of data ecosystems to explore the role of these characteristics in determining the effectiveness of levers for intervention. Future work may expand on these initial findings by investigating additional data ecosystems. This may entail considering other ecosystem characteristics,

including the following: the jurisdiction of the organisations and data concerned (confined to the UK or international); the nature of the data collection mechanism (data volunteered by individuals, generated through tracking, inferred from other sources); the visibility of the data (is it shared with named individuals, specific parts of an organisation, an entire organisation? Can it be reshared by the organisation accessing it?); the technical details of how the data is stored and shared.

5.4.1 Summary of findings from existing ecosystem case studies

We wanted to test how the effectiveness of levers may vary in different ecosystems with a view to developing the assessment criteria set out in Section 5.3. To that end, we selected six existing case studies of data sharing, listed in Figure 33 below. We also report for completeness the type of data shared and the stage of development of each ecosystem.

Figure 33 Summary of ecosystem case studies

Ecosystem	Relationship between organisations	Type of data shared	Stage of development
Energy smart meters: Sharing of data collected through gas and electricity smart meters with third-party organisations	One data provider (data communications company), many users	Consumer data	Developed
APROCON: A platform that aims to speed up the design of complex aircraft components, by allowing partners to access, use and share data in real time	Many providers and users	Business data	Developing
CityVerve: A pilot study for a “platform of platforms” allowing the linking of data between distinct local data sharing initiatives	Many providers and users	Mixed	Nascent
OneTRANSPORT: a data marketplace that aims to enable organisations to share and use data to support intelligent mobility solutions	Many providers and users	Mixed	Nascent

Ecosystem	Relationship between organisations	Type of data shared	Stage of development
HiLo Maritime Risk Management: HiLo is a joint initiative in the maritime industry aiming to improve risk modelling in this sector. Participating shipping companies share data on accidents and safety incidents with HiLo.	Many providers, one data user	Business data	Developed
The Weather Company: A commercial platform that combines weather data and other information to aid decision-making in sectors where weather has an impact (e.g. transport, oil and gas, agriculture)	Many providers, one data user	Business data	Developed

Source: Frontier Economics analysis of case studies in Royal Academy of Engineering (2019) and ODI (2020)

Figure 34 Summary of findings from existing case studies

Lever	Drivers of effectiveness
Improve /demonstrate incentives	This lever may be more likely to lead to additional data sharing where: <ul style="list-style-type: none"> the data is dispersed (there are many organisations potentially involved in the ecosystem); using the data requires integration to reach a sufficient scale; and the benefits from using the data are not immediately obvious/rely on innovation/take some time to materialise
Reduce costs of data sharing through better data foundations	This lever may be more likely to lead to additional data sharing where trusted relationships between data users and providers already exist or can be developed/nurtured. Vice versa, the additionality of this lever may be lower where data is being shared by many owners with one large-scale intermediary or user who can monetise the data
Improve /demonstrate incentives and Improve knowledge / understanding of data sharing	These levers can be used to support the development of data ecosystems where data-driven products have not yet been created and/or customer take-up of those services is likely to be relatively slow in the absence of intervention.
Addressing risk	Where trust between data users and providers has not yet been established, using this lever could help interventions that reduce the cost of data sharing to achieve the objective of increasing access to data.

Lever	Drivers of effectiveness
Reduce (perceived) regulatory burden, mandate data sharing	Limited evidence from existing ecosystem case studies. Could be explored in future case studies.

These findings are based on our interpretation of the six case studies. We apply economic theory to draw general lessons based on material in the ODI and Royal Academy of Engineering reports. Annex B provides further detail on each study, including a summary of the two organisations' findings.

The OneTRANSPORT case study suggests that **demonstrating benefits** from data sharing may be particularly helpful in the following circumstances: **when the data is dispersed (many organisations may be involved in the ecosystem)**; when using the data requires integration to reach a sufficient scale; and when the benefits from using the data are not immediately obvious/rely on innovation/take some time to materialise. In such a situation network effects could be triggered through demonstration. Demonstration could also be beneficial where these characteristics are not present, but in such cases there may be a less direct link between demonstration and increased data sharing. For example, the APROCONE project could demonstrate the benefits of sharing data for product design between supply chain partners in aerospace or other sectors. However, demonstration may not necessarily lead others to join the ecosystem, with the potential for creating a positive feedback loop, because its size may be fixed by existing supply chain relationships.

A comparison of the APROCONE, OneTRANSPORT and CityVerve cases suggests that **lowering the cost of data sharing** may be most effective when **trusted relationships between data users and providers** already exist or can be developed. Trust may enable providers and users to take advantage of the reduced costs more quickly and efficiently than would be the case if there were concerns about commercial, ethical and reputational risks. If there is trust, cutting the cost means sharing can be done more efficiently. It can also lead to increased sharing, bringing about in some cases a step change in the usefulness of the data. For example, the APROCONE project aimed to improve the timeliness of data being shared; in the event, real-time sharing was achieved. In APROCONE, a relationship already existed between at least some of the organisations in the data ecosystem; in the other two case studies, there were more participants and the relationships between them may not have been as firmly established.

Comparing ecosystems with many data users to those with just one, it appears that government **intervention to support data foundations** may be **needed less in situations where data is being shared by many owners with one large-scale intermediary or user** who can monetise the data. In the HiLo and The Weather Company studies, the data user can monetise the data and appears to have the incentive and the resources to make the data received from different providers interoperable. This may not be the case to the same extent where there are multiple data users, where the scale of data used is smaller, or where there isn't a clear, immediate commercial benefit to be gained from integrating the data. However, even in these circumstances, interventions that make it easier for data

owners to invest in interoperability may avoid duplication of efforts and improve the efficiency of data sharing.

In several cases (e.g. OneTRANSPORT, CityVerve and to a lesser extent smart meters), the ultimate development of a data ecosystem relies on the roll-out of data-driven services and how they go down with the public. This suggests that interventions that raise awareness of data-driven services plus, where appropriate, steps that incentivise their take-up, could be part of a mix of policies to support increased access to data under the “knowledge” or “incentives” levers.

6 Initial application of the framework

6.1 Initial market sizing

As described in Section 5.2, assessing how much economic activity in aggregate could be positively affected by each lever involves:

- Mapping each of the possible data sharing issues (as defined in Section 3) to the levers that may tackle the issue.
- Mapping data sharing issues (and therefore levers) to the various areas of the economy (industries and/or cross-industry ecosystems) where these arise.
- Measuring the value of economic activity affected by the issues.

6.1.1 Mapping levers to areas of the economy

For this analysis, we relied on a rapid review of available evidence on the prevalence of barriers in specific data ecosystems. Most of the literature provides evidence at the industry level, with the exception of reports on the geospatial data market and on smart city ecosystems. It is possible that the levers could impact markets/data ecosystems that are not included in our analysis because, as yet, no research has been published on data sharing barriers specific to those additional sectors. And owing to our focus on ecosystem-specific evidence, the review may not fully cover evidence on cross-cutting issues, such as barriers to data sharing related to difficulties in assessing compliance with the EU General Data Protection Regulation.

We were able to find evidence covering all industries defined under the Standard Industrial Classification of Economic Activities (SIC), with the exception of professional and scientific activities, education, arts and retail. We did not look for evidence on data sharing issues in the public sector as this is beyond the scope of our analysis. We were able to find evidence of issues in the geospatial, smart cities and home Internet of Things ecosystems as well as in consumer smart data (specific to finance, banking, pensions and communications). The evidence we have identified primarily concerns the private sector, with the partial exception of data sharing in health (which includes third-sector research organisations) and in smart cities (which includes local government). A full list of the evidence included in this assessment is shown in Annex 1).

Our findings by lever are summarised in Figure 35 below. Two ticks (✓✓) indicate that the market or data ecosystem is relevant for the primary lever, and one tick (✓) for the secondary lever. The “address risks”, “reduce costs” and to a lesser extent “improve incentives” levers are pertinent for most industries. “Improve incentives” and “Mandate sharing” are significantly more germane for their potential role as secondary levers than they are solely in their primary role. In both cases, we have found relatively limited industry-specific evidence of the issues that these levers are primarily intended to address. However, there is much more evidence of issues around the risks of data sharing, which these levers can address through secondary effects, as described in Section 4.2.2.

Figure 35 Industry or ecosystem relevant for each lever

	Improve knowledge / understanding of data sharing	Reduce costs of data sharing through better data foundations	Support (new) ways to address the risks associated with data sharing	Improve / demonstrate incentives for data sharing	Reduce (perceived) regulatory burden associated with data sharing	Mandate data sharing where this is in the public interest
Agriculture	✓ ✓	✓	✓ ✓	✓ ✓		
Automotive	✓	✓ ✓	✓ ✓	✓ ✓		✓
Banking / FS		✓ ✓	✓ ✓	✓		✓ ✓
Construction	✓		✓ ✓			✓
Energy		✓ ✓				
Extractives	✓ ✓	✓	✓ ✓			
Healthcare	✓ ✓	✓ ✓		✓		✓ ✓
Geospatial	✓ ✓	✓		✓ ✓		
Infrastructure	✓		✓ ✓			✓
Manufacturing (less food and automotive)	✓	✓ ✓	✓ ✓	✓ ✓		✓
Platforms			✓ ✓			✓
Produce (food manufacturing)	✓	✓	✓ ✓	✓ ✓		✓ ✓
Real estate	✓ ✓		✓ ✓			
Smart cities	✓ ✓	✓ ✓		✓		
Smart devices/IoT	✓	✓ ✓	✓ ✓			✓
Telecommunications		✓ ✓		✓		✓ ✓
Transport	✓	✓ ✓	✓ ✓		✓ ✓	

6.1.2 Sizing activity relevant for each lever

Having identified the relevant industries for each lever, the next step involves assessing the extent of economic activity in each of the industries that could be affected by each lever. Given the available data, we are able to consider two metrics to assess the relative pertinence of different levers: Gross Value Added generated in the industries concerned and the value of the data market in those sectors. This exercise focuses on data sharing in the private sector, as the available data does not capture third-sector activities.

Below is a brief description of our methodology along with the advantages and disadvantages of each metric:

- 1. Gross Value Added generated in industrial sectors relevant for each lever.** Information on the GVA for industries that have an assigned SIC is taken from the ONS's GVA per industry data. Information on industries or data types that do not have an assigned SIC code - including platforms, smart cities, home IoT and geospatial data – comes from published

research.³⁷ The GVA metric likely overestimates how much economic activity might be affected by each lever. Even if one used the most granular standard industry definition (four-digit industrial codes, e.g. “manufacture of footwear”), it is likely that a lot of the GVA that most sectors generate is attributable to activities beyond data sharing and access.

2. **The value of the data market in industries relevant for each lever.** Information on the value of the data market is drawn from the European Data Market monitoring tool.³⁸ The European data market is defined as “the marketplace where digital data is exchanged as “products” or “services” as a result of the elaboration of raw data”. The value of the market is calculated as the sum of the revenues generated by data suppliers for these products and services. We have estimated the UK’s share based on the size of each industry. This metric may underestimate how much economic activity is potentially affected by each lever. That is because it captures exchanges where there is a sale of products or services based on the elaboration of data. However, data sharing and access can take other forms. For example, a supplier and a customer can decide to share data on product development without a sale for the data being recorded. Moreover, the metric gives us specifically the value of existing, rather than potential, data sharing.

Figure 36 below shows results from our sizing exercise using both metrics. We also report both the value of industries for which the lever is primary (left-hand column under both methods) and the value of all industries for which the lever is relevant (both primary and secondary relevance, right-hand column).

The relative impact of each lever is broadly consistent across the two methods. When considering primary effects only, the “reduce cost” and “address risk” levers rank highest under both the data market value (DMV) and GVA methods. When the secondary effects are also taken into account, the value associated with the “incentives”, “knowledge” and “mandate data sharing” levers increases substantially under both methods. This reflects the fact that all three levers have secondary effects on at least one other data-sharing issue. For the “knowledge” lever, the increase is larger under the GVA method than under the DMV method, but this discrepancy is purely due to missing DMV data for some industries where improving knowledge can act as a secondary lever.

Figure 36 Size of markets relevant to each lever

Lever	Data Market Value Method (£m)		GVA Method (£m)	
	Primary effect	Primary + Secondary	Primary effect	Primary + Secondary
Improve knowledge / understanding of data sharing	331	2,241	225,808	631,748
Reduce costs of data sharing through better data foundations	6,342	6,570	556,966	615,355
Support (new) ways to address the risks associated with data sharing	4,647	4,647	651,602	651,602
Improve / demonstrate incentives for data sharing	1,429	5,982	241,658	525,565

³⁷ For example, the GVA of geospatial data is derived from estimates provided in the Cabinet Office’s *Analysis of the Potential Geospatial Economic Opportunity* (2018).

³⁸ The Lisbon Council and International Data Corporation (2020)

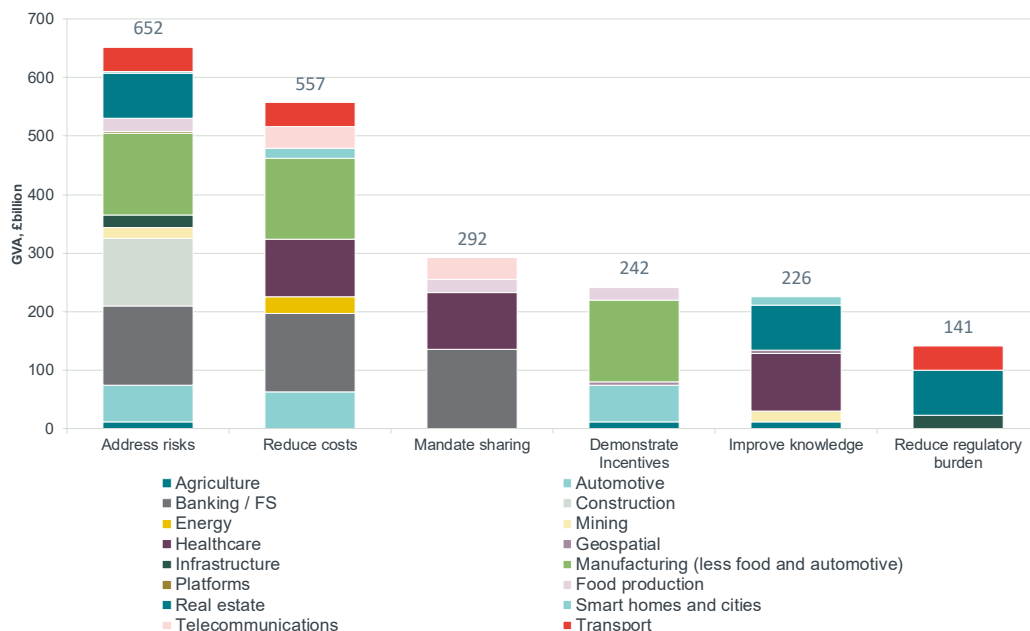
Reduce (perceived) regulatory burden associated with data sharing	427	427	140,841	140,841
Mandate data sharing where this is in the public interest	4,780	6,036	292,036	637,747

Source: Office for National Statistics, Regional Gross Value Added estimates 2018; European Data Market Monitoring Tool

Note: The values in each row should not be added because we count the value of each industry every time this is relevant. Adding rows would lead to double counting.

Figure 37 summarises the relevant GVA for each lever, with colour-coding to indicate the contribution of each industry to the GVA estimate. As an example, we can see from the chart that the “addresses risk” lever has larger contributions from financial services, construction and manufacturing, and smaller contributions from agriculture, automotive, mining, geospatial data, food production, smart homes and cities, and transport.

Figure 37 Size of markets by lever, GVA method



Source: GVA estimates based on ONS GVA by industry estimates for 2018

Note: The latest ONS GVA estimates disaggregated by industry and used in our analysis are from 2018. Therefore, any growth that has occurred since 2018 is not captured.

These results should be interpreted with caution for several reasons. First, as described in Section 5, there is relatively limited evidence about where in the economy data-sharing issues arise; this gap could be filled through additional primary research. Further evidence would also give us a better understanding of what levers might be of most use in tackling each of the issues. Specifically, it is possible that our assessment underestimates somewhat the activity relevant for the “knowledge” and “reduce (perceived) regulatory burden” levers:

- We are able to list industries/ecosystems as pertinent for the “knowledge” lever only if a lack of knowledge or understanding of data sharing in that area has been reported, typically by stakeholders. However, it is precisely in industries/ecosystems where there is very limited knowledge of data sharing that evidence is lacking about the issues preventing sharing.

- For the “reduce (perceived) regulatory burden” lever, it is possible that legal and regulatory risks may affect a broad range of industries (e.g. in connection with data protection). As a result, these risks, while important for many organisations, may not be fully reflected in evidence related to the specific issues that arise in particular industries.

Moreover, there may be some overlap between the “Smart homes and cities” and “Geospatial” ecosystems GVA and the GVA of the Energy, Transport and Real Estate industries. As a result, there may be limited double-counting of GVA for levers that affect at the same time one or both of the ecosystems and one or more of the other industries listed above. However, this is unlikely to affect significantly our results given the relatively limited size of the Smart homes and cities and Geospatial ecosystems compared to the industries considered in this exercise.

6.2 Applying the assessment criteria

As described in Section 5.3, our framework includes an assessment of the levers against general criteria (across the economy), and an ecosystem-specific evaluation. In this report, we provide an initial appraisal against the general criteria. As described below, owing to time constraints for this project, we offer only a partial assessment against the following criteria:

- **Strength of evidence:** here we consider evidence on data-sharing-specific interventions but not potentially relevant evidence on the impact of interventions in other policy areas (e.g. innovation support). This is because deciding what existing interventions are germane requires more detailed definition of the levers. For example, evidence pertaining to innovation policy could fit under any of the “knowledge”, “incentives”, “costs” and “risks” levers as currently defined. In the case of funding a research programme on data sharing, the associated lever would depend on the focus of the programme.
- **Cost of the lever:** again, this requires the levers to be defined more precisely, as each could entail very different costs depending on how specifically it is implemented. For example, the “knowledge” barrier may involve relatively light-touch interventions (disseminating existing knowledge) or more intensive intervention (generating new knowledge with significant investment in research & development).

Figure 38 General lever assessment criteria

Criterion	Assessed in this report?
Strength of evidence	In part: assessed evidence of data sharing-specific interventions but not of wider potentially relevant policies
Timing of impact	Yes
Indirect effects of lever	Yes
Unintended consequences and trade-offs	Yes
Cost of the lever	No – could be assessed in future work

We summarise our initial assessment in Figure 39 below and provide more details of each criterion in the remainder of this section.

Figure 39 Summary of general lever assessment

Lever	Key advantages	Key disadvantages
Improve knowledge / understanding of data sharing	Some evidence of positive impact; possible large indirect effects	Relatively slow impact
Reduce costs of data sharing through better data foundations	Some evidence of positive impact; possible indirect effects; relatively quick impact in improving the efficiency of existing data sharing	Relatively slow impact in leading to additional data sharing;
Support (new) ways to address the risks associated with data sharing	Relatively swift impact once ways to address risk are established	Impact may be relatively slow if developing new ways to address risk requires significant innovation
Improve / demonstrate incentives for data sharing	Possible large indirect effects	Trade-offs with data protection; risk of incentivising data sharing that may not ultimately have significant economic impact
Reduce (perceived) regulatory burden associated with data sharing	Relatively quick impact (once intervention is in place) ³⁹	Trade-offs with data protection
Mandate data sharing where this is in the public interest	Some evidence of positive impact	Risk of disincentivising data collection; risk of unfairly putting incumbents at a competitive disadvantage

Beyond the indirect effects described below and in following sections of this report, it is worth noting that trustworthy sharing that generates economic and/or social benefits, and is understood by others to do so, is likely to lead more organisations to consider sharing data.

6.2.1 Strength of evidence

There is very limited **evidence** available on existing data policy interventions. Our review of the evidence has included both targeted searches for evaluations of policies across the OECD, identified in the initial phase of our research, and general searches for academic research on this topic (e.g. via Google Scholar). Results from this exercise include the following findings.

There is evidence of a positive impact from an EU-funded accelerator programme, **DataPitch**, which provides start-ups with ideas for data-driven products together with support including initial funding and matching to potential data providers. We categorise this as a “knowledge” intervention because of this match-making component as well as the access it gives start-ups to peer networks and training materials. It is worth noting, however, that the intervention also includes help in

³⁹ Major changes to regulatory and legal frameworks may take a long time compared to other interventions, e.g. funding innovation programmes.

drafting data sharing agreements between start-ups and data providers, which could be classified as a “cost” or “regulatory burden” intervention.⁴⁰

There is evidence of positive impact of standards for data on opportunities for physical activity in the UK.⁴¹ This type of intervention improves data foundations and therefore fits under the “cost” lever.

There is limited initial evidence on the impact of the UK’s Open Banking initiative, which requires the country’s nine largest banks to grant users the right to share current account data with authorised third-party providers (TPPs) in a standardised way. While there has been no evaluation of the effect of Open Banking on outcomes such as competition in retail banking, there is evidence that data is being shared and that the volume of sharing is increasing over time. In September 2020, there were 541m successful API calls (a proxy for the amount of data being shared), a more-than-threefold increase from September 2019, when 138m API calls were completed.^{42 43}

The take-up of Open Banking was initially slow due to: (i) a lack of trust, which the CMA has sought to overcome using tools such as a codified approach to consent;⁴⁴ and (ii) limited investment in data foundations. From conversations with stakeholders, we have learnt that regulatory mandates have to be designed correctly in order to make the process for sharing data as easy as possible. For example, in Open Banking there was initially no rule for the authentication process. As a result, banks developed relatively cumbersome processes which have since been improved. Similarly, the revised Payment Services Directive (PSD2)⁴⁵ mandated that data had to be shared via Application Programming Interfaces (APIs), but it did not stipulate that banks had to use the same APIs. The result was that each bank developed its own API.

6.2.2 Timing of impact

The likely timing of the impact of an intervention is assessed on the basis of the stage of a “typical” data sharing journey. In practice, this journey will vary across ecosystems and may not always consist of wholly discrete steps. For example, an organisation may choose to consider the commercial risks from data sharing at the same time as defining the business case for sharing, rather than as a following step. As for other components of our framework, the assessment can be more precise when the framework is applied to a specific ecosystem. Figure 40 below maps each lever against the stage of the data sharing journey it would affect.

⁴⁰ London Economics (2020)

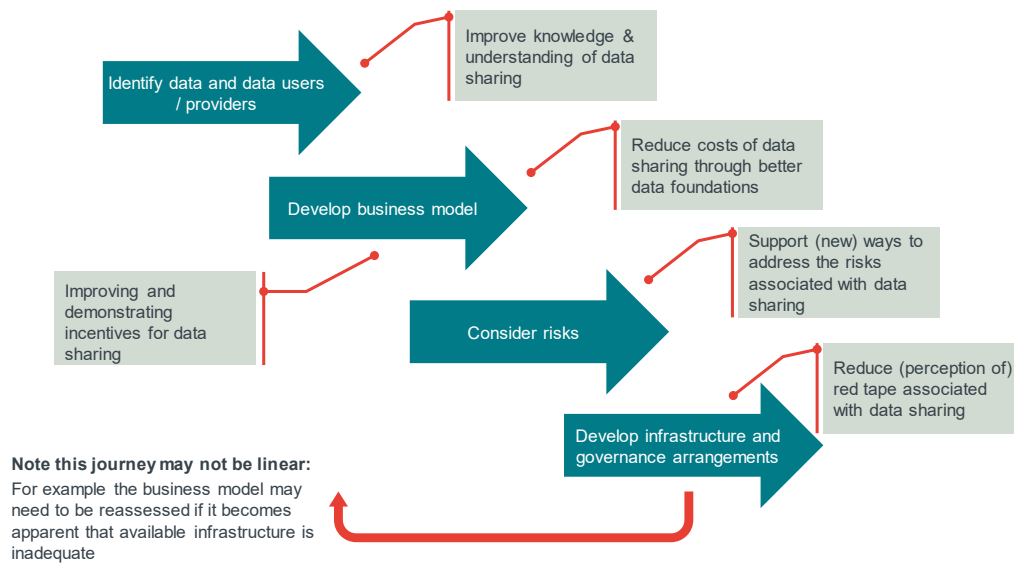
⁴¹ ODI – *OpenActive* (2020)

⁴² Open Banking – *APIs* (2020)

⁴³ The Open Banking Implementation Entity estimates the potential annual benefit from the initiative at £12bn for consumers and £6bn for SME users. Reynolds et al. (2019)

⁴⁴ Open Banking – *Trust* (2020)

⁴⁵ “The revised Payment Services Directive (PSD2) updates and enhances the EU rules put in place by the initial PSD adopted in 2007 [...]. The PSD2 supports innovation and competition in retail payments and enhances the security of payment transactions and the protection of consumer data” ECB (2018)

Figure 40 Mapping of levers to stages of data sharing journey

Note: “Mandate data sharing” could be relevant across all stages of the data-sharing journey

Interventions to **increase knowledge and understanding** of data sharing act on early stages of this journey, when organisations are exploring what data exists, how they could use it for their benefit and how sharing could be undertaken. Therefore, it may take some time for these interventions to generate more data sharing even if they succeed in improving knowledge of sharing. Data providers, users and intermediaries will still need to quantify the likely benefits and costs of sharing data, identify technical solutions and take other steps before proceeding.

Reducing the cost of data sharing may have a slightly quicker impact compared to improving knowledge. Interventions of this type would first have an effect in instances where organisations have identified the data they could access/share, and possibly the benefits that could arise, but find the costs too high to justify access.

Improving and demonstrating incentives for data sharing may act on different stages of the journey. Demonstration could lead more organisations to start considering sharing/accessing new data, and therefore the timing of impact might be similar to the “knowledge” lever. Improving incentives would have a more direct impact on the expected balance of costs and benefits from sharing. That could produce a relatively quick impact in cases where organisations have identified the data providers/users they would like to engage with but have held back because they judge the benefits to be limited.

Supporting ways to address risk would generally act on a relatively late stage of the data sharing journey. Commercial and technical risks may be considered once possible data providers/users, uses of data and business models have been identified. However, in some cases, e.g. where the data is particularly sensitive, high risks may prevent organisations from even considering the use of data. In those instances, new ways of addressing risk could have a significant, albeit delayed, impact on data sharing.

Reducing the (perceived) regulatory burden from data sharing is likely to have an effect on late stages of the journey. As a result, its impact is likely to be felt more quickly than other levers.

The timing of impacts from the “**mandate data sharing**” lever is harder to assess and may depend on the specifics of the intervention. For example, it may hinge on whether the specific uses of data that is currently not being shared are clear in principle and could be undertaken once the data is made available, or whether they would need to be identified after data sharing has been mandated.

6.2.3 Indirect effect of levers

All interventions can potentially increase data sharing beyond their effect on a specific issue. This may happen, for example, for the following reasons:

- There are economies of scale and/or scope in the use of a particular type of data. This could mean that when additional data is shared, the benefits from using it increase, which in turn may lead others to share, and so on. In principle, any intervention to increase data sharing may play a role in supporting this virtuous cycle.⁴⁶
- Increased data sharing may lead to a change in attitudes. Any change is likely to a positive one (more organisations considering data sharing) in the case of trustworthy sharing that generates economic and/or social benefits. Conversely, in the opposite scenario any change is likely to be negative.

For the purpose of assessing the levers defined in this report, however, we are more interested in the indirect effects of each lever.

Interventions that **improve knowledge and understanding of data sharing** also have the potential to help address all other barriers (lack of incentives, actual and perceived risks, retaining data for commercial interest). This is mainly for two reasons:

- A minimum level of knowledge and understanding of data sharing is necessary for organisations to start considering sharing. Once it has acquired this minimum knowledge, organisations can then assess the benefits, costs and risks of data sharing.
- Greater knowledge is likely to help organisations pinpoint the problems they need to consider when deciding whether to share and access data. For example, an understanding of how anonymisation and possible re-identification happen is likely to facilitate an effective assessment of the ethical, reputational, legal and regulatory risks involved in sharing personal data.

Interventions that **improve and demonstrate incentives** also have the potential to help overcome other barriers. That is because the greater the likely benefit from sharing data, the more likely organisations are to find ways to address the risks (and possibly to invest in data foundations). Demonstrating benefits can also make more people interested in data sharing and thereby increase the knowledge and understanding of data.

Interventions that **reduce the cost of data sharing** may have an indirect effect on the knowledge and incentives for sharing. Improved data foundations can lead to greater discoverability of the data. With lower costs and better foundations in place, organisations may find it easier to run pilots to explore new forms of data sharing

⁴⁶ This virtuous cycle assumes that those who share data also benefit, which may not always be the case.

and new combinations of data. This may be particularly beneficial because it is often claimed that the most valuable uses of data are not necessarily those that have been initially foreseen.⁴⁷ However, this “experimental” approach is possible only if the risks have been adequately addressed and if there is trust in data sharing.

It is less clear how interventions that **address the risk** of data sharing, **reduce the (perceived) regulatory burden** or **mandate data sharing** might have an indirect impact on other barriers, over and above the general indirect effects described above. In some ecosystems, addressing risk may be the intervention that “unlocks” data sharing, but this really depends on the specific data sharing barriers. Therefore, while addressing risks could have a large indirect impact on an ecosystem, surmounting other barriers is likely to have a greater indirect impact on the economy as a whole.

6.2.4 Trade-offs and unintended consequences

Interventions to increase access to data may lead to unintended adverse consequences. They may also entail trade-offs against other policy objectives. With every lever, the protection of data and of intellectual property may at times need to be traded off against increased access to data. This trade-off is particularly relevant for some of the levers. Figure 41 describes our assessment, based on the initial logic models described in Section 4 of this report. These trade-offs and possible unintended consequences may warrant defining mitigation strategies when proceeding to implement a set of interventions. Defining the levers in greater detail and applying them to specific ecosystems would allow a more granular assessment of the possible unintended consequences and of the risk that they might occur.

⁴⁷ OECD (2015).

Figure 41 Potential unintended consequences and trade-offs in using levers

Lever	Potential unintended consequences and trade-offs
Improve knowledge / understanding of data sharing	<ul style="list-style-type: none"> ■ Reduce scope for improving knowledge – There is a risk that ‘best practice’ guidelines, e.g. on technical infrastructure for data sharing, could quickly become outdated and hinder adoption of innovative solutions. ■ Disincentivise effort to find data/providers/users – “Matchmaking” activities may reduce the incentives of firms to put effort in themselves as they are less likely to gain a competitive advantage.
Reduce costs of data sharing through better data foundations	<ul style="list-style-type: none"> ■ Reduce scope for innovation – Mandating standards might limit the potential/incentive to explore alternative ways of storing, sharing and reusing data.
Support (new) ways to address the risks associated with data sharing	<ul style="list-style-type: none"> ■ Reduce incentive for data users – Over-protecting intellectual property could reduce incentives for data users to access and manipulate the data if they don’t own the outputs. ■ Create barriers to entry – New, potentially expensive technologies could create barriers to entry which favour incumbents.
Improve / demonstrate incentives for data sharing	<ul style="list-style-type: none"> ■ Encourage the over-collection of data – Incentives could incentivise data collection that has limited economic and social benefit or even net social costs. ■ Encourage data hoarding – If organisations believe that tax incentives for sharing some types of data will eventually be extended to others, they may choose to wait for that to happen before sharing their data.
Reduce (perceived) regulatory burden associated with data sharing	<ul style="list-style-type: none"> ■ Privacy concerns – Revisions of legislation could harm data privacy. This could in turn undermine trust and so lead to less data sharing. ■ Competition concerns – Revisions of competition legislation to enable coordination and data sharing could facilitate collusion. ■ Regulatory uncertainty – Risk that regulatory changes may lead to increased uncertainty on how to comply with regulation.
Mandate data sharing where this is in the public interest	<ul style="list-style-type: none"> ■ Disincentivise data collection and maintenance – Firms may stop collecting or maintaining data in order to avoid being subject to the mandate. ■ Distortion of competition – Risk that intervention puts incumbents at an unfair disadvantage or that it increases the regulatory burden on new entrants; risk of encouraging collusion.

7 Conclusions

7.1 Summary of our findings

This report aims to provide a framework to prioritise possible interventions to increase access to data from the private and third sectors. The framework starts by considering the reasons why government may be well placed to intervene in this domain and then picks out the key issues that may prevent optimal data sharing. We then identify six possible levers for intervention and provide an initial assessment of how much economic activity they could affect, alongside their advantages and disadvantages.

The analysis included in this report suggests that improving knowledge and understanding of data sharing, improving or demonstrating incentives, supporting ways to address risk, reducing the cost of sharing through data foundations, reducing the (perceived) regulatory burden and mandating data sharing could all form part of a package of effective interventions.

Addressing risk and reducing the cost of sharing are the levers that, based on our initial assessment, are likely to have a direct effect on the largest amount of economic activity. However, improving incentives and increasing knowledge of data sharing are particularly likely to have wide-ranging, indirect effects. There is very limited research to hand on the effectiveness of existing interventions, but there is some evidence for the effectiveness of demonstration activities and for mandating data sharing where there is a clear case that this could lead to the development of additional services (e.g. current account comparison services in the case of Open Banking) or to increasing choice and competition.

However, given the breadth of economic activity where additional data sharing may be beneficial, the array of issues that may prevent sharing, and the sparsity of the evidence base, these findings should be interpreted with caution. This study provides a starting point for the development of public policy in this area rather than a set of firm conclusions that, by themselves, can inform the elaboration of evidence-based policy interventions. The next section provides recommendations for further research that could help fill the existing evidence gaps.

7.2 Opportunities for future research

Research on the extent of and issues in data sharing

There is limited evidence on the issues that prevent data sharing, and in particular on where they arise. Which industries, ecosystems, types of organisation and types of data are particularly affected by each issue? It would be possible to fill the gaps in the existing evidence base through primary research, including a survey of organisations in the private and third sectors and/or further stakeholder engagement in the industries for which there is currently inadequate evidence.

The sizing analysis described in Section 6.1 could be enhanced through a more precise definition of data ecosystems and the activities they cover. This could involve:

- A more precise assessment of the current size of the data ecosystem:
 - Measuring the activities of “core” data providers in the ecosystem, that is organisations for which the supply of data is a core part of their product or service offering. The metrics used could include revenues and/or funding raised by “core” organisations; and
 - Assessing the extent to which providing data is a significant part of the activities of other (non-core) organisations. This assessment could be based on qualitative interviews or from surveys asking organisations the extent to which they engage in data sharing. This would likely include exchanges of data within existing relationships (e.g. supply chains). This analysis may allow part of the revenues or investment of non-core organisations to be allocated to the value of the data ecosystem.
- A more precise assessment of the importance of data sharing issues in the ecosystem and of the activities that may be unlocked by addressing those issues. Existing evidence mainly describes in very general terms the types of issues that arise in an industry. Stakeholder engagement and/or primary data collection could strengthen the evidence base by delving into how many organisations are affected by each data sharing issue and what would change if the issue was resolved or mitigated.

In addition, future work could further investigate how the characteristics of data ecosystems influence the effectiveness of different levers. This would build on our initial findings by:

- studying additional data ecosystems; and/or
- considering other characteristics of data ecosystems, potentially including the jurisdiction of organisations and data involved (confined to the UK or international), the nature of data collection mechanisms (data volunteered by individuals, generated through tracking, inferred from other data); data visibility (Is it shared with named individuals, specific parts of an organisation, an entire organisation? Can it be reshared by the organisation accessing it?); how the data is stored and shared from a technical point of view.

Research into the impact of data sharing

This report focuses on the issues that may prevent optimal data sharing and on how government interventions could address them. However, it could be useful to prioritise policy interventions in those areas where the resulting data sharing might have the greatest economic and social impacts. Our review of the evidence suggests that additional research into these impacts could be useful, potentially including:

- Quantitative analysis to investigate the impact of data sharing on company performance.
- Developing a framework to complement work undertaken in this study, in order to try and prioritise interventions that are more likely to produce a specific type of benefit, such as:
 - Increasing the efficiency of existing processes or the quality of existing products;

- Innovation (enabling the development of new or better products and services);
- Supporting market competition;
- Enabling the redistribution of benefits and costs from the use of data (e.g. so that consumers, or specifically vulnerable consumers, benefit more); and
- Addressing specific policy issues (e.g. reducing regional inequality).

8 References

- Acquisti, A., & Grossklags, J. (2007). What can behavioral economics teach us about privacy. *Digital privacy: theory, technologies and practices*, 18, 363-377
- Alan Turing Institute, 2020. Road-mapping Uses of Advanced Analytics in the UK Food and Drink Sector. <https://science-council.food.gov.uk/sites/default/files/appofadvancedanalyticsfinalreport.pdf>
- BBC, 2018. Fitness app Strava lights up staff at military bases. <https://www.bbc.co.uk/news/technology-42853072>
- BCG, 2020. Share to Gain: Unlocking Data Value in Manufacturing. http://www3.weforum.org/docs/WEF_Share_to_Gain_Report.pdf
- BEIS, 2018. Modernising Consumer Markets. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/699937/modernising-consumer-markets-green-paper.pdf
- BEIS, 2020. Smart Data Impact Assessment. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/915974/smart-data-impact-assessment.pdf
- BEIS, 2020. Next Steps for Smart Data. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/915973/smart-data-consultation-response.pdf
- Cabinet Office, 2018. An Initial Analysis of the Potential Geospatial Economic Opportunity. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/733864/Initial_Analysis_of_the_Potential_Geospatial_Economic_Opportunity.pdf
- CECE, 2019. Digitising the Construction Sector. <https://www.cece.eu/publications/digital-reports>
- CCDR, 2019. Reaping the benefits of Open Data in public health. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6781855/#r20>
- CMA, 2020. Online platforms and digital advertising. https://assets.publishing.service.gov.uk/media/5fa557668fa8f5788db46efc/Final_report_Digital_ALT_TEXT.pdf
- Coyle, D., Diepeveen, S., Widowin, J., Tennison, J., Kay, L., 2020. The Value of Data – Policy Implications. https://mk0nuffieldfounpg9ee.kinstacdn.com/wp-content/uploads/2020/02/Value_of_data_Policy_Implications_Report.pdf
- Deloitte, 2017. Data Sharing in Infrastructure. https://nic.org.uk/app/uploads/Data_sharing-in-infrastructure.pdf
- Department for Digital, Culture, Media and Sport, 2020. National Data Strategy. <https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy>

Digital Competition Expert Panel, 2019. Unlocking Digital Competition.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf

European Commission (EC), 2018. Staff Working Document - Guidance on sharing private sector data in the European data economy.
<https://ec.europa.eu/digital-single-market/en/news/staff-working-document-guidance-sharing-private-sector-data-european-data-economy>

EC, 2019. ME panel consultation - B2B Data Sharing.
<https://ec.europa.eu/digital-single-market/en/news/sme-panel-consultation-b2b-data-sharing>

EC, 2020. Digital Economy Working Paper 2020-05: Business-to-Business data sharing: An economic and legal analysis.
<https://ec.europa.eu/jrc/sites/jrcsh/files/jrc121336.pdf>

EC, 2020. The European Data Market Monitoring Tool.
http://datalandscape.eu/sites/default/files/report/D2.9_EDM_Final_study_report_16.06.2020_IDC_pdf.pdf

European Central Bank (ECB), 2018. The revised Payment Services Directive (PSD2) and the transition to stronger payments security.
https://www.ecb.europa.eu/paym/intro/mip-online/2018/html/1803_revisedpsd.en.html

Fabrizio, K. R. (2013). The effect of regulatory uncertainty on investment: evidence from renewable energy generation. *The Journal of Law, Economics, & Organization*, 29(4), 765-798

Frontier Economics (2020). Geospatial Data Market Study. Report prepared for the Geospatial Commission.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/937025/Frontier_Economics_-_Geospatial_Data_Market_Study.pdf

Hall et al, 2017. Growing the artificial intelligence industry in the UK.
<https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk>

HMG, 2017. Industrial Strategy.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf

International Data Corporation (IDC), 2016. Opening up Private Data for Public Interest.
http://datalandscape.eu/sites/default/files/report/Story_1_New_format.pdf

Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the Economics of Data. *American Economic Review*, 110(9), 2819-58

Lisbon Council, 2020. Final Study Report: The European Data Market Monitoring Tool Key Facts & Figures, First Policy Conclusions, Data Landscape and Quantified Stories.
https://datalandscape.eu/sites/default/files/report/D2.9_EDM_Final_study_report_16.06.2020_IDC_pdf.pdf

London Economics, 2019. Independent assessment of the Open Data Institute's work on data trusts and on the concept of data trusts.

<https://theodi.org/wpcontent/uploads/2019/04/Datatrusters-economicfunction.pdf>

London Economics, 2020. Data Pitch evaluation. Report prepared for Data Pitch.

<https://datapitch.eu/wp-content/uploads/2020/01/London-Economics-Data-Pitch-evaluation-FINAL-PDF.pdf>

McKinsey, 2019. Smarter cities are resilient cities.

<https://www.mckinsey.com/business-functions/operations/our-insights/smarter-cities-are-resilient-cities>

National Audit Office (NAO), 2014. Using alternatives to regulation to achieve policy objectives.

<https://www.nao.org.uk/wp-content/uploads/2014/06/Using-alternatives-to-regulation-to-achieve-policy-objectives1.pdf>

National Audit Office, 2017. Vulnerable consumers in regulated industries.

<https://www.nao.org.uk/wp-content/uploads/2017/03/Vulnerable-consumers-in-regulated-industries.pdf>

Nelson, P. (1970). Information and consumer behavior. *Journal of political economy*, 78(2), 311-329.

OECD, 2015. Data-Driven Innovation: Big Data for Growth and Well-Being.

OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264229358-en>.

OECD, 2018. Enhancing Access to and Sharing of Data. OECD Publishing, Paris

https://www.oecd-ilibrary.org/science-and-technology/enhancing-access-to-and-sharing-of-data_276aaca8-en

Open Data Institute (ODI), 2018a. Mapping Data Ecosystems.

<https://docs.google.com/document/d/1vSqoHOYT5u6vrCHlebCS0rze0gWwXOspEeowWzwake8/edit#>

ODI, 2018b. What organisations need in order to share more data.

<https://theodi.org/article/what-organisations-need-in-order-to-share-more-data-our-research/>

ODI, 2019. Data trusts: lessons from three pilots. <https://theodi.org/article/odi-data-trusts-report/>

ODI, 2019. The ODI calls on companies and governments to build trust by

engaging with their customers and citizens about how they use personal data.

<https://theodi.org/article/nearly-9-in-10-people-think-its-important-that-organisations-use-personal-data-ethically/>

ODI, 2019. ODI Food Waste data trust. <https://theodi.org/article/data-trusts-food-waste/>

ODI, 2020. Data Toolkit for Businesses. <https://theodi.org/service/tools-resources/data-toolkit-for-business/>

ODI, 2020. The Data Spectrum. <https://theodi.org/about-the-odi/the-data-spectrum/>

ODI, 2020. Activating sector change: what we've learnt from OpenActive.

<https://theodi.org/article/activating-sector-change-what-weve-learnt-from-openactive/>

Office of Fair Trading, 2020. Government in markets: why competition matters – a guide for policy makers.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/284451/OFT1113.pdf

Open Banking, 2020. Open Banking APIs Performance.

<https://www.openbanking.org.uk/providers/account-providers/api-performance/>

Open Banking, 2020. Trust in Open Banking.

<https://www.openbanking.org.uk/insights/targeting-trust-in-open-banking/>

Policy Lab, 2020. Introducing a Government as a System toolkit.

<https://openpolicy.blog.gov.uk/2020/03/06/introducing-a-government-as-a-system-toolkit/>

Rovatsos M., Mittelstadt, B. & Koene, A. (2020). Landscape Summary: Bias in Algorithmic Decision-Making. Report prepared for the Centre for Data Ethics and Innovation.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819055/Landscape_Summary_-_Bias_in_Algorithmic_Decision-Making.pdf

RICS, 2019. The use and value of commercial property data.

<https://www.rics.org/uk/news-insight/research/insights/the-use-and-value-of-commercial-property-data/>

Reynolds et al (2019). Consumer priorities for Open Banking.

<https://www.openbanking.org.uk/wp-content/uploads/Consumer-Priorities-for-Open-Banking-report-June-2019.pdf>

Royal Academy of Engineering, 2019. Towards trusted data sharing: guidance and case studies. <http://reports.raeng.org.uk/datasharing/cover/>

Royal College of Psychiatrists, 2020. Technology use and the mental health of children and young people. <https://www.rcpsych.ac.uk/docs/default-source/improving-care/better-mh-policy/college-reports/college-report-cr225.pdf>

Royal Society, 2018. AI and data governance from principles to practice: auto insurance. <https://royalsociety.org/-/media/policy/topics/open-science-data/workshop-notes.pdf>

Transport Systems Catapult, 2017. The case for government involvement to incentivise data sharing in the UK intelligent mobility sector.

<https://ts.catapult.org.uk/intelligent-mobility/im-resources/opendata/>

Turner, S., Galindo Quintero, J., Turner, S., Lis, J., & Tanczer, L. M. (2020). The exercisability of the right to data portability in the emerging Internet of Things (IoT) environment. *New Media & Society*, 1461444820934033

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, 185(4157), 1124-1131

Annex A Further detail on evidence of data sharing issues

Figure 42 Summary of evidence on data sharing issues

#	Data sharing issue	Is there evidence of this issue in specific areas?
1	Data providers unwilling to make data available due to uncertainty as to benefits (relative to costs)	<p>Evidence is specific to geospatial and fresh produce (food manufacturing), agriculture, automotive and manufacturing.</p> <p>The ODI Food Waste data trust pilot found that fresh produce manufacturers need to better understand the costs to the business of sharing data, what the benefits are and when they will accrue in order to share more data. (ODI - Food Waste, 2019)</p> <p>Frontier Economics, Geospatial Data Market review (2020)</p> <p>An EC panel consultation with SMEs found that respondents from the agricultural, automotive and manufacturing sectors are most likely to try to acquire data held by another company. Many businesses in these industries encountered difficulties in obtaining data due to unreasonable acquisition practices, including high licensing fees, long lead times, unfavourable contracts, etc. In addition, many firms do not include data sharing in their business models and so are unable to put a value on sharing. (EC, 2019)</p>
2	Data providers unwilling to make data available because the benefits of investment do not outweigh the costs	<p>Evidence is specific to automotive insurance and, to a lesser extent, to agriculture, automotive and broader manufacturing industries.</p> <p>Royal Society interviews with stakeholders in the automotive industry found that they were dissuaded from sharing data due to a lack of established business models that could identify whether the benefits of sharing were greater than the costs. (Royal Society, 2018)</p> <p>See final point under data sharing issue 1 above (EC, 2019)</p>
3	Organisations sharing / accessing data do not know where to find users / providers	<p>Evidence is specific to the extractive and healthcare industries.</p> <p>The ODI found that interviewees in these sectors had "experienced challenges in finding relevant data". (ODI, 2018a)</p>
4	Organisations sharing / accessing data do not know whether data is fit for purpose	<p>Evidence is specific to geospatial data.</p> <p>Frontier Economics, Geospatial Data Market review (2020).</p>

#	Data sharing issue	Is there evidence of this issue in specific areas?
5	Organisations do not appreciate the benefits of using / accessing data.	<p>Evidence is specific to smart cities and agriculture.</p> <p>The Copenhagen Hitachi City data project held workshops and interviews with businesses in the Danish capital and found users often did not know which data was useful for them. (IDC, 2016)</p> <p>An EC panel consultation with SMEs found that respondents from the agricultural sector would most value data-sharing guidance and support that focused on including data in business models, best practice examples of data sharing and elements that should be considered in contractual negotiations. (EC, 2019)</p> <p>In general, it was particularly hard to find to evidence this problem. It is likely that if an organisation/industry is unaware of the benefits of sharing data then there will be little research and evidence on data sharing in that industry.</p>
6	Organisations perceive high risks in data sharing	<p>Evidence is specific to transport, infrastructure, food manufacturing and construction.</p> <p>The Transport Systems Catapult interviewed stakeholders in transport and found that: "A number of industry respondents highlighted ...risks around transport data being breached and used maliciously as the main reason for not sharing data. The concern was expressed that if data they had generated or were responsible for was compromised, they would be liable for any consequences arising from a loss of privacy, reduced security and reduced safety." (Transport Systems Catapult, 2017)</p> <p>Interviews with stakeholders and academics in the infrastructure industry found that recent privacy legislation (e.g. GDPR) had encouraged a risk-averse approach to data sharing in the sector. (Deloitte, 2017)</p> <p>The Committee for European Construction Equipment interviewed stakeholders across the building sector and found that the perceived risks associated with data sharing (in particular around cyber security) were a barrier to sharing. To overcome this hurdle, the CECE recommended that properly specified data protection and access permission controls should be developed and taken into account in data sharing agreements. (CECE, 2019)</p>

#	Data sharing issue	Is there evidence of this issue in specific areas?
7	Organisations have or perceive a lack of trust in data sharing	<p>Evidence is specific to the extractive, transport, automotive, agricultural and manufacturing industries.</p> <p>The ODI Data Access interviews found that in the extractives industry companies aren't trusted to release unbiased data. In addition, data released by governments which are not generally seen as credible will not be trusted. (ODI, 2018b)</p> <p>Interviews with stakeholders and academics in the infrastructure industry identified a culture of opposing data sharing due to underlying trust issues. There is also a culture of suspicion and a lack of trust in infrastructure organisations on the part of customers. (Deloitte, 2017)</p> <p>An EC panel consultation with SMEs found that respondents in the automotive, agriculture and manufacturing sectors value trust and transparency in data sharing agreements. They indicated that such agreements should include safeguarding and protection of companies' commercial secrets; they should also be transparent about who can access relevant data and not aim to distort competition or to lock companies into deals with one provider only. (EC, 2019)</p> <p>Transport Systems Catapult (2017) surveyed transport organisations and found that: "A number of industry respondents highlighted ...risks around transport data being breached and used maliciously as the main reason for not sharing data. The concern was expressed that if data they had generated or were responsible for was compromised, they would be liable for any consequences arising from a loss of privacy, reduced security and reduced safety." (Transport Systems Catapult, 2017)</p>

<p>8 Private companies may retain data because they are worried about losing a potential source of competitive advantage</p>	<p>Evidence is specific to infrastructure, finance, extractives, manufacturing (including food manufacturing), automotive insurance, IoT, platforms and construction.</p> <p>Interviews with stakeholders and academics in the infrastructure industry identified the perceived loss of competitive advantage as a barrier to data sharing. In addition, "Commercial barriers are higher where data sharing and collaborative approaches are not the norm, as there may be a 'free rider' problem where data shared by one firm is used by other parties that do not reciprocate by sharing their own data." (Deloitte, 2017)</p> <p>The ODI Data Access Interviews found the extractives industry is particularly unwilling to share data due to a long history of secrecy both within companies and their government counterparts, and for fear that sharing reduces competitive advantage. (ODI, 2018)</p> <p>Interviews with stakeholders in manufacturing identified fear of losing negotiating power or competitive advantage and fear of unintentionally giving away valuable or sensitive business data as the top trust-related barriers to data sharing. (BCG, 2020)</p> <p>An EC panel consultation with SMEs found that the majority of respondents agreed that data generated by the IoT presents particular challenges for market fairness. (EC, 2019)</p> <p>Open Banking is an example of greater data sharing and access improving competition. The Smart Data Review said there is an opportunity for a similar data sharing initiative (Open Finance) to increase competition in other areas of finance, including savings, mortgages, pensions, insurance, consumer credit and investments. (BEIS, 2020)</p> <p>The Committee for European Construction Equipment interviewed construction companies about the digitisation of the sector. It found that firms are reluctant to share data for fear of surrendering their know-how. Some players also believe IT companies could erode part of their market share if data is not adequately protected. (CECE, 2019)</p> <p>The Royal College of Psychiatrists recently called for an urgent review to establish a protocol for the mandated sharing of social media companies' data with research organisations in order to safeguard the mental health of children and young people. Currently, data from social media platforms is provided at a population level and is concerned only</p>
--	---

#	Data sharing issue	Is there evidence of this issue in specific areas?
		<p>with screen times. The college argues that meaningful analysis requires more detailed data across a range of indicators (Royal College of Psychiatrists, 2020)</p> <p>The Alan Turing Institute interviewed food sector practitioners and found that companies are often reluctant to share production data because it is typically commercially sensitive. (Alan Turing Institute, 2020)</p>
9	Risks associated with data sharing has eroded trust	<p>Evidence specific to consumer data and transport. The Transport Systems Catapult found that: "A number of industry respondents highlighted ...risks around transport data being breached and used maliciously as the main reason for not sharing data. The concern was expressed that if data they had generated or were responsible for was compromised, they would be liable for any consequences arising from a loss of privacy, reduced security and reduced safety." (Transport Systems Catapult, 2017)</p> <p>Frequent high-profile data breaches can increase the concern over reputation for firms and discourage them for sharing data. E.g. the Marriot Hotel Group suffered from hacking of sensitive customer booking data. They consequently incurred a fine of 4% of turnover and the reputational damage, as measured by a YouGov survey, was a shift of respondents having a positive to negative impression of the brand after the incident.</p>
10	Organisations are concerned that data sharing may lead to harm / breach of the law, regulations or intellectual property rights	<p>Evidence is specific to the transport industry.</p> <p>The Transport Systems Catapult Report found that stakeholders expressed concern about the application of competition law, "...for fear of their organisation being seen to use data sharing as a means of collusion. This concern was influenced by previous competition investigations in the sector and new concerns that have been raised relating to telematics data from vehicles." Local bus franchises have struggled to share fare data because of competition law concerns. (Transport Systems Catapult, 2017)</p>

#	Data sharing issue	Is there evidence of this issue in specific areas?
11	There is a lack of coordination in investment in data foundations	<p>Evidence is specific to smart devices/IoT and manufacturing.</p> <p>The London Economics-ODI report on the data trust pilots concluded that the absence of standardisation on data access structures is a coordination failure. (London Economics, 2019)</p> <p>A study of smart home devices found that when companies were requested by customers to transfer data to a rival's device, they all said they could not share data directly with competitors due to differences in technical infrastructure; they said they could share via individual customers. (News Media & Society, 2020)</p> <p>Differing levels of digital maturity and the cost of switching technologies (or fear of technological lock-in) are barriers to data sharing in manufacturing (BCG, 2020)</p>
12	Organisations do not invest sufficiently in data foundational issues that would increase efficiency	<p>Evidence is specific to healthcare, smart data (consumer utilities) and transport industries.</p> <p>In healthcare, "The collection, processing, storing and access of complex data coming from different structured (e.g. national health records) and unstructured sources present challenges to use." (EC, 2020)</p> <p>The BEIS Smart Data Impact Assessment found that, "While the private sector could in principle develop effective standards for customers to share data with third parties, in key markets they have failed to materialise. This may be due to concentrated costs and dispersed benefits – i.e. whilst customers and innovative companies would benefit, the implementation costs narrowly fall upon incumbents. Therefore, because incumbents are uncertain the benefits will accrue to them, implementation is not in their best interests." (BEIS - Impact Assessment, 2020)</p> <p>Interviews with stakeholders and academics found that inadequate data formats in the private and public sectors were restricting data sharing and use in the infrastructure industry; there was also a lack of commercial incentives to make the investments needed to overcome technical hurdles. (Deloitte, 2017)</p> <p>The OECD reviewed data practices worldwide and found that: "Interoperability is a major issue for data re-use across applications. Even when commonly used machine-readable formats are employed, interoperability is not guaranteed." (OECD, 2018)</p>

#	Data sharing issue	Is there evidence of this issue in specific areas?
13	Organisations lack access to technologies/infrastructure that support data sharing	<p>Evidence is specific to smart cities, energy and automotive.</p> <p>The Copenhagen Hitachi City data project held workshops and interviews with businesses in the Danish capital and found that both data users and providers lacked the time and resources to invest in data. Companies also lacked the skills to understand and manipulate the data they do have. The Open Data for Energy Conservation project identified an additional barrier to sharing: the significant time needed to collate and format data. (IDC, 2016)</p> <p>An EC panel consultation with SMEs found that respondents from the automotive sector felt that technical guidance would be among the most useful means of data sharing support. (EC, 2019)</p>
14	There may be opportunities to use data to enhance competition, aid regulatory enforcement and support research (but there is no direct benefit for data holders)	<p>Evidence is specific to smart data (consumer utilities), health data and food manufacturing.</p> <p>Evidence reviewed within the BEIS Smart Data Impact Assessment found that, "while the private sector could in principle develop effective standards for customers to share data with third parties, in key markets they have failed to materialise. This may be due to concentrated costs and dispersed benefits – i.e. whilst customers and innovative companies would benefit, the implementation costs narrowly fall upon incumbents. Therefore, because incumbents are uncertain the benefits will accrue to them, implementation is not in their best interests" (BEIS-Impact Assessment, 2020)</p> <p>There are many examples of data sharing that is in the public interest. Canada's Communicable Disease Report identified two: in healthcare (by combining social media hits with sexually transmitted diseases to improve the effectiveness of disease outbreak control) and in health and safety (using data from satellites and weather indicators to help predict extreme weather events). (CCDR, 2020)</p> <p>The Alan Turing Institute interviewed food sector practitioners and found that data users often faced problems in accessing data in time for it to be used effectively. Some progress has been made by using nowcasting techniques to provide estimates of current values. (Alan Turing Institute, 2020)</p>
15	Organisations are reluctant to share data that may expose them to unwanted scrutiny, even when in the public interest	See row 14 above

Source: Frontier Economics, rapid review of literature on data sharing issues

Annex B Further detail on case studies of data sharing ecosystems

B.1 One-to-many case study 1: smart meters

As part of the government's drive to upgrade the UK's national energy infrastructure, around 4.3m smart meters had been installed and connected to the national network as of the end of March 2020.⁴⁸ A new four-year framework will be introduced from June 2021 which will give each energy supplier an annual installation target to help speed up the roll-out further.

As well as providing a link between the smart meter and in-home display, a communications hub allows information to be exchanged between smart meters and third-party organisations that use the data. These include competing energy suppliers, energy network operators and other authorised parties, such as third-party intermediaries that offer services to do with energy saving, switching suppliers or load shifting and demand-side responses.

The design of the communications infrastructure gives any party access to any piece of data, subject to consumer authorisation, except for sensitive data such as security information.

Data Chain

There are four different types of players that benefit from smart meters:

- **Consumers:** buy energy from suppliers, send (receive) data on consumption (prices) from smart meters
- **Energy suppliers:** buy energy from operators, receive (send) data on consumption (prices) to consumers
- **Network operators:** transport the energy to suppliers and can access anonymised/aggregated smart meter data
- **Third parties:** these are external players interested in smart meter data, directly from consumers or indirectly from energy suppliers

Smart DCC Ltd has built the national infrastructure which enables communications between smart devices, energy suppliers, network operators and third parties.

What data sharing issues does this project aim to address?

The UK government viewed smart meters as vital to upgrade the national energy infrastructure and underpin the cost-effective achievement of its goal to reduce net zero carbon emissions to zero.⁴⁹ Smart meters were also perceived to offer consumers the benefit of being able to better manage their energy use, make savings and reduce emissions. It was also hoped that switching suppliers would

⁴⁸ This figure relates to "Smet2" meters. In total around 21.5m smart meters had been installed but most of these were the earlier model (Smet1 meters) which do not have full functionality.

⁴⁹ https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/893124/delivering-smart-system-post-2020-govt-response-consultation.pdf

be smoother and faster with smart meters, making it easier to get the best deals.

50

Data sharing was undoubtedly a necessary component to achieve these objectives. However, unlike the other case studies we have reviewed, it was not the primary objective.

With this in mind, there were a number of obstacles in the way of data sharing that smart meters have begun to address:

- Reduce the risks of data sharing:
 - Through localised storage: DCC will not maintain a centralised database of consumption data, which is stored instead in the smart meters on the consumer's premises;
 - Through maintaining data access control: consumers must give their consent to allow parties to access detailed consumption and tariff information from their electricity and gas meters;
 - Through a training and accreditation obligation: users of DCC data are required to undergo security and privacy assessments to demonstrate compliance with obligations;
 - Through a self-governing code of conduct: the Smart Energy Code (SEC) sets out the rights and obligations of energy suppliers, network operators and other relevant parties involved in the management of smart metering in Great Britain; and
 - Through investment in strict security measures: data is transmitted from the device in the form of encrypted packets to which only authorised users have access.
- Reduce the lack of incentive to share (and access) data:
 - Through additional functionality by sharing data: so-called consumer access devices (CADs) can be added to the smart meter network. Today's CADs are small boxes which connect to WiFi routers to stream energy data, but the CADs of tomorrow could be anything from a tumble drier to a home automation hub. This should make smart meters more valuable for consumers, for example if they can programme their drier to turn on when energy prices are low.
- To adjust market outcomes (increase competition):
 - Mandate data sharing: certain organisations within the energy industry, such as domestic energy suppliers and network operators, are obliged through their licences to become DCC users. In March 2020, five suppliers had failed to comply and were banned from taking on new customers until they became DCC users.

What other data-sharing issues may need addressing in the future?

While the roll-out of smart meters has increased since government intervened, the objectives of the scheme are far from having been fully met. In 2019 take-up was still below expectations and half of the larger suppliers missed their self-imposed annual targets by more than 10%.

⁵⁰ <https://www.gov.uk/guidance/smart-meters-how-they-work#benefits-of-smart-meters>

A fundamental reason for the disappointing progress is likely to be that consumers still lack the incentives to install a smart meter. Here are some levers that could increase take-up and the barriers to acceptance they can help to overcome:

- Reduce the perception of risk associated with data sharing:
 - Lack of trust: consumers do not always trust energy suppliers with their data; third parties are trusted even less. This does not appear to be a barrier for the majority of consumers. However, distrust may have increased following the Cambridge Analytica scandal.⁵¹
- Reduce the regulatory burden and cost of sharing data
 - Cost/hassle of set-up: consumers might perceive the installation of a smart meter as an inconvenience; and
 - Circumstantial barriers: some consumers (e.g. renters) are less likely to want to install a meter.
- Demonstrate the benefits of data sharing
 - Awareness of the benefits: consumers might not be informed about the monetary, societal or convenience benefits that can arise from data sharing and using third-party services.
- Reduce the lack of incentive to share (and access) data:
 - The advent of half-hourly measurements potentially greatly increases the volume, quality and use cases of data available. However regulatory changes are required for domestic tariffs to vary on a half hourly basis. Without this, there is a reduction in the demand for consumers to share their data.

B.2 One-to-many case study 2: APROCONE

The Advanced Product Concept Analysis Environment project (APROCONe) is a £19.2m exercise jointly funded by government, universities and industry. Within APROCONE sits a proof-of-concept co-design platform whose objective is to speed up the design and testing of complex aircraft components⁵². This is achieved by letting partners access, use and share data in real time through the platform.

The platform can enable better and more timely data sharing. This allows supply chain partners to carry out design studies in parallel as they no longer need to wait for the latest information to be sent⁵³. This can result in faster design processes, better product design and ultimately cheaper products.

What data-sharing issues does this project aim to address?

The APROCONE project aims to tackle the following problems.

- Reducing the risks of data sharing:

⁵¹ <https://www.smartenergydatapiag.org.uk/> - Stimulus Paper 6 - Consumer Research on Access to Smart Meter Data

⁵² <https://www.cranfield.ac.uk/press/news-2016/aprocone-new-aerospace-research-investment-to-use-cranfield-technology>

⁵³ <http://reports.raeng.org.uk/datasharing/case-study-6-aprocone/>

- Through investment in security measures: ID-based user access; adding or removing systems and partners is straightforward; and
- Through commercial frameworks which enable companies to maintain their intellectual property rights (although these are still being refined).
- Reducing the cost of data sharing:
 - The platform has enabled real-time data sharing. Previously, design data was shared digitally at pre-arranged times or in response to formal requests. To meet these requests data had to be processed and prepared, which took time. Eliminating these additional steps significantly speeds up the design process, rendering the supply chain more efficient and reducing costs.

What other data-sharing issues may need addressing in the future?

The case study suggests that improvements are being sought in two areas.

- Improvement to data foundations:
 - Through the establishment of data standards: how best to define and record data provenance; identifying the data formats required to create the right level of information that is appropriate to the process.
- Improve incentives for data sharing:
 - Through investigation and development of commercial frameworks: the case study suggests that the use of shared databases potentially transforms the commercial framework, since the sharing of information in this way is not compatible with current transaction-based processes.

B.3 Many-to-many case study 1: OneTRANSPORT

OneTRANSPORT is a data marketplace that aims to enable organisations to share and use data to promote “intelligent mobility”. One of the platform’s intended uses is to support access to data from and by local authorities, particularly those that cannot afford to set up their own data sharing initiatives. To that end, a project funded by Innovate UK helped four local authorities to share data on the platform, in return for visualisation services and access to other local authorities’ data.

What data-sharing issues does this project aim to address?

The OneTRANSPORT project aimed to help address the following issues.

- Reducing costs of data sharing:
 - Through investment in data foundations: improving the findability, accessibility and interoperability of data by providing a common platform using an international standard; and
 - By reducing the number of data-sharing agreements. The platform acts as a hub, whereby local authorities (and other data providers) have a single data-sharing agreement with oneTRANSPORT rather than separate agreements with all potential users of the data.
- Improving and demonstrating the benefits of data sharing, for example providing access to data from several local authorities. The aim is to achieve a

critical mass of large-scale data to design data-driven intelligent mobility services.

What other data-sharing issues may need addressing in the future?

The case study suggests that there is potential to add data to the platform from other data owners, such as public transport authorities, logistics companies and private car parks. However, at least some of these data owners may have limited incentives to do so. They may not be aware of the opportunity, they may require further evidence of the platform's benefits, or they may be concerned about sharing sensitive data of a commercial and/or personal nature. These challenges may need to be tackled in order to keep improving data sharing via this platform.

B.4 Many-to-many case study 2: CityVerve

CityVerve Manchester was a £10m project funded by DCMS and Innovate UK that ran between July 2016 and July 2018. CityVerve was linked to the broader Internet of Things UK Programme (IoT UK), which seeks to promote development of the IoT in the UK. CityVerve aimed to provide a “platform of platforms” allowing the linking of data between distinct local data-sharing initiatives. One example: researchers collected data on people experiencing chronic obstructive pulmonary disease and combined it with environmental data, e.g. on the weather.

What data-sharing issues does this project aim to address?

The CityVerve project aimed to reduce the cost of data sharing by providing infrastructure for sharing between platforms at local level. The purpose was to demonstrate the feasibility and benefits of data sharing to other cities working on smart city projects and services.

What other data-sharing issues may need addressing in the future?

According to the case study, some relevant aspects (e.g. legal agreements) were being developed over the course of the project. The success of this ecosystem also appeared to rely on the development of underlying platforms.

B.5 Many-to-one case study 1: HiLo Maritime Risk Management

HiLo is a joint initiative in the maritime industry aiming to improve risk modelling in this sector. Participating shipping companies share data on accidents and safety incidents with HiLo. HiLo analyses the data and produces reports for shipping companies that highlight signs of possible incidents and recommend preventative action. Data sharing enables benchmarking between different companies and the development of risk models based on more comprehensive and more informative datasets.

What data-sharing issues does this project aim to address?

HiLo addresses some of the risks of data sharing and reduces the cost of sharing through investment in data foundations:

- Thanks to HiLo, shipping companies can share data without worrying about revealing commercially sensitive information to competitors and/or about the possibility that information sharing may be anticompetitive.

HiLo has determined what data needs to be shared and in what format. It also provides the requisite infrastructure (a portal or an Application Programming Interface). More generally, HiLo is a mechanism to coordinate sharing between many organisations. This is particularly beneficial since the data being shared concerns high-impact, low-frequency events of which any one company, by definition, has limited experience.

According to the case study, the project drew on previous successes in the rail and aviation sectors. This suggests that there may be a demonstration effect for similar initiatives in other industries.

What other data-sharing issues may need addressing in the future?

Our reading of this case studies has not identified any specific outstanding data-sharing issues in the ecosystem. It is possible that greater awareness of the initiative and of its benefits could boost participation, but we are not aware of any evidence that this is the case.

This is an example of a data-sharing ecosystem that, to the best of our knowledge, has developed without the help of a specific policy intervention or government funding. Factors that may be behind the success of the ecosystem include:

- Data sharing in this ecosystem aims to mitigate a tangible high-cost problem (safety incidents in the maritime sector). Therefore, the intended benefits of the data sharing are clear and their “value at stake” could be quantified.
- The cost of safety incidents is largely borne by individual shipping companies, so the benefits of data sharing are clearly allocated. However, there may also be negative externalities, e.g. shipping companies may not bear the full cost of ocean cargo spills.
- The organisations that own the data are those that bear the costs of safety incidents and can take action to prevent them. That means there is generally no need for coordination between organisations in acting upon the insights gleaned from the data.

In other ecosystems, these factors may not be present and therefore the private sector may not deliver a similar level of data sharing. For example, there may be even larger externalities, limiting the private incentives to share data relative to the social benefit; or it may be necessary for several organisations to band together to act upon the insights generated by the data, in which case coordination failures may arise.

In such ecosystems, a range of interventions may be helpful to support data sharing.

B.6 Many-to-one case study 2: The Weather Company

The Weather Company (TWC) is a data-centric company owned by IBM that provides insights to aid decision-making in sectors where weather has an impact (e.g. transport, oil and gas, agriculture, etc.). TWC includes two business models:

- A B2C business based on advertising which uses a global data platform to send weather information to smart devices through various free apps (e.g. weather apps, storm Radar, systems built into cars). The platform absorbs very large amounts of data from personal weather stations, satellites and national weather infrastructure. This is cleaned and blended using a combination of weather prediction algorithms to produce targeted weather forecasts for various sectors.
- A B2B business that provides organisations with blended weather data and related insights.

Our focus in this case study is to explore the first step of the data-sharing process, whereby TWC receives data from a variety of sources.

What data-sharing issues does this project aim to address?

The Weather Company aims to help address the following issues.

- Reducing the risks of data sharing:
 - Through investment in security measures: dedicated experts monitor attempted data breaches and conduct penetration testing; and
 - Standardised T&Cs and quality measures have enhanced trust. Furthermore, a two-way agreement is required before data sharing occurs, with obligations and permissions on both sides.
- Reducing the costs of data sharing:
 - The platform is able to process large volumes of data and is resilient to spikes in data sharing, e.g. when major weather events lead to high demand for TWC's services.
- Demonstrate the benefits of data sharing:
 - TWC has helped customers to explore the monetisation of their data and to determine in advance the value of that data by means of pilot projects.
- Improvement to data foundations:
 - For commonly used data fields, such as location, standards are in place. However, progress in this area is hampered by the fact that a lot of the relevant commercial data is proprietary.

What other data sharing issues may need addressing in the future?

The case study highlights a number of ongoing challenges to data sharing:

- GDPR will require a lot of new infrastructure.
- Business leaders do not always have a mature understanding of the value of data they hold.

- Different industries have different views on the value of data and the conditions under which data should be collected and used. Healthcare has its own regulations.
- Establishing ownership rights on derivative work and understanding the chain of custody is important to facilitate data sharing.

