Research into the cost considerations of data sharing

Final report

London Economics

June 2022

About London Economics

London Economics is one of Europe's leading specialist economics and policy consultancies and has its head office in London.

We advise clients in both the public and private sectors on economic and financial analysis, policy development and evaluation, business strategy, and regulatory and competition policy. Our consultants are highly-qualified economists with experience in applying a wide variety of analytical techniques to assist our work, including cost-benefit analysis, multi-criteria analysis, policy simulation, scenario building, statistical analysis and mathematical modelling. We are also experienced in using a wide range of data collection techniques including literature reviews, survey questionnaires, interviews and focus groups.

Head Office: Somerset House, New Wing, Strand, London, WC2R 1LA, United Kingdom.

w: londoneconomics.co.uk　　e: info@londoneconomics.co.uk　　🐦: @LondonEconomics
t: +44 (0)20 3701 7700　　f: +44 (0)20 3701 7701

Authors

Moritz Godel Divisional Director

Ryan Perkins Senior Economic Consultant

Clio von Petersdorff Economic Consultant

Executive Summary

The Department for Digital, Culture, Media and Sport (DCMS) commissioned London Economics (LE) to produce research into the costs facing businesses and organisations that share data. Mission 1 of the National Data Strategy is focussed on unlocking the value of data across the economy by creating an environment where data is appropriately usable, accessible and available. Understanding how organisations consider the costs of data sharing is integral to supporting this Mission and the wider National Data Strategy. This report presents LE's research findings on cost considerations of data sharing, including a taxonomy detailing the related types and categories of costs, as well as evidence substantiating the scale and size of these costs.

This research aims to answer the following questions:

1) What types of costs accrue to organisations at each stage of the data lifecycle?
2) What factors affect these costs, and the organisational perceptions of these?
3) What are sensible upper and lower £ estimates for these costs?

These research questions were answered by conducting a literature review, discussing findings with experts, and conducting a survey of IT decision makers from 500 private UK organisations, as well as eleven in- depth qualitative interviews.

Where possible, these costs are placed along the Government Data Quality Framework's 'Data Lifecycle', a lifecycle used as a way of describing the different stages that data will go through, from collection to dissemination[1].

Data lifecycle
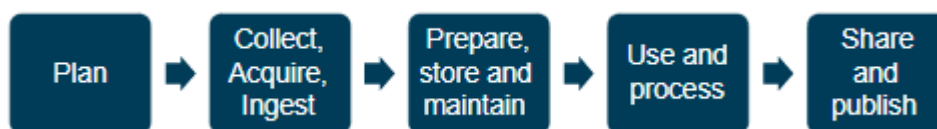
Source: *Government data quality hub (2020) Government data quality framework*

The five cost categories of data sharing identified were

■ **Strategy and planning costs**, such as identifying data sources, planning data sharing strategy, validating the business case, and liaising with internal and external stakeholders.

■ **Legal and regulatory costs**, which include the legal costs for setting up data sharing agreements and the time costs of ensuring that data architecture is secure and compliant.

■ **Technical and organisational costs**, including setting up sharing infrastructure, and the expertise needed to clean raw data and ensure that it is usable for external parties.

■ **Operational costs**, which refer to the costs that are incurred after the data sharing process has begun, such as ongoing data quality monitoring and maintenance of the sharing interface and architecture.

---

[1] For the purposes of this analysis, we have removed the sixth stage - 'Archive and destroy' - as it lies beyond the scope of this research.

■   Implicit costs, which are discussed in the economic literature on data sharing, but do not necessarily appear explicitly on the balance sheets of organisations. These include costs relating to:

  □ a loss of competitive advantage,

  □ reputational costs, and

  □ the cost of cyber security breaches.

**Factors that influence the size of data sharing costs**

Based on the combined evidence from the literature, expert interviews, and stakeholder consultations, the size of data sharing costs is context dependent – each unique dataset is expected to cost a different amount. The factors affecting these costs include the **sector in which the firm operates**, how data is shared (**sharing mechanisms**), and the **size of the firm**.

Businesses deal with large quantities of data, or with sensitive data may require additional spending on security measures. The sharing mechanisms, such as the type of interface required for sharing, and frequency of data sharing will also contribute to the costs. For example, data shared as a once off will involve lower costs than if it is shared in real-time. The size of the firm may impact the amount of data that is shared, as well as the number of employees with data expertise. Finally, data sharing costs will vary depending on the number of parties with whom data is shared. Sharing data openly will cost more than transferring data to one data sharing partner.

Estimates of spend of organisations on data sharing

Based on the survey analysis, **the (weighted) average cost of all data sharing activities across all respondents is around £60,000**. Costs also differ across sizes of firms, with the **highest nominal costs of data sharing incurred by large firms**. Large firms may have higher costs of data sharing due to higher numbers of staff, software licences and potentially, higher complexity of legal agreements. Micro businesses potentially have a higher cost of data sharing compared to Small Businesses due to proportionately higher fixed costs, as well as a need to hire external consultants to fulfil skill and knowledge requirements not held by staff.

Average cost of data sharing among organisations sharing data by size of firm
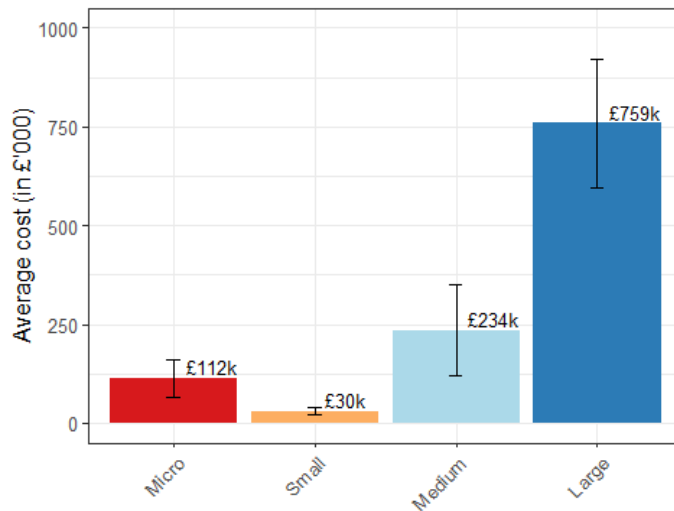
Note: Respondents could choose from a choice of ranges, the midpoint of these ranges was taken to produce the chart above. The largest option (£5m+) was transformed to 7.5m – therefore, these charts are potentially a conservative estimate. Error bars represent standard error around the mean.

Source: LE analysis of survey data

From the survey's five most represented sectors (Information and Communication; Financial and insurance; Wholesale and retail trade; Professional, scientific and technical; and Manufacturing), **the (weighted) average cost of data sharing ranges from roughly £21,000 to £79,000. The Information and Communication, and Financial and Insurance sectors have slightly higher costs** than the other three sectors within the band of represented sectors. Although these are sectors in which higher levels of data sharing and data activities may be expected[2], the average costs across the top five sectors are not significantly different from one another.

Across the five cost categories above, the **highest reported cost was reputational (with a weighted average cost of around ~£50,000)**, the lowest costs were Technical and Organisational (with a weighted average cost of around **~£15,000**). Reputational costs, as an **implicit** cost, are harder to accurately estimate compared to direct costs. Firms may easily recall incidents that have had major reputational costs (such as high-profile scandals or data breaches), and therefore think that reputational costs of data sharing are particularly high. This may have led to respondents over-estimating reputational costs.

The largest proportion of costs across all respondents fell within the sharing stage of the data lifecycle, with the smallest proportion of costs falling in the planning stage. This proportion across the lifecycle is similar for firms of different sizes, however, smaller firms tend to incur a slightly higher proportion of costs in the sharing stage and lower proportion in the collecting stage when compared to larger firms. Large firms experience economies of scale if they spend more on data infrastructure generally (and thereby have more well-developed data infrastructure), which could result in lower sharing costs. Conversely, large firms experience higher costs in the planning stage.

---

[2] The Information and Communication sector includes activities such as publishing, including software publishing, telecommunications, and information technology. Due to the nature of these activities, organisations within this sector may inherently be involved in data sharing, which may mean that these respondents have a greater awareness of their costs. Financial and insurance activities often involve the use and sharing of sensitive financial datasets which may explain higher average sharing costs.

Conclusion and areas for future research

Areas for future research include examining the differences between personal and non-personal, as well as sensitive and non-sensitive, data sharing. This research has found that sharing of personal data has its own unique challenges and costs, especially in terms of regulatory and legal compliance. If the intention is to encourage further data sharing, research into understanding the motivations behind data sharing would be useful to ensure the success of any future policy interventions.

The expert interviews indicated that organisations making the decision to share data usually approach the process by establishing a business case, or by determining a clear benefit or value from this data sharing; therefore, organisations often do not look at data sharing primarily through the lens of cost consideration. If the business case can be justified, costs are unlikely to be sufficiently high to be an impediment preventing data sharing. The main problems preventing data sharing include establishing a business case, structuring data sharing relationships, and regulatory uncertainty.

Research into the mechanisms of sharing (such as bilaterally, multilaterally, or as open data) would also be useful to understand how costs and motivations for sharing data may differ depending on the mechanics of the data sharing itself.

This report is an exploratory step into quantifying the costs of data sharing. The taxonomy serves as a verified classification of costs for future research; an understanding of different types of costs will also be useful for any potential policy interventions used to reduce the barriers and cost of data sharing. The quantitative results in this report provide an overview of the relative magnitudes of data sharing costs, as well as where along the data lifecycle these costs are incurred.

Table of content

Introduction

The Department for Digital, Culture, Media and Sport (DCMS) commissioned London Economics (LE) to produce research into the costs facing businesses and organisations that share data. Mission 1 of the National Data Strategy is focused on unlocking the value of data across the economy by creating an environment where data is appropriately usable, accessible, and available. Understanding how organisations consider the costs of data sharing is integral to supporting this mission and the wider National Data Strategy. This report presents LE's research findings on the cost considerations of data sharing, including a taxonomy detailing the related types and categories of costs, as well as estimates of the potential scale and size of these costs to firms. Where possible, these costs are placed along the Government Data Quality Framework's 'Data Lifecycle', a lifecycle used as a way of describing the different stages that data will go through, from collection to dissemination.

Figure 1   Data lifecycle
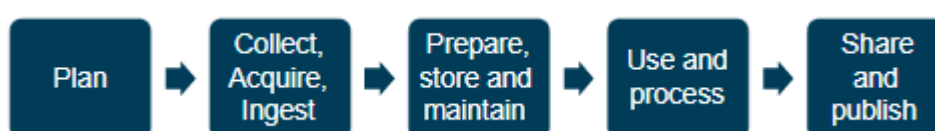
- **Plan** – This stage is where an organisation (intending to collect, store and use data) must plan their processes and data storage. It involves determining business needs, identifying existing data and what needs to be collected.
- **Collect, acquire, ingest** – This stage involves the acquisition and collection of data.
- **Prepare, store, and maintain** – This stage involves the preparation of data, including formatting data for later use, and storing and maintaining databases appropriately.
- **Use and process** – This stage involves the use of data for an organisation's business needs. This could include analysis of data or production of outputs.
- **Share and publish** – This stage involves preparing data for publication and sharing, as well as sharing data with users outside your organisation (whether this data is shared/ sold for monetary gain or otherwise).

Note: This lifecycle also has been adapted to remove the sixth stage 'Archive and destroy' due to this stage lying outside the research scope

Source: Government data quality hub (2020) Government data quality framework

This research lies within the context of DCMS's wider research focused on data sharing. This report is positioned to inform future policy interventions and research targeted at increasing the accessibility of data.

Research outline and methodology

This research aims to answer the following questions:

1)   What types of costs accrue to organisations at each stage of the data lifecycle?
2)   What factors affect these costs, and the organisational perceptions of these?
3)   What are sensible upper and lower £ estimates for these costs?

These questions were targeted with the following research strands:

1)   **Systematic literature review**: this review targeted existing literature and studies relating to costs of data sharing. This review was used as the basis of the taxonomy of costs.

2)   **Expert interviews**: interviews with data sharing experts guided the literature review, as well as substantiating and confirming the findings and conclusions drawn.

3)   **Survey**: an online survey with 500 IT decision makers conducted with the aim of gathering quantitative data for the taxonomy's identified costs.

4)   **Stakeholder interviews**: interviews with 9 organisations that share data to gather additional data on the cost of data sharing.

5)   **Synthesis and quantification**: synthesis of previous strands to obtain a range of quantitative estimates and qualitative descriptors of the factors affecting these costs.

6)   **Reporting and presentation**: results and analysis are presented in this report, as well as in an accompanying data dashboard.

Figure 2   Overview of Methodology
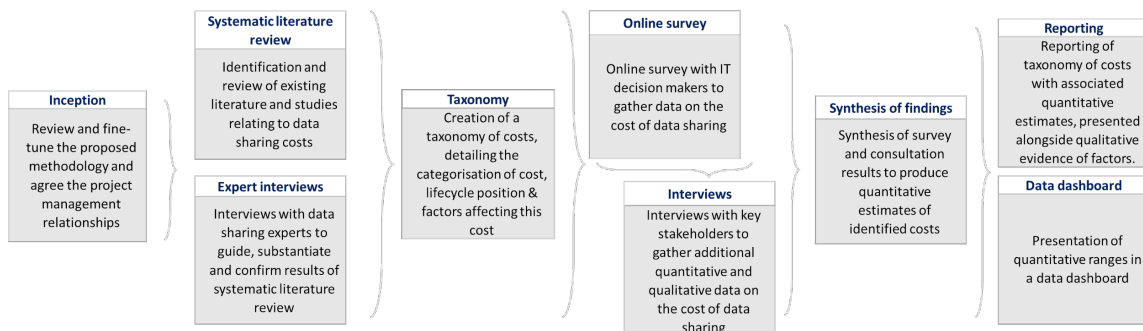
These research strands produced the following key research outputs:

1. A **taxonomy of costs** detailing the categorisation of each cost, the position along the data lifecycle and factors affecting this cost. A spreadsheet detailing the taxonomy in further detail – along with a bibliography of literature that was used to create this taxonomy – is provided alongside this report.
2. This report, which contains
   a. A **range of quantitative estimates** for each of these costs (where possible), presented within this report and an interactive data dashboard.
   b. **Additional qualitative findings** substantiating these costs.

3.1 Literature review methodology

We used a **systematic literature review approach** as part of the literature identification, review and subsequent taxonomy creation. This process involved cataloguing policy reports, academic papers and findings published by data sharing experts, reviewing these reports, and detailing the findings from each source in a spreadsheet which was provided alongside this report. These findings were then further analysed and synthesised with expert interviews to form a taxonomy of data sharing costs. More information on the steps involved and the structure of the spreadsheet can be found in Annex 5.

3.2 Expert interviews

As part of the taxonomy creation process, four in-depth interviews were conducted with the following data sharing experts:

Table 1 Details of expert interviews

Contact	Organisation	Role
Olivier Thereaux	Open Data Institute	Head of Research and Development
Aditya Singh	Open Data Institute	Researcher for the Data Institutions programme
Zee Kin Yeong	Personal Data Protection Commission Singapore	Deputy Commissioner
Elena Simperl	King's College London	Professor of Computer Science

These interviews aimed to substantiate our understanding of data sharing costs and were used to direct and verify our thinking throughout the process. Additionally, **these interviews also provided valuable feedback on our taxonomy, confirming the costs identified in our research**.

3.3 Survey

An online survey was administered by IRBureau to a panel of IT decision makers[3] at 500 organisations across several sectors. The aim of this survey was to gather data to create a range of quantitative estimates for each of the costs identified in the taxonomy.

The survey can be found in Annex 1.

3.3.1 Overview of survey questions

The aim of the survey was to provide insights into the costs and barriers that organisations face when sharing data. Respondents were first asked to describe some of the characteristics of their firm, such as its size, the sector(s) in which it operates, and where it is headquartered. Respondents were then asked whether their firm collects digital data, whether the data that is collected could be valuable to external users, and whether it is shared externally.

---

[3] IT decision makers refer to people in leadership positions responsible for making decisions about their organisation's IT activities – such as data sharing.

The survey included questions on the overall costs that firms face when they collect, hold, and use digital data. These costs include the number of employees that are employed to process data and the cost of the data infrastructure required, such as data storage, security, and software costs. Respondents whose firms did not share data were asked to provide details about which types of costs may present barriers to sharing. Respondents that work in firms that share data provided estimates of the data sharing costs that were identified in the taxonomy and attributed those costs across the data lifecycle.

Respondents in this survey could provide cost figures by selecting from a list of cost categories.[4] The midpoints of each of these categories were taken to generate further analysis and statistics. As a result, the estimates provided are conservative in value due to the potentially limitless cost for respondents in this final cost category.

3.3.2   Sample targeting overview

The sample consists of 500 individuals who completed the online survey about the costs of data sharing in their firms. The survey was distributed by IRB to their panel, **targeting individuals from UK private organisations with job titles which suggest responsibility over their organisation's data use and sharing activities**. Targeting was necessary to ensure that the survey was completed by respondents able to successfully answer the data sharing questions. This approach may have led to selection bias – organisations that employ individuals dedicated to IT activities may be overrepresented in specific sectors and among large firms.

This targeting led to two fifths (190) of individuals completing the survey having an 'IT decision maker' role, and another fifth (96) being general managers.

3.3.3   Overview of achieved sample

The sample of businesses captured through this survey appears skewed in some ways:

- Business size: Large firms comprise more than two fifths of the sample, compared to less than 1% of the business population.
- Business sector: All major types of business sector are represented in the sample, but there is bias towards businesses operating in Information and Communication (23% in the sample compared to 6% in the business population).
- Business location: Businesses from all regions responded to the survey, but the sample is skewed in a number of ways (e.g. 28% of the sample is in the London region, compared to 19% of businesses)
- Data sharing activity: businesses which share data are likely to have been over sampled in this study. While methodologies vary, the UK Business Data Survey 2022 indicates 48% of the business population share data, compared to 55% in this sample; and 28% of small businesses in the business population share data, compared to 48% in the sample.

This skew in respondents is likely to impact how applicable the findings of this survey are to the general business population.

---

[4] These categories were: (i) Zero, (ii) Less than £1,000, (iii) £1,000- £4,999 (iv) £5,000- £9,999, (v) £10,000- £49,999, (vi) £50,000- £99,999, (vii) £100,000- £499,999, (viii) £500,000- £999,999, (ix) £1m- £4,999,999, (x) More than £5m.

For the headline summary figures quoted in this report, findings have been weighted to account for the skew to large business sizes found in the sample. Further weighting, to account for the other biases noted above, wasn't feasible.[5] All findings should be treated as a summary of the sample, rather than of the general business population. Research on a larger scale is needed to draw more generalisable conclusions.

The detailed overview of the achieved sample is presented in Annex 2.

3.4   Interviews

Interviews were conducted to substantiate the taxonomy of costs with additional qualitative and quantitative data by OMB Research. Several sources were targeted in efforts to recruit respondents for the interviews, and interviewees were offered a £50-£100 incentive. The sample of organisations interviewed included organisations from different sectors, of varying sizes which shared different types of data.

Organisations interviewed were sampled from two sources:

1. Organisations identified from previous data-focused research. These organisations had previously participated in the interviews related to data sharing – these respondents were confirmed to have an awareness and familiarity with data use and data sharing issues.
2. Survey respondents who consented for a follow-on interview after survey completion.

Participants recruited from the survey were also recruited based on their ability to provide quantitative cost figures in their responses. Eleven interviews were completed. Details of the interview recruitment process, panel and the interview guide can be found in Annex 3.

---

[5] The number of respondents in each subcategory (sector, geographical location, data sharing behaviour etc) becomes very small as the overall sample is 500. This in turn would make the error margins too large to draw meaningful conclusions from the data.

Evidence review and taxonomy

This section details the findings of the literature review and expert interviews in relation to market failures in the market for data and barriers to data sharing (section 3.1) the role of the data lifecycle in data sharing considerations (section 3.2), and the taxonomy of data sharing costs (section 3.3).

4.1 Key findings of market failures and barriers to data sharing

Many of the barriers to data sharing for firms stem from market failures in the market for data. Market failures occur when goods and services are inefficiently distributed in the free market and can lead to over- or under consumption of resources. Market failures that may exist in the market for data:

- **Asymmetric information** between organisations, or between an organisation and its consumers can lead to worse outcomes for the parties with less information (Ctrl-Shift, 2018). For example, where the value of data is better understood by data driven businesses than by "analogue" businesses, the data driven businesses can prevent the others from accessing and creating value from the market opportunity (Clarkson, Jacobsen and Batcheller, 2007). Information asymmetries can also exist between businesses and the consumers from which data are obtained. Consumers may be unaware of how much data is being gathered, who it is being shared with and what it is being used for, this may lead to outcomes that consumers would prefer to avoid.

- **Principal-agent problems** occur in data sharing when the incentives of the organisations that collect data, the third parties who access the data and the consumers from whom data is gathered are not in alignment (Coyle et al, 2020). Related to the principal-agent problem, frictions arise from 'incomplete contracts' when the agreements between data sharers and data users don't specify how data can be used in every possible contingency (London Economics, 2019). This presents a hold up for firms wishing to use data as well as for firms that wish to share it.

- **Coordination failures** occur where market participants fail to align on issues, for example on data standards, and are therefore unable to share data (Ctrl Shift, 2018). Coordination failures can also occur when data holders are unaware of other parties who may be interested in obtaining data, or when businesses who are interested in using data do not have information about who collects it.

- **Externalities** may arise through data sharing[6] (Coyle et al, 2020; Ctrl Shift, 2018). Positive externalities can arise from the social or environmental benefits that occur due to the additional insights or improved decision making when datasets are linked together or shared. Negative externalities arise when harms come about through data being collected or used, or through the amount of data being collected and stored.
  - ☐ When data sharing leads to **positive externalities, society loses out when** there is not enough data sharing (Coyle et al, 2020).
  - ☐ When **negative externalities** arise, there is a chance that organisations will share too much data, as the costs of the harms generated through sharing are not felt by the sharing organisation.
  - ☐ In either case the market may not provide the best overall outcomes.

---

[6] In economics, externalities refer to the consequences of an industrial or commercial activity that are not reflected in market price.

4.1.1   Other barriers to data sharing

In addition to the market failures described above, there may be other barriers to data sharing that are not directly related to the economic characteristics of data.

Firms often do not see a **clear business case for data sharing** (Ctrl-Shift, 2018), because the returns to sharing data for the firm are uncertain (Frontier Economics, 2020).

For smaller organisations, the preconditions needed for data sharing, such as the hardware, software and skills required, are often lacking (London Economics, 2019). In these companies and ones that do not rely heavily on data, it may be more difficult to use and share data for value creation (London Economics, 2019). In particular, the **lack of technical knowledge** about the uses of data or what data is available may hamper data sharing (London Economics, 2019; Frontier Economics, 2021).

The costs of data sharing are not linear throughout the sharing journey. There are **high upfront costs** to data sharing, which may deter companies if there is no clear business case for sharing (Frontier Economics, 2021). These costs may be related to the technical set-up costs, to the internal coordination required such as identifying relevant datasets, the stakeholders involved and negotiating access, which can be difficult and resource-intensive (London Economics, 2020).

Firms experience additional barriers to sharing data when the data relates to individuals. This is both because of the **risks of breaching data protection** (Frontier Economics, 2021), and because of consumers' reactions to their information being shared. Organisations face concerns about privacy from their customers (Acxiom and Foresight factory, 2019), and collecting and sharing their data may lead to a lack of trust and engagement from customers (Data and Marketing Association, and Acxiom and Foresight Factory, 2018).

4.1.2   Evidence about market failures from expert interviews

The expert interviews, conducted with four data sharing experts, revealed information on additional barriers and costs facing businesses who wish to share data. Due to the low number of interviews, these findings are not intended to represent the views or experiences of all companies, but to corroborate the points found in the literature. As such, the findings from the expert interviews should be considered holistically and together with other evidence and findings. Where possible, these findings are signposted with relevant literature supporting these.

Respondents in the expert interviews commented that the barriers preventing data sharing include:

- Lack of digital literacy[7]
- Lack of awareness of the value of data sharing[8]
- Lack of knowledge about strategic reasons for data sharing[9]
- Perception that data sharing lowers competitive advantage[10]

---

[7] Problems related to lack of hardware, software and skills detailed in London Economics, 2019; Frontier Economics, 2021

[8] Coordination failures detailed in Ctrl Shift, 2018

[9] Asymmetric information failures detailed in Ctrl Shift, 2018. Lack of knowledge of reasons to share data, as well as demand for data can prevent organisations from sharing this data

[10] Fear of loss of competitive advantage detailed in London Economics, 2019

■   Concern over the risks of data breaches[11]

The expert interviews indicated that organisations making the decision to share data usually approach the process by establishing a business case, or by determining a clear benefit or value from this data sharing; therefore, organisations often do not look at data sharing primarily through the lens of cost consideration. If the business case can be justified, costs are unlikely to be sufficiently high to be an impediment preventing data sharing. The main problems preventing data sharing include establishing a business case, structuring data sharing relationships, and regulatory uncertainty;[12] regulatory uncertainty and failure to establish data sharing relationships between market participants may be an example of **coordination (market) failures**. Concern about misuse of data may also be evidence of **principal-agent problems** occurring from 'incomplete contracts', as well as an awareness of potential negative **externalities** arising from misuse and data-breaches.

Organisations choosing to share data do so for the following reasons:

■   Organisational responsibility, sharing data to help improve innovation within the sector, or to gain trust.

■   Commercial, transactional, or financial reasons.

■   Required due to regulation and licence conditions.

Costs which face businesses data sharing include:[13]

■   Legal and regulatory costs such as the time costs of compliance and ensuring data architecture is secure and compliant.[14]

■   Technical costs, such as setting up infrastructure, ensuring data is clean and ensuring that this data makes sense for users outside the organisation.[15]

■   Strategic and planning costs such as identifying data sources, planning data sharing strategy and validating business cases.[16]

■   Operational costs such as ensuring ongoing data quality and maintenance of sharing interface and architecture.[17]

According to the experts that were interviewed, data sharing agreements often fail to materialise due to the sharer having low awareness of the full range of costs involved with data sharing. Organisations with lower levels of digital literacy, skills, processes, and infrastructure often agree to share data, but then pull-out of agreements after realising the full scope of costs. For example, data collected for internal usage may require additional licensing agreements if this data is to be shared or transformed for commercial use. This **information asymmetry** between data sharing participants may be a cause for these agreements – and subsequent data sharing – failing to come to fruition.

---

[11] Cost of cyber security breaches and reputation costs detailed in Acxiom and Foresight factory, 2019; Frontier Economics, 2021

[12] Corroborated in Ctrl-Shift, 2018; Frontier Economics, 2020

[13] The costs below reflect the taxonomy classifications, further sources and details on the underlying literature can be found in the project taxonomy spreadsheet

[14] Related costs detailed in Coyle et al, 2020; McKinsey, 2013

[15] Related costs detailed in OECD, 2019; OECD, 2020; Bald, 2020; Figueiredo, 2017

[16] Related costs detailed in London Economics, 2019; Frontier Economics, 2020

[17] Related costs detailed in Figueiredo, 2017

Factors which affect these costs include:

- The size of the company – smaller businesses will often not have the skill set in-house necessary to set up and maintain data sharing infrastructure.
- Size and complexity of the data shared – including characteristics such as whether the data is personal.
- How 'digital' a company is, a company that is tech savvy – with a lot of data-related skills – will not struggle with much of the technical costs related to data sharing.
- Nature of access of this data, such as full or partial access, or a necessity for real time data.
- How this data is shared – data shared through an API may be expensive.

Raising the awareness of the value of data sharing – along with testimonials – would help increase management buy-in and increase awareness of the associated costs of data sharing. The lack of awareness of costs is also sometimes a barrier preventing data sharing – this low awareness affects businesses effectively forming the business case to share data. These factors are an example of continued **information asymmetry** market failures – a lack of awareness of the value of data between data controllers and users which can attain value from this data hampers the rate of data sharing.

4.2   Data lifecycle

All data used for productive purposes will go through a similar journey through organisations (Government data quality hub (2020)). This journey is described by the data lifecycle. While the actual lifecycle of data will be specific to an organisation (and may differ across sector/ industry), the framework is a useful tool to identify points where data quality issues may arise.

The first stage in the lifecycle is the **planning stage**, which involves determining business needs, identifying existing datasets and what needs to be collected. During the second stage, data is **collected or acquired, and ingested**. Data then needs to be **prepared, stored and maintained** so that it can be useful to the organisation. During this stage it is important that consistent standards are applied and that data is cleaned and linked with other records where applicable. The fourth stage of the data lifecycle is related to the **use and processing** of the data, which includes exploration and analysis, and production of outputs. Data can also be **shared and published** where appropriate for secondary purposes. Finally, when data is no longer in active use, the data owner should determine whether it should be **archived or destroyed**.

Figure 3   The data lifecycle

Source: Government data quality hub (2020) Government data quality framework

While the data lifecycle can be mapped out into separate stages and presented in a linear form, a single type or category of cost that arises from data sharing is often present in multiple stages. For example, the same software and hardware may be needed to collect data, to prepare and maintain it, as well as to use, process and share it. The licensing costs of software can therefore not be attributed to a specific stage in the data lifecycle.

4.3   Taxonomy of data sharing costs

The Infocomm Media Development Authority (IMDA) has developed a useful framework for data sharing which may be used to group/categorise sharing costs. The framework describes the considerations for firms that wish to share data (IMDA, 2019).[18] We have built on their framework, supplementing it with other findings from the literature review.

The costs of data sharing, as identified in the literature and through expert interviews, have been split into explicit and implicit costs. **Explicit costs** are costs that organisations experience directly, and are likely to be present in an organisation's balance sheet. An overview of these costs can be found in Table 2. The explicit costs have been grouped based loosely on the trusted data sharing framework described above.

Table 2   Overview of the explicit costs in data sharing

Cost	Description
Data sharing strategy	**Part 1 in data sharing framework**
Planning and coordination costs	Cost of identifying data to be shared, and coordinating among various internal and external stakeholders
Data collection costs	Cost involved in collecting data that is to be shared. This applies in cases where additional data must be collected for the purpose of sharing and excludes data collection required for business processes.
Procurement costs	Cost of purchasing data from a third party/vendor.
Legal and Regulatory considerations	**Part 2 in data sharing framework**
Compliance costs - collection	Ensuring that the collection of data to be shared is compliant with regulation.
Compliance costs - sharing	Ensuring that data can be shared, and that it will be shared in a way that adheres to GDPR and other legislation.
Data sharing agreements	Administrative and legal cost of creating contracts between data sharers and data receivers.
Technical and organisational considerations	**Part 3 in data sharing framework**
Software costs	Software that may be required for data collection, cleaning and sharing, including the cost of licences, as well as the cost of developing programmes in-house.
Security costs	Administrative cost of ensuring that the interface used for sharing is secure and the risk to the company is limited.
Storage costs	Cost of storing the data to be shared (on-premises or cloud).
Data preparation costs	Cost of cleaning data, preparing metadata and formatting data to adhere to standards. This includes preparing data for machine learning, such as labelling.
Operationalising data sharing	**Part 4 in data sharing framework**
Data quality monitoring	Ongoing monitoring of data quality.
Data sharing interface maintenance	Ongoing maintenance of interface used for sharing.

Source: London Economics

---

[18] The framework was developed as a planning tool for organisations and has therefore needed to be modified slightly to be appropriate for the cost considerations.

The **implicit costs** of data sharing are those that do not explicitly appear on the balance sheets of organisations.

Table 3   Overview of implicit costs related to data sharing

Cost	Description
Competitive disadvantage	Loss to a business if data leads to a loss in competitive advantage.
Reputational cost	Firms that share data may experience a reputational cost if the nature of the data is sensitive, or if there are other objections from customers/partners/other stakeholders.
Cost of cyber security breaches	Firms may incur high losses if they are subjected to a hacking attack facilitated or enabled by the sharing arrangement.

Source: London Economics

The following paragraphs explore these categories in more detail.

Developing a data sharing strategy

In this part of the data sharing framework, the firms need to evaluate their current and potential data sharing capabilities and establish the value of their data. The firms must also understand the attributes of different data sharing models and engage with various internal stakeholders. The firm can also consider engaging with a data service provider to facilitate data sharing.

The identified costs associated with this part of the framework are costs associated with planning the data sharing process, coordinating between various internal and potentially external stakeholders, and the cost of collecting or procuring the data that will be shared.

Identifying valuable datasets and coordinating with other market participants reduces the potential market failures of **information asymmetries**, **principal-agent problems,** and **coordination failures**. Harmonisation of incentives between parties participating in data sharing resulting from effective strategizing enables data sharing by reducing the barriers associated with these failures.

Legal and regulatory considerations

The legal and regulatory considerations include determining if the data can be shared, and the regulations that need to be adhered to in the sharing process. The size of the costs involved in this stage will depend on the type of data that is being shared. Personal and sensitive data will have higher compliance costs than other types of data due to the additional security measures required.

If data is going to be shared bilaterally, or only with specific individuals/firms, then data sharing agreements will need to be drawn up in this stage as well.

Regulation and laws reduce misuse of data and subsequently reduces potential **negative externalities**. Failing to comply with these considerations may lead to greater risk of these externalities – the importance of satisfying these considerations is twofold: fulfilling legislative requirements and reducing these **externalities**.

Technical and organisational considerations

This part of the data sharing framework is concerned with the technical considerations of data sharing, such as preparing data for sharing, and the broader organisational considerations such as the security of the interface that will be used to share data.

Preparing data and ensuring that data structures are appropriate for sharing can help assuage potential **principal-agent problems.** Sufficiently prepared and secured datasets can help to reduce the risk of data being used for 'unspecified contingencies' as part of 'incomplete contracts', where data sharers and data users fail to specify exactly how data can be used (London Economics, 2019). Reducing these contingencies (and associated potential misuse) can also reduce potential **negative externalities**.

Operationalising data sharing

Once the legal, technical and organisational factors associated with data sharing have been determined, the data can be shared. In this part of the framework, data sharing is operationalised – this stage includes ensuring transparency and accountability, continually monitoring the quality of the data being shared and maintaining the interface used to share.

A focus on continued transparency and accountability of this data use can allow for a reduction of potential **principal-agent problems** – ensuring incentives remain aligned between data sharers and third parties using this data. This transparency and accountability also assuage the impact of **information asymmetries** between these data users and consumers which may be affected by this data use.

Implicit costs of data sharing

Implicit costs refer to external costs which may be incurred on an organisation sharing data. For example, reputational costs, loss of competitive advantage and potential cyber breaches. These costs are not directly incurred on a firm but may be included in an organisation's decision to share data or not.

Negative externalities inflicted on consumers and external parties by an organisation's data sharing activities would result in a potentially higher reputational cost whereas **positive externalities** generated from data sharing may result in a loss of competitive advantage – which may lead to these organisations choosing to share less data.

4.4   Factors affecting costs of data sharing

Data sharing costs can be affected by the characteristics of the data and organisations sharing data. The following characteristics were discussed during the expert interviews:

- **Data type** – different characteristics of data incur different profiles of costs. For example, sharing personal data involves compliance with GDPR and other regulations – this type of data may also incur greater reputational risks in the event of misuse or data breaches.
- **Data use** – the uses of the shared data may also incur different levels of costs. For example, processing data for machine learning would involve different skills and procedures compared to processing data for marketing purposes – and therefore incur different technical and operational costs.
- **Method and period of sharing** – the periodicity and method of sharing data can also affect the cost profile. Data which needs to be shared frequently requires additional operational costs to maintain interface and quality of data. Additionally, the method of sharing and steps taken to make data interoperable may incur different costs. For instance, an organisation would have to use resources to ensure data standards are adhered to to enable easier access to that data by third parties.

- **Firm sector** – the sector of firms may also be an indicator of costs of data sharing. Sectors which involve significant levels of data use and related activities may already have existing infrastructure to support data sharing. Sectors dealing with sensitive data, such as financial services and healthcare may also have higher costs due to the legal and regulatory requirements on sharing and using this data.

- **Firm size** - firm size may also result in different profiles of costs of data sharing. Fixed costs (such as software licences) for smaller firms are disproportionately higher than for larger firms. Different types of costs are greater for firms of different sizes, which would affect a business' decision to share data.

Empirical findings

Quantitative (and qualitative) data on costs were gathered from a survey of 500 UK business respondents as well as 11 in-depth interviews.

While efforts have been made to ensure representation across a number of different business sectors, this sample is not a representative sample of the UK. These results reflect a specific composition of sectors and firm sizes (detailed in Section 2.4.2). The firms in the sample all collect digital data and employ individuals with job titles that suggest their organisation uses and/or shares data. Subsequently, total and average cost figures are more representative of firms that are data-driven. In other words, although these results may not reflect the costs of the "average" UK firm, they are indicative of the costs that an average data-sharing firm may face.

As discussed in section 3.4, data sharing costs can be affected by different characteristics of both the data and organisations sharing data. These factors include the type of data, what the data will be used for, the method and frequency of sharing, as well as characteristics of the firm, such as the sector it operates in and its size. The following empirical findings may therefore be shaped by these factors, and where possible, the assessment of these costs is nuanced with respect to them- consideration of these factors is important when interpreting the results below.

5.1   Data sharing

The majority of respondents (over 85%) agreed that the data their firm collects may be valuable to an external user and were able to identify which data type of firm data would be most valuable (see Figure 5). There were 33 firms (6.6%) that did not consider any of the data their firm collected to be valuable to external users. Overall, larger firms were more likely to consider their data to be valuable to external users.

Figure 4   Number of firms that consider their data to be valuable, by firm size
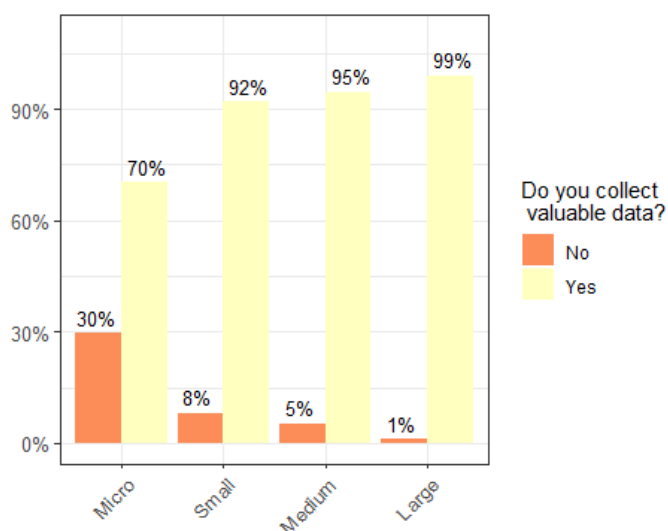
Note: This data is unweighted.

Source: LE analysis of survey data

The most frequently identified type of data that is considered most valuable to external users is contact information, which was identified by 74 firms, followed by financial/accounting data (71 firms), and sales or transaction data (64 firms).

Figure 5   Type of data that is considered most valuable to external users
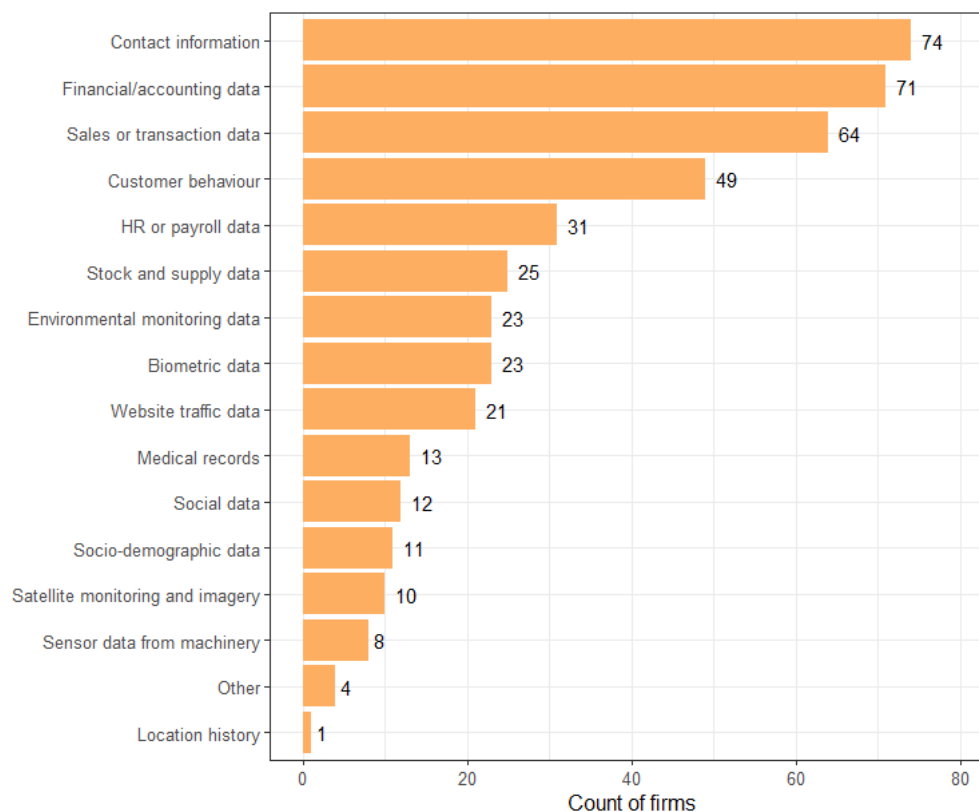
Note: These counts are unweighted and may not represent the UK business population.

Source: LE analysis of survey data

The respondents who were able to identify a valuable type of data were asked whether their firm shares this data externally. **Just over 50% of firms share data, 47% of firms do not share data, and 7 respondents were unsure.** Firms of all sizes were similarly likely to share data with external users. According to the UK Business Data Survey (2022), 28% of small firms, 43% of medium firms and 48% of large firms share either personal or non-personal data with external organisations. The UKDBS results may indicate that the number of data sharing firms in this sample are slightly higher than the population average, particularly for small firms.

Figure 6   Number of firms that share data
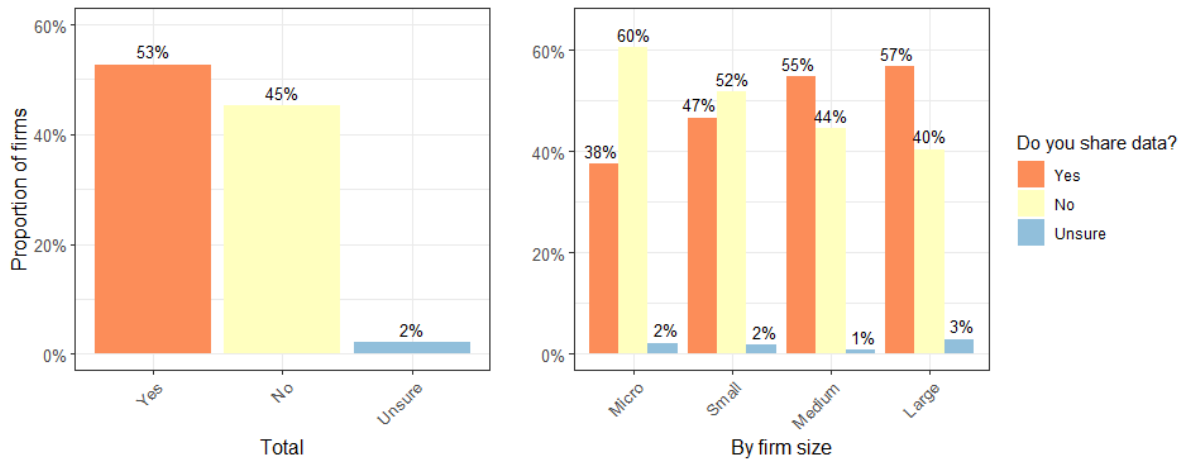
Note: Chart on the left is weighted by firm size. The chart on the right is unweighted.

Source: LE analysis of survey data

Figure 7 shows variation in the answer to whether valuable data is shared split by the type of sector the firm operates in. Based on the results from this sample it does not seem as though the sector in which firms operate contributes significantly to their decision to share data. The number of firms that share data is similar across all sectors in the sample (barring those for which there are only very few firms in our sample).

Figure 7   Propensity to share data by sector
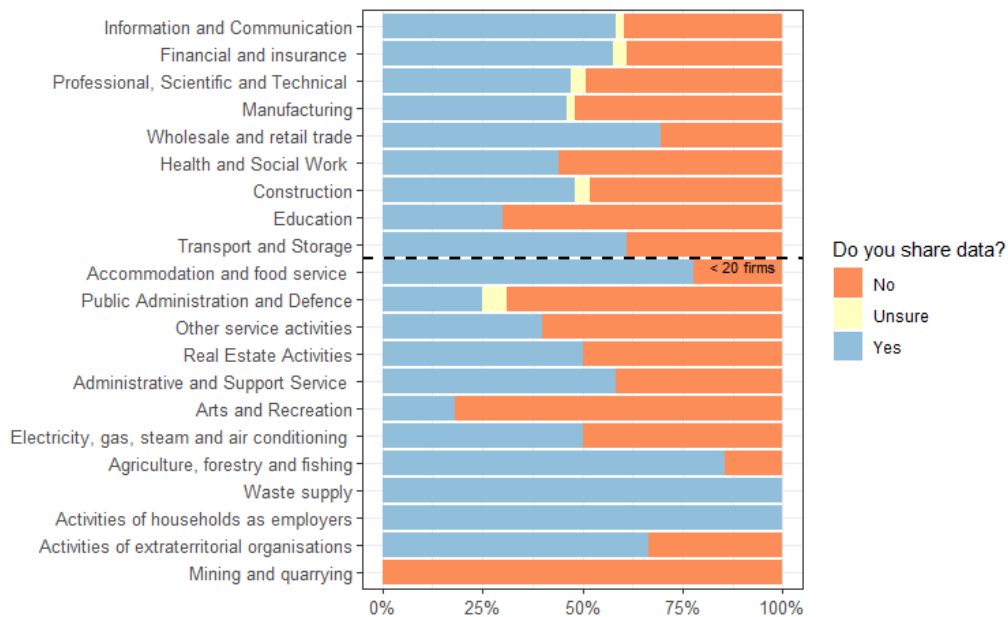
Note: Fewer than 20 firms selected into each sector appearing below the dotted line. This data has been weighted by firm size.

Data can be shared in different formats:

- **cleaned data** is data that is ready for digital processing and analysis and may refer to data which has undergone an intermediary process such as labelling.

- **processed data** is cleaned data with additional analysis/value added,

- **summary data** is data with reduced granularity, and
- **raw data** is in the format in which it is generated/acquired.[19]

According to the survey, all these data formats are fairly commonly used to share data, with 131 firms sharing clean data, 124 sharing processed data, 111 sharing summary data and 78 respondents stating that their firm shares raw data with external parties. In this question firms were able to select all data formats in which they share data.

Figure 8   Format in which data is shared
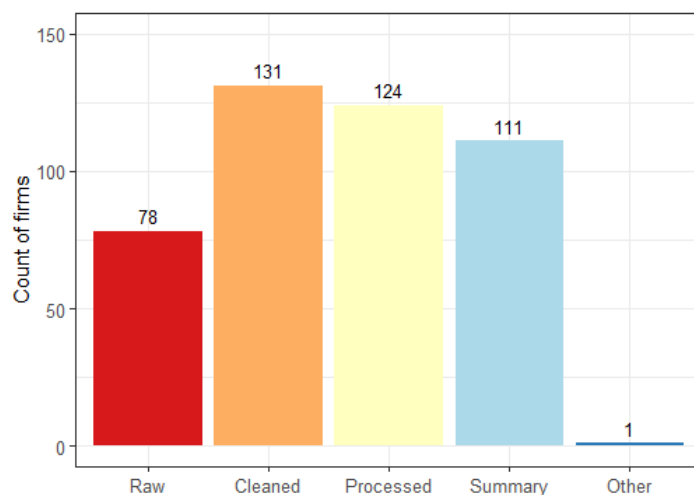
Note: These counts are unweighted, and therefore may not be representative of the population.

Source: LE analysis of survey data

---

[19] Machine learning processes require differing steps of preparation depending on the process and use by the recipient. Data shared for machine learning purposes may therefore be shared in different formats depending on the complexity and maturity of the machine learning process. For example, labelling data involves tagging raw data by adding meaningful label(s) to provide context for this data (AWS, n.d.,); cleaned data

5.2 Total costs / overview

5.2.1 Cost of data infrastructure

The first category of costs of data sharing involves the cost of data infrastructure: security, storage and software costs. **Across all firm sizes, security costs were on average the largest data infrastructure cost.**

Table 4 Summary statistics of infrastructure costs

	Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
Software	0	920	3,068	65,291	22,993	12,985,759	458
Security	0	690	6,898	82,714	22,993	12,985,759	461
Storage	0	920	2,902	34,277	12,986	3,895,728	486

Note: This data has been weighted by firm size

Source: LE analysis of survey data

Figure 9 Total infrastructure costs for organisations sharing data
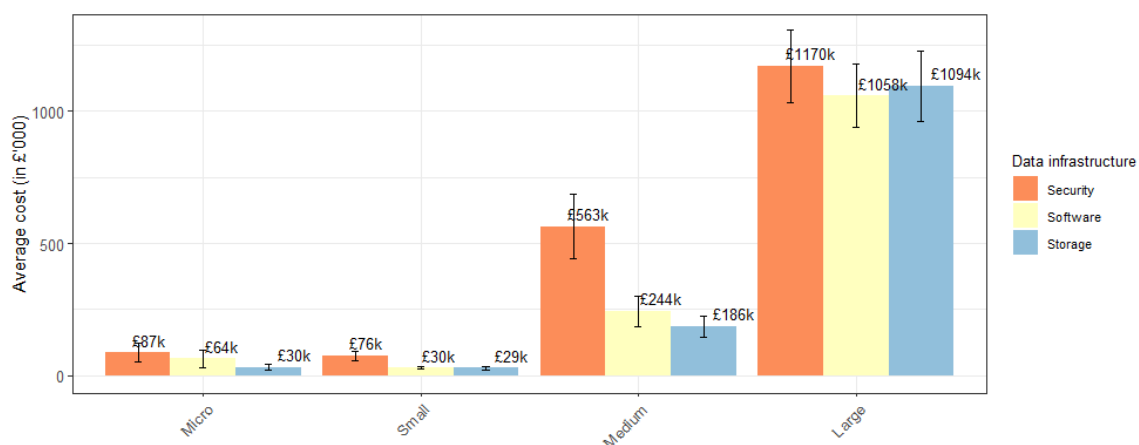
Note: Respondents could choose from a choice of ranges. The midpoints of these ranges were taken to produce the chart above. The largest option (£5m+) was transformed to 7.5m – therefore, these charts are potentially a conservative estimate. The error bars represent the standard errors around the mean.

Source: LE analysis of survey data

There was no statistically significant difference on spend on infrastructure between firms that do and do not share data. Our literature review and findings from interviews suggest that firms that share data need to spend more on software to protect the data and facilitate safe sharing. It was suggested in the expert interviews that software costs may be particularly high if data is shared through interfaces such as APIs. These discrepancies could be explored in further research.

5.2.2 Total cost of sharing data

Firms spend a (weighted) average of almost £57,559 on their data activities that relate to data sharing (but a median of only £3,068), including infrastructure costs, staff costs, legal costs and other related costs used explicitly for sharing data – with large firms spending considerably more than SMEs.

Table 5 **Summary statistics of cost of data sharing**

Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
6	690	3,068	57,559	22,993	3,895,728	221

Note: Data has been weighted by firm size

Source: LE analysis of survey data

Figure 10 **Average cost of data sharing among organisations sharing data by size of firm**
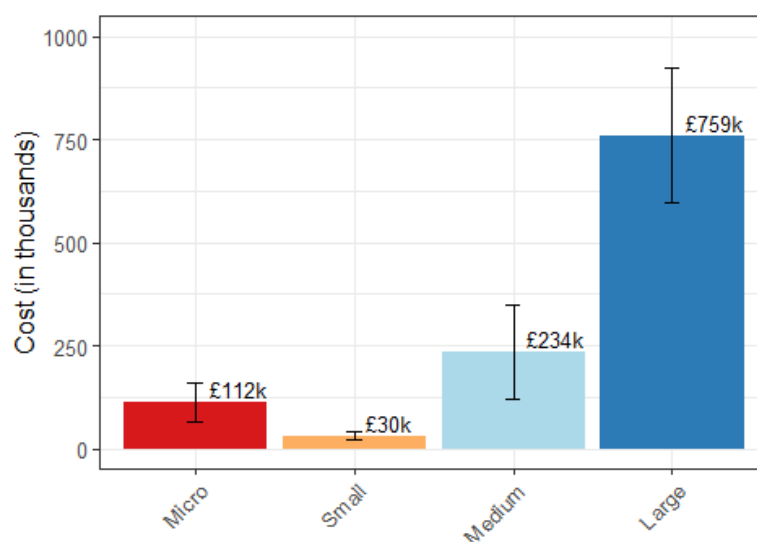
Note: Respondents could choose from a choice of ranges. The midpoints of these ranges were taken to produce the chart above. The largest option (£5m+) was transformed to 7.5m – therefore, these charts are potentially a conservative estimate. The error bars represent the standard errors around the mean.

Source: LE analysis of survey data

Table 6 **Summary statistics of cost of data sharing per firm size**

Size	Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
Micro business	500	7,500	25,000	112,000	118,750	750,000	16
Small SME	500	2,500	7,500	29,900	25,000	250,000	26
Medium SME	500	7,500	25,000	234,000	206,250	7,500,000	66
Large firm	500	25,000	75,000	759,000	250,000	7,500,000	113

Source: LE analysis of survey data

Figure 11   Cost distribution by category of costs

Note: Respondents could choose from a choice of ranges. The midpoints of these ranges were taken to produce the chart above. The largest option (£5m+) was transformed to 7.5m – therefore, these charts are potentially a conservative estimate. The error bars represent the standard errors around the mean.

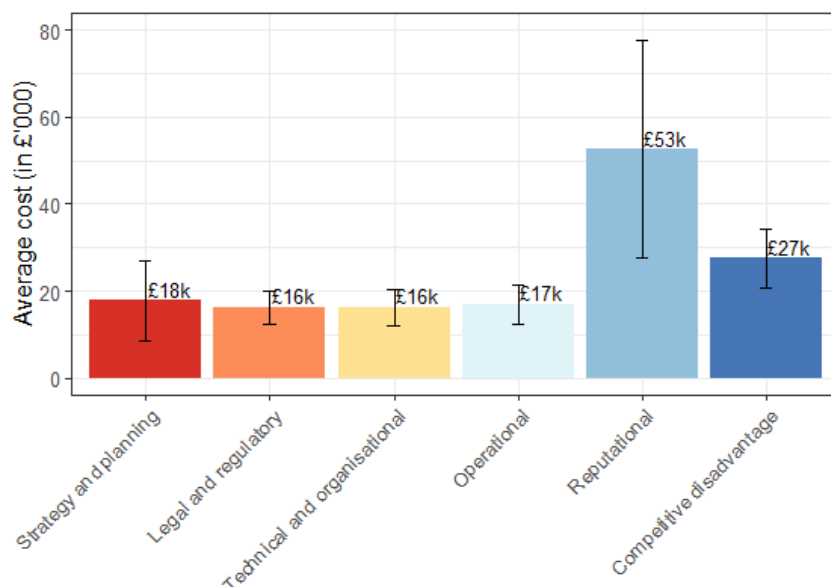*Source: LE analysis of survey data*

There is no significant difference between the different categories of explicit costs or implicit costs.

Table 7   Summary statistics of each category of costs

Size	Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
Strategy and planning	0	230	920	17,804	3,068	0	146
Legal and regulatory	0	288	2,299	16,198	8,707	0	196
Technical and organisational	6.14	307	920	16,127	6,898	6.14	170
Operational	6.14	307	2,299	16,857	8,707	6.14	133
Reputational	0	307	2,597	52,572	22,993	0	128
Competitive disadvantage	0	307	3,068	27,486	22,993	0	77

Source: LE analysis of survey data

**Contrary to the quantitative findings, a theme from the interviews was legal costs are perceived as the largest explicit cost factor.** Legal costs of data storage and sharing are high, and potentially prohibitive for smaller firms. A supportive regulatory and legal environment was described as being key to encouraging increased sharing of data. GDPR was seen as complex and potentially burdensome for compliance.

In the interviews, technical costs (and management costs) are perceived to have a similar magnitude to legal costs in some cases, but this category of cost was considered an integral part of the business model rather than as costs uniquely related to data sharing.

The discrepancy between the findings from interviews and the survey could be explored in further research.

The average cost of data sharing can also be split by sector:

Figure 12   Average cost of data sharing by sector
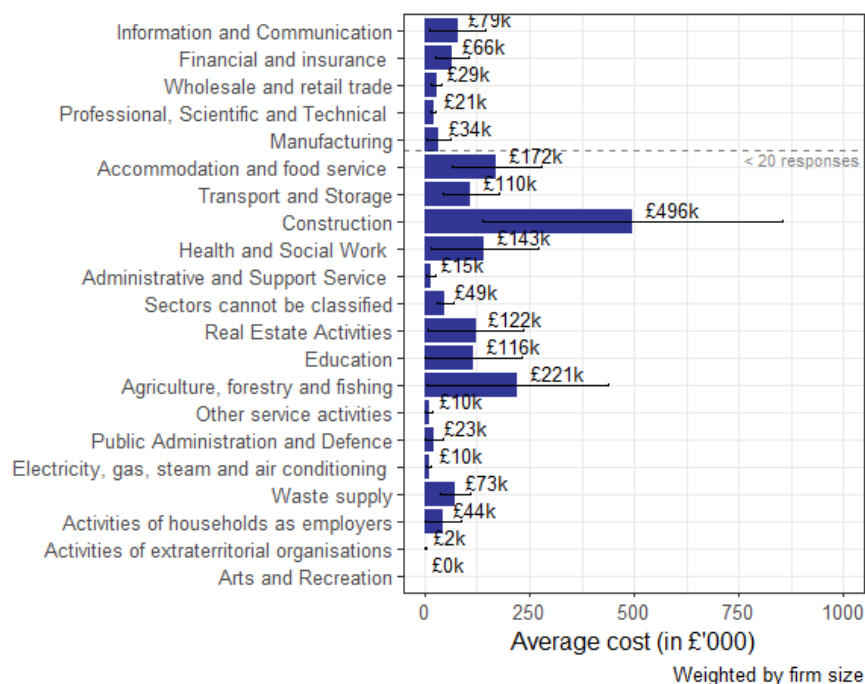
Note: As respondents could select multiple sectors, for the above analysis results were split into multiple observations. For example, a respondent within the Information and Communication and Manufacturing sectors is treated as a respondent in the former and the latter for the purposes of creating these average costs. The chart also denotes sectors with less than 20 responses. The error bars represent the standard errors around the mean.

Source: LE analysis of survey data

From these five most represented sectors, the average cost of data sharing ranges from roughly £21,000 to £79,000, however, these differences are not statistically significant. The Information and Communication (£79, 000), and Financial and insurance (£66,000) sectors have higher costs than the other three sectors within the band of represented sectors. Although these differences are not significant, this result is in line with what we would expect to see in a larger sample, as these are sectors which may have higher levels of data sharing and data processing (UKBDS, 2022).

Box 1   Sensitive data

Firms sharing sensitive and personal data have their own concerns related to sharing. A common theme identified from the interviews is a lack of easily identifiable use cases for commercialising or sharing data.

"When I think about data sharing, I get these negative images in my mind about privacy […] I can't think of a positive use for commercialising data" – Interview participant from large firm in Professional, scientific and technical activities sector

Data sharing is prevented by data controllers worried about potential reputational costs from perceived or actual misuse of data – this is especially true for controllers of sensitive data.

Changes in the definition of personal and sensitive data are also a concern for businesses sharing data, with one respondent commenting on the definition of personal data in global markets becoming "very wide".

Sharing data between markets within and outside the EU are also affected by international data sharing and privacy policies. For example, China has introduced strict privacy laws which limit the movement of personal data between the UK and their market – resulting in a prohibition of data sharing without significant infrastructure investment. For one particular respondent, they had to rebuild their data sharing computer infrastructure to comply with Chinese laws.

5.2.3 Data lifecycle

The costs across each step of the data lifecycle, are not significantly different from one another, however, firms reported that slightly higher costs were experienced in the 'Use' stage. The proportion of costs between these stages increases until this stage (Use), and then falls for the 'Share and Publish' stage. This could suggests that for organisations sharing data, the costs inherent with sharing the data itself is a fraction (at less than 20%) of the total costs required for sharing.

This proportion across the lifecycle is similar for firms of different sizes, however, smaller firms tend to incur a higher proportion of costs in the sharing stage and less in the collecting stage compared to larger firms. Large firms experience economies of scale by spending more on data infrastructure, resulting in lower sharing costs. Consequently, large firms experience higher costs in the planning stage – possibly because of these firms needing to ensure data sharing is consistent with company policy, procedure, and existing legal and regulatory agreements.

Figure 13   Proportion of costs across the data lifecycle for organisations sharing data by firm size
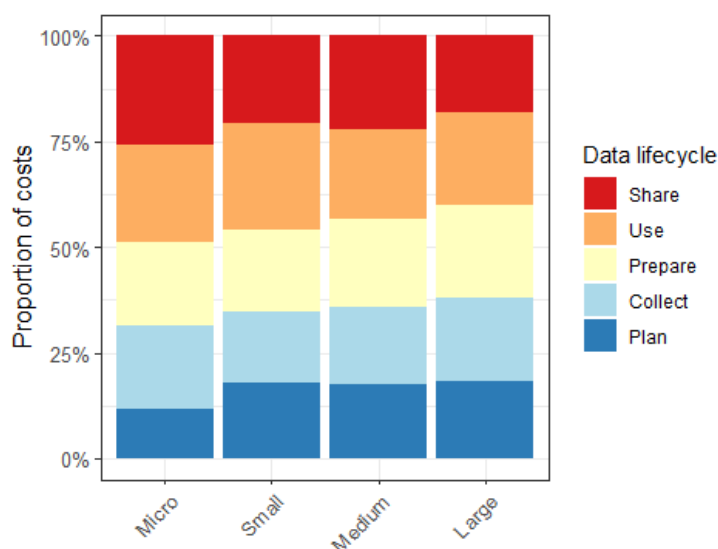
Source: LE analysis of survey data

The proportion of costs across the data lifecycle for different types of data can be seen in Figure 14. There are few significant differences between where in the data lifecycle the costs are between the different types of data being shared.

Figure 14   Proportion of costs across the data lifecycle for different types of data
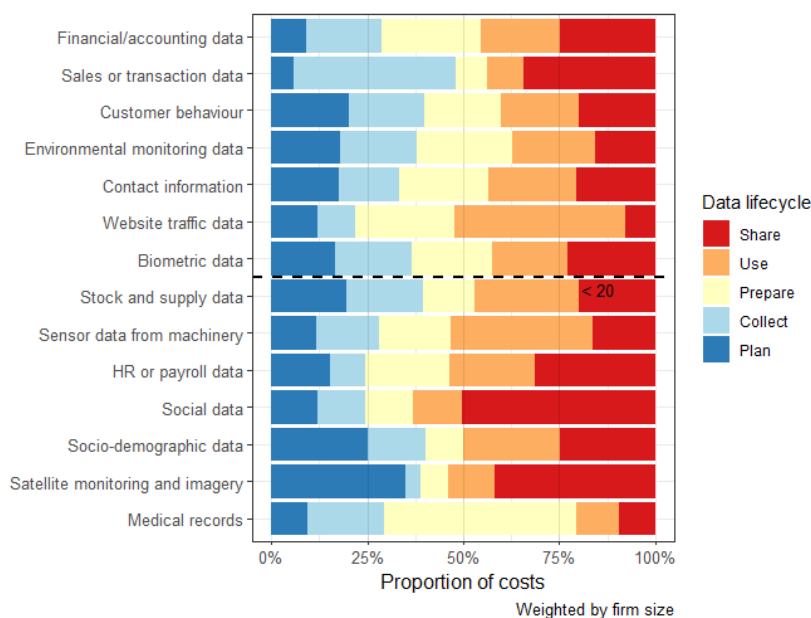
Note: Data types on y-axis are ordered from data types with highest firm count to lowest. All data types appearing below the dotted line represent the views of fewer than 20 firms.

Source: LE analysis of survey data

5.3   Data sharing strategy costs

Organisations will need to develop a strategy for data sharing, which should include what data will be useful to be shared, how it can be valued and the various structures/interfaces that can be used to share it. Data sharing strategy includes the (previously discussed) costs of:

- Planning and coordination costs
- Data collection costs
- Procurement costs

5.3.1   Overall range of costs

The interquartile range of strategy costs lie between £230 and £3,068 – with a mean weighted cost of £17,804. The average cost is much higher than median cost, suggesting that a small number of firms have large strategy costs, and many firms have relatively small strategy costs.

Table 8   Summary statistics for data sharing strategy costs for firms sharing data

Type	Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
Strategy and planning	0	230	920	17,804	3,068	1,298,576	146

Source: LE analysis of survey data

Overall, the costs estimates gathered from the interviews are slightly lower than the weighted averages from the survey results, however, they confirm the finding that there is a lot of variation between firms. Some of the estimates provided are listed below:

■ Staff costs spent on engaging with data sharing partners amount to around **£50,000** per year – *small firm sharing customer behaviour and events and booking data*

■ Ad-hoc projects are undertaken to look at specific elements of data sharing strategy, which cost around **£30,000** per year - *medium sized firm sharing data on research grant availability and outcomes*

■ Administrative burden associated with gathering data into a shareable format would amount to around **£3,200**[20] - *large research institute working with survey/research data and customer data*

■ In terms of staff costs, an estimated **30% internal analysts' time** is spent dealing with data requests and enquiries related to strategy and planning – *large training provider and IT consultancy sharing learner data, course data and client data*

Figure 15   Average strategy and planning costs for firms sharing data per size of firm
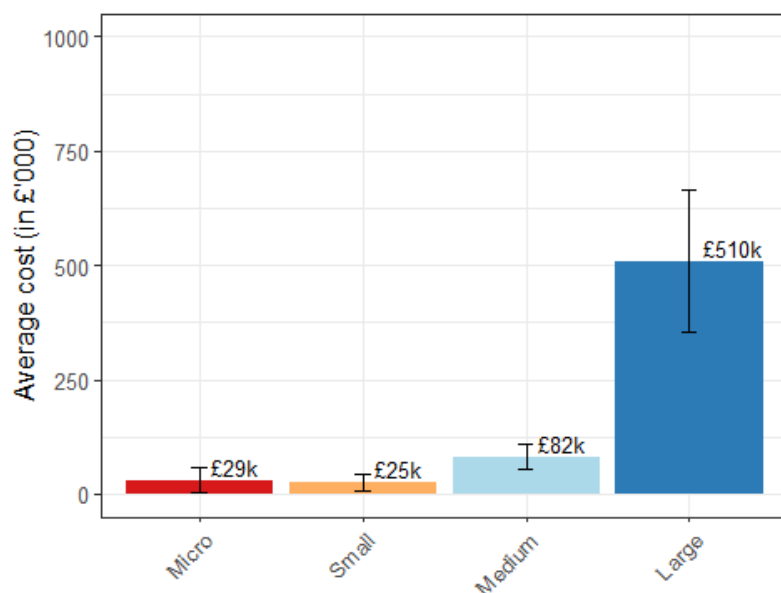
Note: Respondents could choose from a choice of ranges. The midpoint of these ranges was taken to produce the chart above. The largest option (£5m+) was transformed to 7.5m – therefore, these charts are potentially a conservative estimate. The error bars represent the standard errors around the mean.

Source: LE analysis of survey data

Table 9   Summary statistics for data sharing strategy costs for firms sharing data by firm size

Size	Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
Micro business	500	500	500	29,000	500	250,000	9
Small SME	0	500	2,500	25,286	6,250	250,000	14

---

[20] Calculated based on response of 2 months of a half-time role @ £20 per hour.

Size	Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
Medium SME	500	2,500	7,500	81,820	25,000	750,000	50
Large firm	500	7,500	75,000	509,541	250,000	7,500,000	73

Source: LE analysis of survey data

5.3.2 Lifecycle costs

Along the lifecycle, more respondents stated that data sharing strategy costs were incurred in the 'Plan' stage compared to other stages. The planning stage of the lifecycle involves determining business needs and identifying datasets – which is closely aligned with the costs included as part of the data sharing strategy.

Most firms in the stakeholder interviews also reported that data sharing strategy costs are mainly incurred towards the beginning of the data lifecycle, in the 'plan' stage as well as into one or more of the later stages of the data lifecycle. In particular, firms in which data sharing is integrated into their business model stated that strategy and planning costs are incurred in almost all stages of the lifecycle, which indicates that data sharing costs do not fit neatly within a linear data lifecycle model.

Figure 16   Number of firms selecting the data lifecycle stages in which data sharing strategy costs are incurred
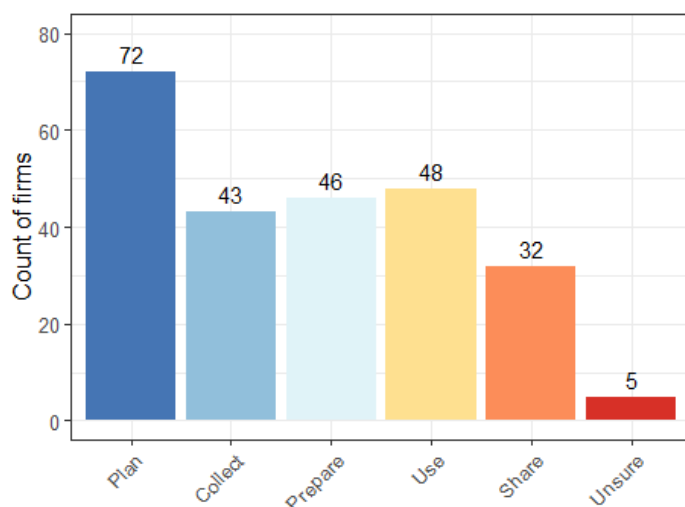
Note: Respondents selected which stages these costs were incurred at – they were given the option to select multiple stages. These counts are unweighted, and may not represent the UK business population.

Source: LE analysis of survey data

Despite a higher number of respondents stating that these costs were incurred in the 'Plan' stage, **the highest proportion of data strategy costs were found to be at the 'Collect' stage on average, with the lowest proportion incurred in the 'Share' stage**. However, these differences were not found to be significant, which again indicates that the data lifecycle may not be an appropriate model for exploration of data sharing costs.

Figure 17   Data sharing strategy costs along the data lifecycle for firms sharing data, by firm size
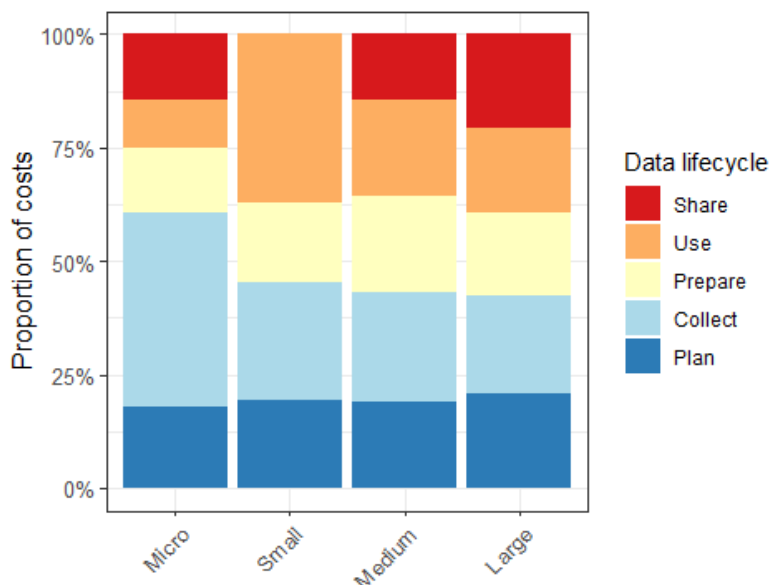
Note: Respondents proportioned costs along each stage they had identified in an earlier survey question.

Source: LE analysis of survey data

5.3.3   Assessment and synthesis with stakeholder interviews

The responses to the survey were quite varied, with the average costs much higher than median costs, suggesting that a small number of firms have very large strategy costs, and many firms have relatively small strategy costs. This is in line with the responses from the stakeholder interviews, which provide more insight and context for the survey results. This section provides a synthesis of the information gathered from the interviews with the view of triangulating the evidence from the survey.

The size of data strategy costs for businesses

From the stakeholder consultations, the size of data sharing strategy costs can vary **depending on whether data sharing is an integral part of the business**, or whether it is seen as an ad-hoc project. Resources invested into the establishment of data standards, the planning of interfaces or the ascertainment of benefits of data sharing is more vital for data-heavy organisations. As a result, these data sharing strategy costs are interwoven into general organisational and business strategy discussions. In these cases, seemingly prohibitive costs may instead be necessary strategy and planning costs for the business as a whole.

Organisations for which data sharing is an integral part of their business activities commented that **data sharing strategy is mostly integrated within their processes**, and that it is sometimes difficult to separate these as distinct costs. For example, some firms stated that a lot of time was spent planning specific data sharing activities and considering the data sharing implications of their wider data strategy, however, this was not something they could measure as these costs were **built into their broader operating costs**.

For other firms, **data sharing occurs as part of a one-off/standalone project**, in which case strategy and planning costs can be significant. One of the firms interviewed mentioned that strategy and

planning costs would involve at least 3-5 days of a manager's time, which may be prohibitive for a micro business or small SME.

The size of strategy costs also depend on the format and mechanism of data sharing. Firms internalising strategy processes incur staff resources, whereas organisations **outsourcing their data strategy** to consultants or external organisations would instead have a purely financial cost to pay. The owner of a small restaurant chain, for example, commented that their strategy cost is their biggest cost area because of the fees paid to a third-party provider to inform and aid them in commercialising their data.

Data strategy costs may differ by firm and data characteristics

Data sharing strategy costs may disproportionately affect **smaller and younger firms**. Based on the stakeholder consultations, for firms that have **existing policies and procedures** in place, such as large firms or firms that have been sharing data for a long time, the time spent understanding and strategizing data sharing implications is negligible. On the other hand, small, data-sharing firms may spend a significant amount of time and resources on establishing data sharing potential and value. One small firm interviewed stated that they spent the first 6 months **setting up their data sharing offering**.

There was no evidence from the interviews that sector has an impact on strategy costs. Data strategy costs may instead depend on whether **data is personal or sensitive** (which could be related to the sector firms operate in). A large software provider working with biometric data said that if they were to share data, there would be significant costs inherent with convincing customers to **share personal data**, as well as solving coordination problems between these data controllers (as many controllers would not want to be the first to agree to share this sensitive data).

5.4 Legal and regulatory considerations

Organisations need to understand the compliance requirements for the type of data that is to be shared, and how to structure the sharing agreements in order to facilitate trusted data sharing between parties (if data is not shared openly).

Data sharing strategy includes the (previously discussed) costs of:

- Compliance costs - collection
- Compliance costs - sharing
- Data sharing agreements

5.4.1 Overall range of costs

The interquartile range of data sharing costs are £288 and £8,707 - with a mean cost of £16,198.

Table 1 **Summary statistics for legal and regulatory costs for firms sharing data**

Type	Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
Legal and regulatory	0	288	2,299	16,198	8,707	389,573	196

Source: LE analysis of survey data

Overall, the costs estimates gathered from the interviews are in line with the weighted averages from the survey results. Based on the information gathered in the interviews there is a lot of variation between firms, with large firms and firms that share personal data incurring significantly higher costs. Some of the estimates provided are listed below:

- Drawing up data sharing agreements involves **£3,000** in legal fees, and the time spent amending internal privacy policies and terms and conditions cost around **£2,000**. – *small firm sharing events and booking data, and website search data*

- Updating policies and procedures to be GDPR compliant cost around **£6,000**[21], however, the stakeholder pointed out that these costs were limited due to the lack of personal data involved – *medium firm sharing data on research grant availability and outcomes*

- Employing a data governance team, which ensures compliance and evaluates the need for legal services on a case-by-case basis, costs between **£1-2 million per year** - *large marketing firm attached to a UK retailer*

Figure 18   Average legal and regulatory costs for firms sharing data per size of firm
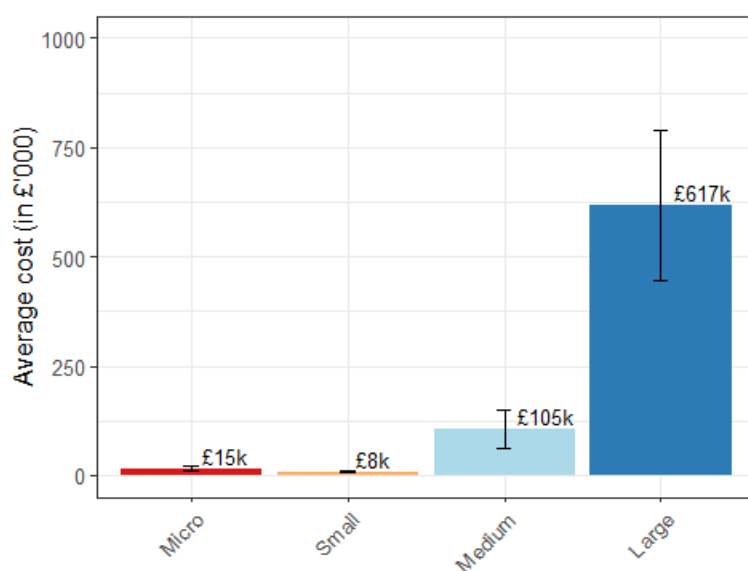
Note: Respondents could choose from a choice of ranges. The midpoint of these ranges was taken to produce the chart above. The largest option (£5m+) was transformed to 7.5m – therefore, these charts are potentially a conservative estimate. The error bars represent the standard errors around the mean.

Source: LE analysis of survey data

Large firms reported greater costs than smaller firms – this may be due to the need for increased legal and regulatory compliance due to the potential for this data sharing to affect many more partners and other external users. One of the large firms from the interviews that collects customer behaviour data mentioned that they would not share data with external parties unless they can exert considerable legal controls on the use of data. For this reason, the firm stated that individual data agreements take months to finalise. Another large firm working with personal data stated that reputable companies would do a full legal due diligence exercise before embarking on a data sharing process.

---

[21] Cost based on information received from stakeholder of 20 days of work @ £300 per day

Table 10   Summary statistics for legal and regulatory costs for firms sharing data by firm size

Size	Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
Micro business	0	500	2,500	15,310	25,000	75,000	21
Small SME	500	500	2,500	8,269	7,500	75,000	56
Medium SME	0	2,500	25,000	105,300	75,000	2,500,000	60
Large firm	0	7,500	25,000	617,225	250,000	7,500,000	89

Source: LE analysis of survey data

5.4.2   Lifecycle costs

Along the lifecycle, a similar number of respondents stated that **legal and regulatory costs are incurred across all stages of the lifecycle**. Most organisations that share data seem to acknowledge the potential cost if data sharing is not compliant with legal and regulatory restrictions. These firms may incur costs by drafting data sharing agreements ('Plan' phase) and obtaining legal counsel to ensure that data is collected, prepared, used and shared in a way that adheres to data protection regulations. As identified in the expert interviews, organisations sharing data sometimes fail to understand the magnitude of these costs, resulting in the cancellation or rescindment of proposed data sharing.

Figure 19   Legal and regulatory costs along the data lifecycle for firms sharing data
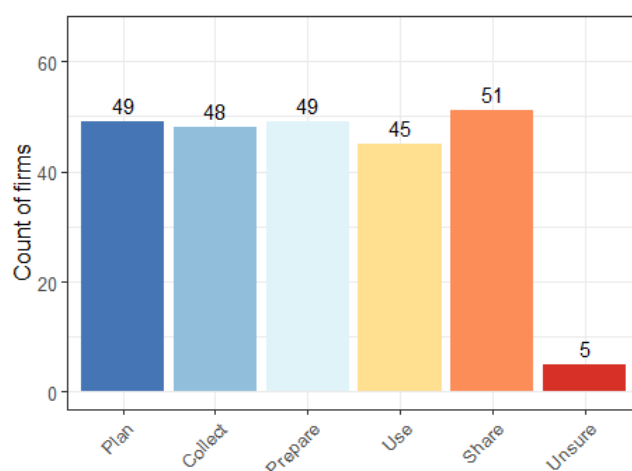
Note: Respondents selected at which stages these costs were incurred. These counts are unweighted and may not represent the UK business population.

Source: LE analysis of survey data

For Micro businesses, the highest proportion of legal and regulatory costs occurs in the 'Prepare' phase. This could be because of the type of data collected by Micro businesses in the sample. If the data that is being collected is personal, additional administrative costs may be present to ensure this data is prepared and used in a responsible and compliant manner.

Some firms that were interviewed remarked that while most legal and regulatory costs occur up-front, these costs can continue to accrue over time. The upfront legal and regulatory costs related to ensuring compliance and finalising data sharing agreements, but they can also be incurred later

if data sharing agreements are updated or when new products are rolled out and data is put to new uses.

Figure 20   Legal and regulatory costs along the data lifecycle for firms sharing data
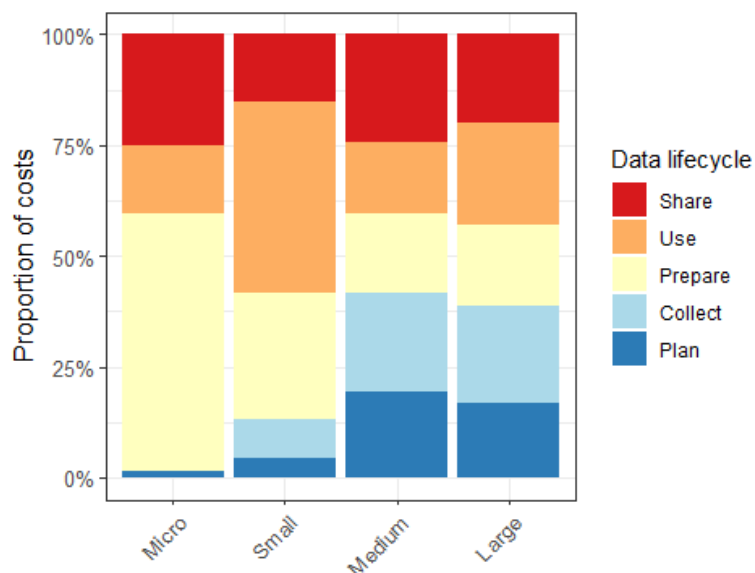
Note: Respondents proportioned costs along each stage they had identified in an earlier survey question.

Source: LE analysis of survey data

5.4.3   Assessment and synthesis with stakeholder consultations

The establishment of data sharing terms and agreements is important for organisations choosing to share data. These costs are incurred from external legal counsel or additional internal staff resources. In some cases, these legal costs represent large up-front costs, which then decrease over time as these agreements only need to be updated periodically. However, these costs may scale depending on the sensitivity of the data and associated legal and regulatory risks.

Legal and regulatory costs may differ by firm and data characteristics

As with data strategy costs, the findings from the interviews suggest that although large firms spend considerably more than small firms on legal and regulatory activities, these costs may **affect smaller firms disproportionately.** Multiple smaller firms noted that legal and regulatory costs were their biggest costs relating to data sharing, or that maintaining compliance required considerable staff time and ad-hoc costs. One large firm, on the other hand, stated that legal and regulatory costs, although high were small compared to the overall budget of data sharing projects.

Firms for whom data sharing is closely integrated with their business model may employ specialised individuals to handle legal and regulatory obligations. For example, one of the large firms that was interviewed employs a data governance team. One of the responsibilities of this team is to manage compliance, however, this is not seen as an additional data sharing cost to firms, as it is part of their business-as-usual expenditure.

Generally, the interviews confirmed the findings from the literature and the expert interviews that **personal and sensitive data** are associated with more risk to companies, and therefore have higher legal and regulatory costs than other types of data. One medium firm noted that their legal and

regulatory costs were very limited due to the lack of personal data involved, while another firm stated that if they were to share their personal data and be compliant, they would need to contact every person from whom they collected data to receive explicit consent.

5.5 Technical and organisational considerations

Organisations need to understand the technical considerations and mechanisms required for sharing data with external stakeholders.

Data sharing strategy includes the (previously discussed) costs of:

- Software costs
- Security costs
- Storage costs
- Data preparation costs

5.5.1 Overall range of costs

The interquartile range of data sharing costs are £307 and £6,898 - with a mean cost of around £16,127.

Table 11   Summary statistics for technical and organisational costs for firms sharing data

Type	Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
Technical and organisational	6.14	307	920	16,127	6,898	389,573	170

Source: LE analysis of survey data

The cost estimates gathered from the interviews are broadly in line with the weighted averages from the survey results, however, based on the interview responses it may be that the survey results underestimate the technical costs of data sharing for large firms. Some of the estimates provided by interviewees are listed below:

- Software licences required to run a data sharing platform cost an estimated **£3,000 - £5,000** per month - *small firm sharing customer behaviour, and events and booking data*
- Building an API to share data in real-time took one week and cost approximately **£1,000 - £2,000** - *small firm, dealing and sharing customer behaviour, events and booking, and website search data*
- Investment in open standard technology of between **£30,000 - £50,000 annually for 6 years** - *small firm sharing customer behaviour and activity data*
- Paying for external management systems (**£4,000 per year**), a data warehouse solution (**£100 per month**), and a business intelligence tool (**£15,000 per year**) - *medium sized firm sharing data on research grant availability and outcomes*
- Investments into sharing data, including staff time, encryption, storage, and external consultants cost roughly **£500,000 – £1 million** - *large training provider and IT consultancy sharing learner data, course data and client data*
- Significant investment was required for new infrastructure and platforms to share data. The costs of this (including staff costs and software licensing costs) were **"in the area of"**

tens of millions of pounds - with about 10% of this investment attributed to costs associated with sharing - *large marketing firm attached to a UK retailer*

Figure 21   Average technical and organisational costs for firms sharing data per size of firm
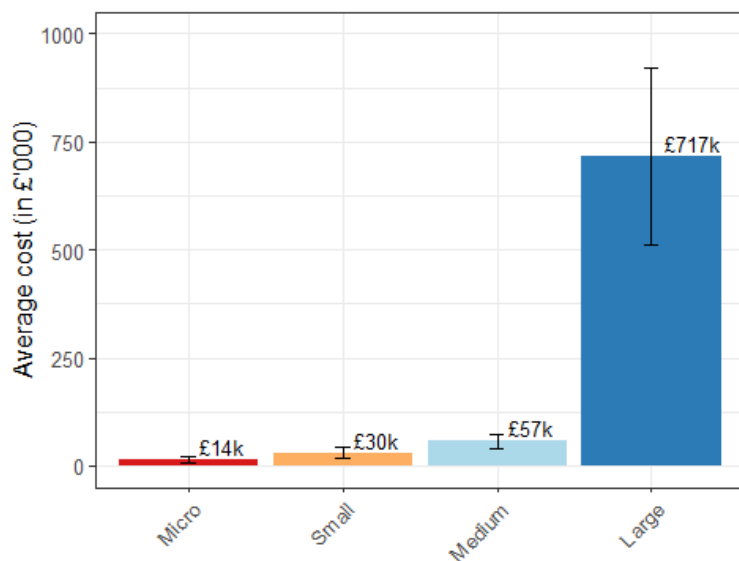
Note: Respondents could choose from a choice of ranges. The midpoint of these ranges was taken to produce the chart above. The largest option (£5m+) was transformed to 7.5m – therefore, these charts are potentially a conservative estimate. The error bars represent the standard errors around the mean.

Source: LE analysis of survey data

As with previously discussed costs, **large organisations report a significantly higher mean than organisations of smaller sizes**. These costs may be explained by large firms needing to scale costs to appropriately manage a larger volume of data sharing. For example, the large firms that interviewed stated that they had made (or would have to make) significant investments into infrastructure to be able to share data.

Table 12   Summary statistics for technical and organisational costs for firms sharing data by firm size

Size	Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
Micro business	500	500	500	13,667	3,750	75,000	12
Small SME	500	500	5,000	30,096	20,625	250,000	26
Medium SME	500	2,500	7,500	57,255	75,000	750,000	51
Large firm	500	7,500	75,000	716,957	250,000	7,500,000	81

Source: LE analysis of survey data

5.5.2   Lifecycle costs

Most survey respondents stated that **technical and organisational costs were incurred in the 'Prepare' and 'Use' stages in the data lifecycle** – with the 'Collect' stage closely behind. This is also reflected in the stakeholder interviews, and is expected as this classification of costs (such as software, security, and storage costs) mostly relate to the technical requirements and resources needed during these lifecycle stages. Data storage costs, for example, would be incurred in the

'Prepare, store and maintain' stage, while specialised software may be required in the 'Use and process' stage.

Figure 22   Technical and organisational costs along the data lifecycle for firms sharing data
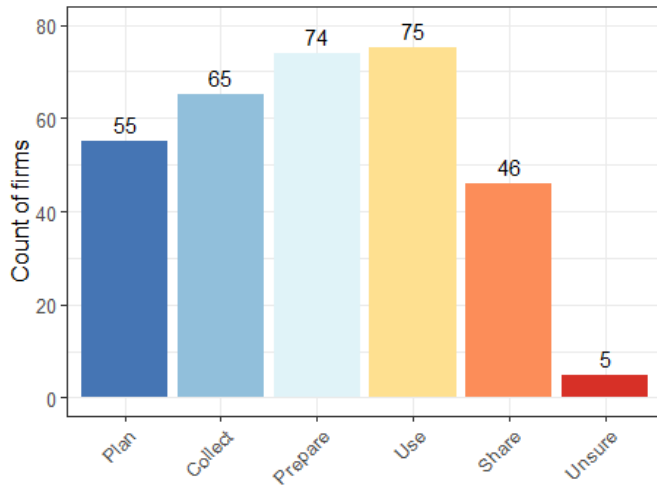
Note: Respondents selected at which stages these costs were incurred. These counts are unweighted and may not represent the UK business population.

Source: LE analysis of survey data

The proportion of these costs across the lifecycle slightly differ between sizes of firms. **Technical and organisational costs are highest in the 'Prepare' stage for Micro, Small and Large firms, with 'Collect' close behind.** The lowest proportion of costs arriving in the 'Share' stage. The relatively lower proportion of technical and organisational costs in the 'Share' stage may be a result from the different modes in which data can be shared. Lower technical costs would be expected for sharing data once-off with a single partner, compared to sharing data openly through an API.

Figure 23   Technical and organisational costs along the data lifecycle for firms sharing data
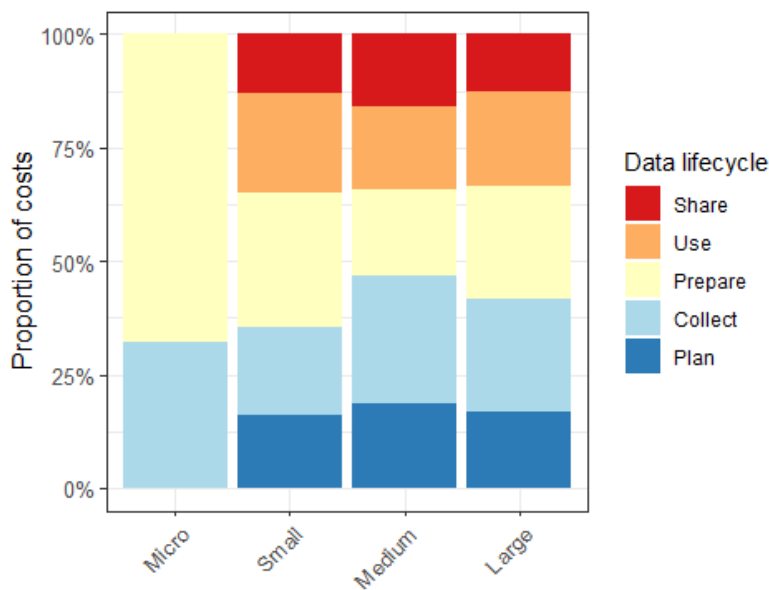
Note: Respondents proportioned costs along each stage they had identified in an earlier survey question.

Source: LE analysis of survey data

5.5.3 Assessment and synthesis with stakeholder interviews

Technical and organisational considerations are costs which are incurred within the middle stages of the data lifecycle – namely the 'Collect', 'Prepare' and 'Use' stages. Due to the potentially higher technical requirements, it is expected that these costs are most present in these stages. These costs range from software licences and data storage solutions, to the establishment of wider sharing structures – such as data standards.

These **costs may also vary depending on the data** itself – unstructured data may require greater resources to clean and prepare it for the purposes of sharing. Large and/or personal datasets may also require bespoke – and expensive – storage solutions to ensure data is stored responsibly while being accessible for other processes, such as further analysis or sharing. A large software provider working with biometric data stated that for them to be able to share that data, they would need to increase the size of their team by between **20-25%** (currently their product and tech team employs around 250-260 individuals).

Technical and organisational costs can also be incurred as a result of **large-scale investment into data sharing infrastructure**. For firms that wish to share data on a wide scale, either with high periodicity or volume, they may make investments into additional staff, data storage and interfaces/platforms to share this data. If data sharing becomes an increasingly important aspect of a firm's business model, they may choose to pay these high investment technical and organisational costs.

The size of technical and organisational costs varies considerably across the smaller firms that were interviewed. For some, these costs were a significant portion of their overall data sharing costs, while for others these costs were minimal. It could be that technical costs are higher for firms when **data sharing is integral to their business model**, compared to when the business focus is unrelated to data. For example, the owner of a small restaurant chain attributed their low technical costs to the fact that they only have simple IT processes, which are largely managed by a third party.

Box 2 Technical costs for the Arts, Entertainment & Recreation sector

Companies interviewed from the Arts, Entertainment & Recreation sector have commented that there is a relative unavailability of technical skills within the sector. As a result, technical costs for firms within this sector are often incurred by commissioning external services and skills.

Technical costs incurred by external consultants and services have included hiring other firms to compress their data for the purposes of sharing, as well as paying for additional external data management systems and storage solutions. These costs are often recurring monthly costs for these firms. Additionally, external developers are often hired to develop prototypes for new data sharing platforms and procedures.

The lack of in-house data and technical specialisation and skills both within these firms and the sector more widely have resulted in the need for these additional external costs.

5.6   Operationalising data sharing

There may be additional considerations for the operationalisation of data sharing and after data sharing has taken place.

Data sharing strategy includes the (previously discussed) costs of:

■   Data quality monitoring

■   Data sharing interface maintenance

5.6.1   Overall range of costs

The interquartile range of operational costs are £91 and £2,737 - with a (weighted) average cost of £16,857.

Table 13   Summary statistics for operational costs for firms sharing data

Type	Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
Operational	6.14	307	2,299	16,857	8,707	389,573	133

Source: LE analysis of survey data

The range of cost estimates gathered from the interviews is quite broad, but overall interviewees did not consider operational costs to be as high as the other types of costs. In contrast, the results from the survey are similar for operational costs and the other explicit costs. Some of the estimates provided by interviewees are listed below:

■   Maintaining APIs to ensure the quality of data cost around **£1,000 per year**, and maintaining databases of business customers' facilities requires about 1 day of staff time per month - *small firm, dealing and sharing customer behaviour, events and booking data and website search data*

■   Paying a development team to keep the data sharing platform running, monitoring its efficiency, and maintaining data quality costs around **£350,000 per year** - *small firm sharing customer behaviour, and events and booking data*

■   The maintenance of systems directly required for data sharing costs about **£1,000 per year** - *medium firm sharing data on research grant availability and outcomes*

Figure 24   Average operational costs for firms sharing data per size of firm
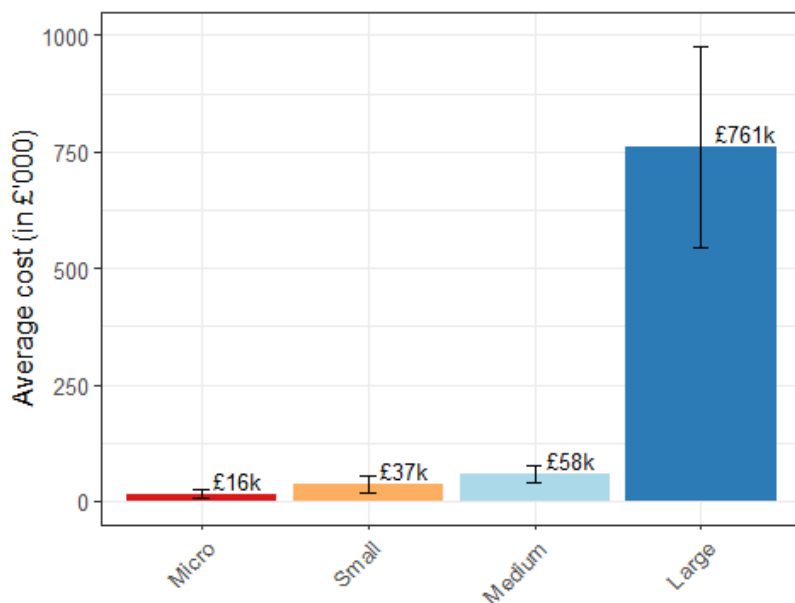
Note: Respondents could choose from a choice of ranges. The midpoint of these ranges was taken to produce the chart above. The largest option (£5m+) was transformed to 7.5m – therefore, these charts are potentially a conservative estimate. The error bars represent the standard errors around the mean.

Source: LE analysis of survey data

The average operational cost increased with firm size, this may be due to potentially larger amounts of data (and subsequent sharing) needing to be monitored. It could be that the methods large firms use to share data are more sophisticated, and therefore have higher maintenance cost for data sharing interfaces. These interfaces require significant labour and technical costs, particularly if it is used by several different users and purposes, which would increase the potential and prevalence of software bugs.

Table 14   Summary statistics for operational costs for firms sharing data by firm size

Size	Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
Micro business	500	2,000	7,500	15,750	11,875	75,000	8
Small SME	500	1,000	2,500	36,806	20,625	250,000	18
Medium SME	500	2,500	7,500	57,911	75,000	750,000	45
Large firm	500	25,000	75,000	760,540	750,000	7,500,000	62

Source: LE analysis of survey data

5.6.2   Lifecycle costs

Operational costs were most often reported to be incurred in the 'Prepare', 'Use', and 'Share' stages – with respondents reporting that these costs were not incurred to the same extent in the 'Plan' and 'Collect' stages. Operational costs refer to costs spent on ensuring data quality and sharing interfaces are maintained once the data sharing is operationalised. Once this data is shared with external users, the need for monitoring potentially increases.

Figure 25   Operational costs along the data lifecycle for firms sharing data
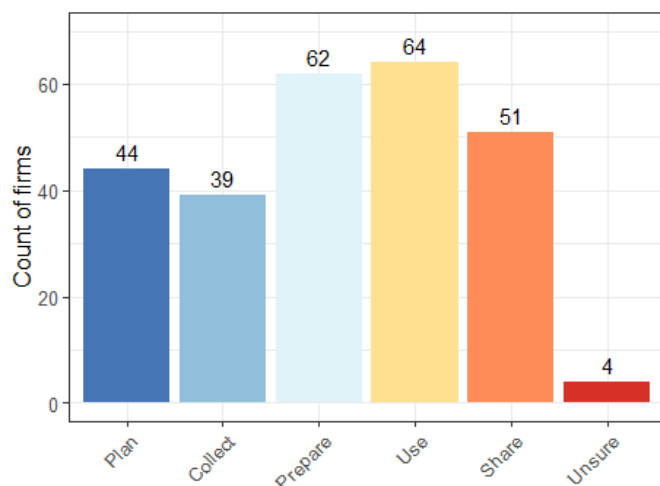
Note: Respondents selected at which stages these costs were incurred. These counts are unweighted and may not represent the UK business population.

Source: LE analysis of survey data

The proportion of these costs across the lifecycle slightly differ between sizes of firms. For micro firms, the largest proportion of operational costs are incurred in the 'Use' stage, for larger firms, they experience a higher proportion of costs in the 'Collect' stages. Micro firms reported that they do not experience operational costs at the 'Collect' stage when sharing data. For some organisations, data quality monitoring may begin as the data is collected, and therefore remain a constant cost over the data lifecycle. The quality of data may need to be higher if it is shared with external partners compared to being used only internally – Figure 8 shows that data is typically shared in a cleaned, processed or summary format and not in its raw form. This means that additional checks may need to be run continuously to ensure the quality of data and that it is anonymised. It is therefore not surprising that a large portion of the operational costs are incurred when data is being prepared and used.

Figure 26   Operational costs along the data lifecycle for firms sharing data
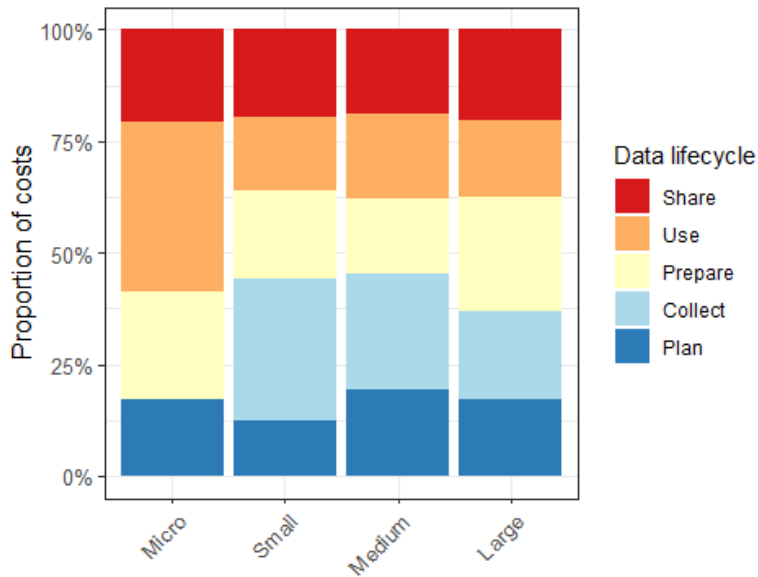
Note: Respondents proportioned costs along each stage they had identified in an earlier survey question.

Source: LE analysis of survey data

5.6.3   Assessment and synthesis with stakeholder interviews

Responsible sharing of data requires on-going monitoring and maintenance to ensure both the data, and the interface used to share this data, are error-free and easy to understand for external users. Although this category of costs refers to costs incurred once sharing is operational, the processes, thinking and set-up may start earlier in the data lifecycle – to ensure a smooth launch, high quality data and continual operation of data sharing processes.

One key finding from the stakeholder interviews is **that operational costs of data sharing may be difficult to disentangle from technical and organisational costs, or from operational costs**. One of the firms interviewed stated some staff and software costs could be allocated to this data sharing cost bucket, but it was not feasible to put separate figure against these activities. A different firm stated that operation costs associated with data sharing were considered operational to the firm's activities, and therefore couldn't be split out from other non-sharing related costs.

5.7   Implicit costs

5.7.1   Reputational

Firms that share data may experience a reputational cost if the nature of the data is sensitive, or if there are other objections from customers/partners/other stakeholders.

The interquartile range of reputational costs are £307 and £22,993 - with a (weighted) average cost of £52,572.

Table 15   Summary statistics for reputational costs for firms sharing data

Type	Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
Reputational	-	307	2,597	52,572	22,993	2,902,344	128

Source: LE analysis of survey data

Between different sizes of firms, **the cost increased as the firm size increased** – this may be potentially explained by the value large firms place on their reputation. A blow to this reputation – especially for large, international firms, may be seen as very costly. Interestingly, small Businesses believed the cost of reputational damage to be higher to them than medium Businesses. This may, however, be due to sample size and weighting issues.

Figure 27   Average cost of reputational damage for firms sharing data per size of firm
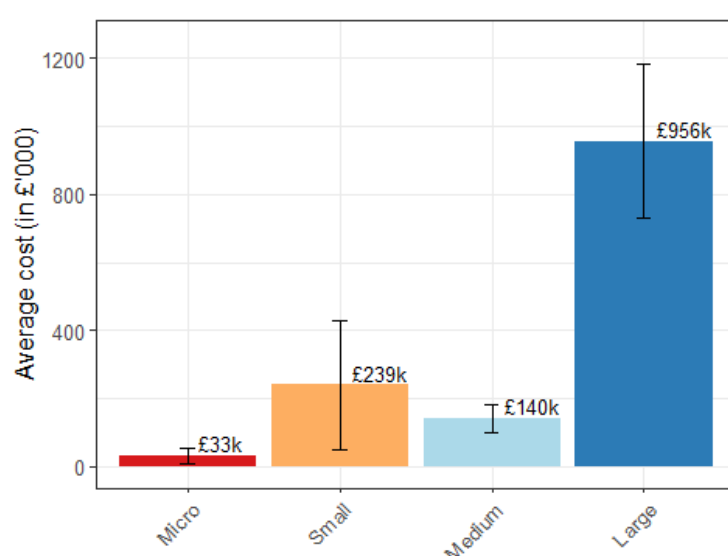
Note: Respondents could choose from a choice of ranges. The midpoint of these ranges was taken to produce the chart above. The largest option (£5m+) was transformed to 7.5m – therefore, these charts are potentially a conservative estimate. The error bars represent the standard errors around the mean.

Source: LE analysis of survey data

Table 16   Summary statistics for reputational costs for firms sharing data by firm size

Size	Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
Micro business	500	500	500	32,545	13,750	250,000	11
Small SME	500	2,500	25,000	239,462	25,000	2,500,000	13
Medium SME	0	7,500	25,000	140,283	250,000	750,000	74
Large firm	0	25,000	75,000	956,358	750,000	7,500,000	30

Source: LE analysis of survey data

Lifecycle costs

While the potential for reputational cost may come from misuse of data, organisations may still receive hostile reactions to the collection of data – especially personal data. As a result, many organisations selected the '**Collect' and 'Share' stages as the stages in which this cost occurs.** Most firms may believe that asking for personal data results in negative reactions, and that further

negative reactions may take place when personal data is shared with other organisations. Other reputational concerns may result from potential misuse of data by the organisation receiving the shared data.

Figure 28   Reputational costs along the data lifecycle for firms sharing data
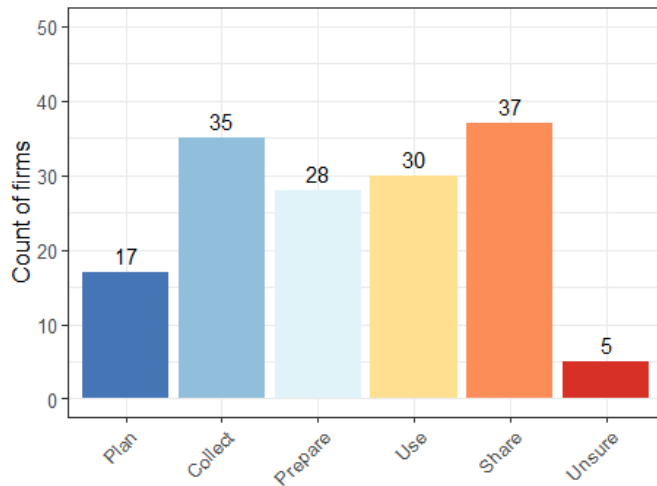
Note: Respondents selected which stages these costs were incurred. These counts are unweighted and may not represent the UK business population.

Source: LE analysis of survey data

Proportion of these costs across the lifecycle differ between firms. Micro and Small Businesses (possibly due to low sample sizes), only reported costs incurred in the 'Use' stage and 'Share' stage respectively. For Medium firms, the 'Collect' stage has the highest proportion of costs but for large firms, these costs are more evenly split (with the 'Use' stage having a slightly larger proportion). Organisations must ensure in the planning stages that any potential reputational harm occurring from this data sharing is minimised as costs in the latter stages may arise from potential misuse.

Figure 29   Reputational costs along the data lifecycle for firms sharing data
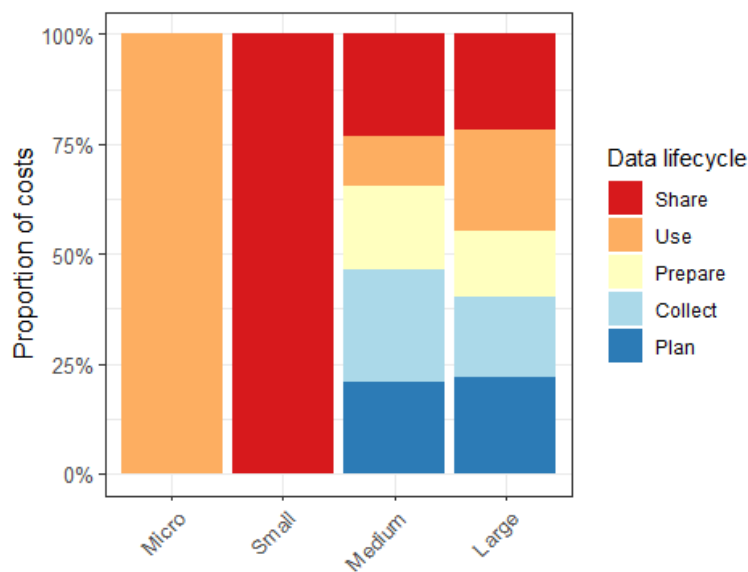
Note: Respondents proportioned costs along each stage they had identified in an earlier survey question.

Source: LE analysis of survey data

5.7.2 Competitive disadvantage

Loss to a business if data leads to a loss in competitive advantage.

The interquartile range of operational costs are approximately £307 and £22,993 – with a (weighted) average cost of almost £27,486.

Table 17   Summary statistics for competitive disadvantage costs for firms sharing data

Type	Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
Competitive disadvantage	0	307	3,068	27,486	22,993	389,573	77

Source: LE analysis of survey data

The costs of competitive disadvantage increase as firm size increases – with large firms reporting a mean cost of over £1.6 million. Large firms are defined as having turnover of greater than £250m, and as such will stand to lose more if their competitive advantage is damaged.

Figure 30   Average competitive disadvantage costs for firms sharing data per size of firm
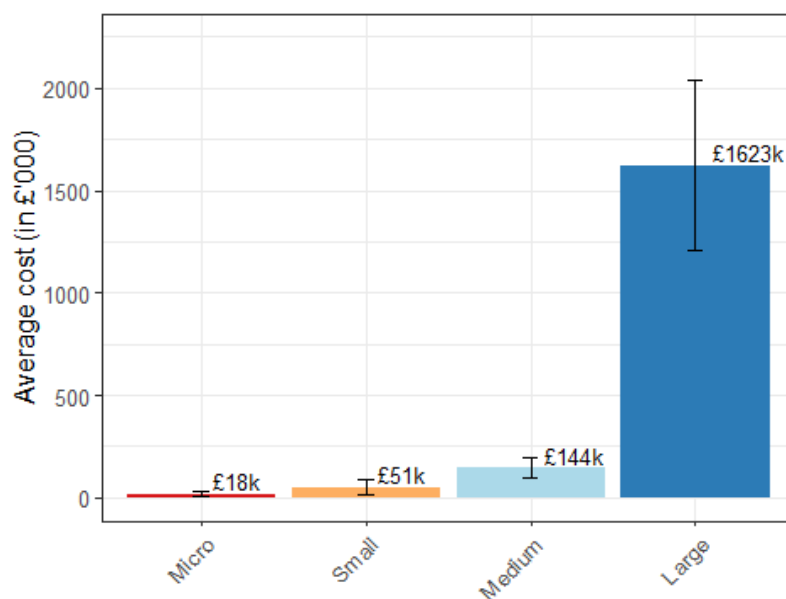
Note: Respondents could choose from a choice of ranges. The midpoint of these ranges was taken to produce the chart above. The largest option (£5m+) was transformed to 7.5m – therefore, these charts are potentially a conservative estimate. The error bars represent the standard errors around the mean.

Source: LE analysis of survey data

Table 18   Summary statistics for competitive disadvantage costs for firms sharing data by firm size

Size	Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
Micro business	500	500	4,000	18,167	20,625	75,000	6
Small SME	0	500	7,500	51,214	50,000	250,000	7
Medium SME	500	7,500	25,000	143,833	250,000	750,000	21
Large firm	500	25,000	250,000	1,622,547	1,625,000	7,500,000	43

Source: LE analysis of survey data

Lifecycle costs

The largest group of respondents reported that competitive disadvantage costs were incurred in the 'Share' stage – this may be because they believe this loss of competitive disadvantage occurs when proprietary data is shared.

Figure 31   Competitive disadvantage costs along the data lifecycle for firms sharing data
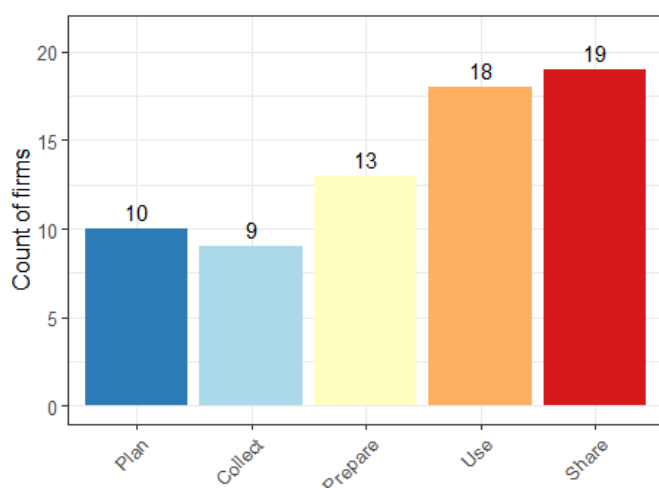

Note: Respondents selected at which stages these costs were incurred. These counts are unweighted and may not represent the UK business population.

Source: LE analysis of survey data

Subsequently, **the greatest proportion of cost arises in the 'Share' stage** – this is consistent with a belief that the sharing of data leads to a competitive disadvantage.

Figure 32   Competitive disadvantage costs along the data lifecycle for firms sharing data
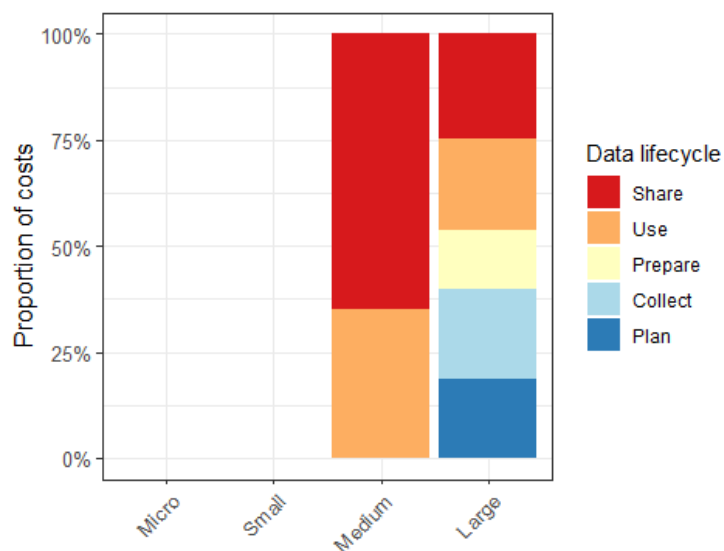
Note: Respondents proportioned costs along each stage they had identified in an earlier survey question.

Source: LE analysis of survey data

5.7.3   Assessment

Implicit costs are valued by respondents as significantly greater than explicit costs. This may be due to these costs being harder to conceptualise – and therefore value – for respondents. These high values may also imply that firms overvalue the potential disbenefit associated with data sharing – this perception of these high implicit costs may serve as a barrier preventing further sharing for these organisations.

Conclusion and possible directions for future research

With the aim of furthering the understanding of data sharing costs, this report presents a taxonomy and quantification of these costs. The taxonomy separates explicit costs of data sharing into four main classifications which have been validated from expert interviews, as well as from evidence of interviewee familiarity with these costs from the primary data collection processes (survey and interviews).

Survey and interview participants were able to quantify the costs they face when sharing data (or deciding to share data) – this research was therefore successful at providing some level of quantification. However, several limitations exist with this research which affect the interpretation and robustness of the results. The quantitative figures reported in the charts in **Section 4** were created from a sample of 500 responses with a bias towards large firms. Although the analysis has been weighted to match UK business population statistics, it should still be acknowledged that the candidates sampled in this research have a specific set of characteristics which are not entirely representative of the UK.

From the survey, the mean cost of data sharing (when weighted by firm size) is around £57,600, with large firms reporting a significantly higher average cost of data sharing than smaller firms. Across the taxonomy classifications, the highest reported cost was Reputational (with an average cost of £52,600), the lowest costs were Technical and Organisational (with an average cost of just over £16,100). While the information gathered from stakeholder interviews provides some insight into why these costs vary between different firms and types of data, the small interview sample does not allow us to generalise these insights.

The findings from the stakeholder consultations suggest that data sharing costs are highly context dependent. Within the small group of firms that were interviewed there was as much variation within a certain size category as there was variation between size categories. Instead, the factors that appeared to have a much greater impact on data sharing costs were the type of data that was being shared (whether the data was sensitive/personal), and whether the data sharing was integral to the firm's business operations (such as firms whose business model depends on data sharing).

The different types of data sharing costs we identified in the taxonomy were all placed along the stages of the data lifecycle by both survey respondents and interviewees. However, it should be noted that some of the costs were one off costs, whereas others were incurred every time data was shared. This indicates the limitations of trying to specify where in the lifecycle some costs occur.

The quantitative figures in this report are exploratory and indicative of the likely costs of data sharing and serve as a starting point for further research within the space. This research elicited cost figures from respondents sharing different types of data within different sectors and domains. Future research involving the creation of case studies showcasing data sharing within specific contexts and data types would not only further the understanding of how data sharing differs between subsets of the population, but also highlight the potential benefits of data sharing.

Other directions for future research include examining the differences between personal and non-personal, as well as sensitive and non-sensitive, data sharing. This research has found that sharing of personal data has its own unique challenges and costs, especially in terms of regulatory and legal compliance. If the intention is to encourage further data sharing, research into understanding the motivations behind data sharing would be useful to ensure the success of any future policy

interventions. Research into the methods of sharing (such as bilaterally, multilaterally, or as open data) would also be useful in understanding how costs and motivations for sharing data may differ depending on the mechanics of the data sharing itself.

Research into how firms access data would also be useful. It may be unclear for organisations who wish to access closed and siloed datasets what steps and processes they need to undertake – for example, who to contact, what compliance requirements they need to be aware of and the terms of access. Understanding the journey these firms need to undergo will help highlight many of the barriers and costs facing organisations on the demand side of data sharing.

Investigating data sharing and related policy solutions in other countries may provide a useful international comparison. For example, the IMDA in Singapore has developed a trusted data sharing framework to guide organisations through the data sharing process, in an effort to increase the prevalence of sharing (IMDA, 2020). The development of further materials (such as the ICO's code of practice for data sharing) (ICO, 2020) and similarly practical tools – combined with effective policy interventions tackling key costs – would help encourage higher rates of data sharing.

The research also faced a number of process-related issues and difficulties which may highlight broader trends in stakeholder understanding of data sharing related costs and issues. There was low evidence that stakeholders took a cost-centric approach regarding their data sharing decisions. Organisations approached the decision to share data by establishing a business case in terms of opportunity and potential benefit – once the decision was made, the associated costs often did not stop data sharing from occurring. This meant that respondents engaging with this research may have had difficulty discussing the costs of sharing data due to their benefit-centric approaches – if decisions are made on the basis of potential firm benefit rather than costs, it would explain why candidates may have elected to not engage with research focusing specifically on costs. Future research within this space may ensure greater uptake by considering other industry approaches to data sharing.

This report is an exploratory step into quantifying the costs of data sharing, the taxonomy serves as a verified classification of costs for future research; an understanding of different types of costs will also be useful for any potential policy interventions used to reduce the barriers and cost of data sharing. The evidence of the level of these costs also provides an overview of their relative magnitudes, as well as where along the data lifecycle these costs are incurred for organisations sharing data. Combined with future research (such as that suggested above), this will provide policymakers with a thorough understanding of the costs associated with organisations sharing data, as well as the other factors driving business decisions to share data.

References

Acxiom and Foresight Factory (2019). *Understanding Attitudes to Data.*

AWS (Amazon Web Services) (n.d.,). *What is data labelling for machine learning?.* Available at: https://aws.amazon.com/sagemaker/data-labeling/what-is-data-labeling/#:~:text=In%20machine%20learning%2C%20data%20labeling,model%20can%20learn%20from%20it.

Bald, J. (2020). *Industry insights – What is the true cost of data management?*

BEIS (2021). *Business population estimates for the UK and regions 2021: statistical release.* Available at: https://www.gov.uk/government/statistics/business-population-estimates-2021/business-population-estimates-for-the-uk-and-regions-2021-statistical-release-html#composition-of-the-2021-business-population

Clarkson, G., Jacobsen, T. E., & Batcheller, A. L. (2007). Information asymmetry and information sharing. *Government Information Quarterly*, *24*(4), 827-839.

Coyle, D., Diepeveen, S., Wdowin, J., Kay, L., and Tennison, J. (2020). *The value of data: summary report 2020*. On behalf of ODI, Bennett Institute and the Nuffield Foundation

Ctrl-Shift (2018). *Data Mobility: The personal data portability growth opportunity for the UK economy.*

Data and Marketing Association, Acxiom and Foresight Factory (2018). *Global Data Privacy: What the consumer really thinks.*

DCMS (Department for Digital, Culture, Media & Sport) (2022). *UK Business Data Survey 2022: detailed findings*. Available at: https://www.gov.uk/government/statistics/uk-business-data-survey-2020/uk-business-data-survey-2020-detailed-findings

Figueiredo, A. S. (2017). *Data sharing: convert challenges into opportunities.* Frontiers in public health, 5, 327.

Frontier Economics (2020). *Geospatial Data Market Study*

Frontier Economics (2021). *Increasing Access to Data Across the Economy*

Government data quality hub (2020). *Government data quality framework*, available at: https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework#The-Data-Lifecycle

IBM Security (2021) *Cost of a Data Breach Report 2021.* Available at: https://www.ibm.com/security/data-breach

ICO (Information Commissioner's Office) (2020). *Data sharing: a code of practice*. Available at: https://ico.org.uk/for-organisations/guide-to-data-protection/ico-codes-of-practice/data-sharing-a-code-of-practice/

IMDA (Infocomm Media Development Authority) (2020) *Trusted Data Sharing Framework*. Available at: https://www.imda.gov.sg/-/media/Imda/Files/Infocomm-Media-Landscape/SG-Digital/Tech-Pillars/Artificial-Intelligence/Trusted-Data-Sharing-Framework.pdf

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

London Economics (2020). *Evaluation of Data Pitch*

McKinsey, (2013). *Open Data: Unlocking innovation and performance with liquid information*.

University of Portsmouth (2020). *Victims of Computer Misuse*. Commissioned by the Home Office and Her Majesty's Inspectorate of Constabulary, Fire and Rescue Services (HMICFRS).

OECD (2019). *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*. Available at: https://www.oecd-ilibrary.org//sites/15c62f9c-en/index.html?itemId=/content/component/15c62f9c-en#

OECD (2020). *Measuring the economic value of data and cross-border data flows*

ANNEXES

Annex 1   Survey

A1.1   Introduction

The Department for Digital, Culture, Media and Sport (DCMS) have engaged London Economics (LE) to explore the costs associated with sharing data between organisations. This research will assist DCMS in supporting innovation from organisations making data available to users outside their organisations.

The survey will provide insights into the costs and barriers facing your organisation and the wider sector regarding data sharing – including pound estimates for these costs. Your input to this study is highly appreciated, and will help inform any future policy interventions.

You can participate in this survey completely anonymously. However, if you provide your contact details and confirm your consent, we may contact you with an invitation for a further telephone consultation (up to 30 minutes in length), for which you will be compensated with a £50 voucher.

London Economics' privacy policy can be found: https://londoneconomics.co.uk/privacy-policy/

A1.1.1   Definitions

Data sharing

For the purpose of this survey, **data sharing** is defined as making data held by your organisation available for outside organisations or users to access.

This data sharing can be done either through providing open access to data (such as hosting downloadable datasets or providing API access) or through providing restricted access (including selling or licensing private datasets).

Data refers specifically to digital, machine-readable data that is created or collected by your organisation and is not readily available otherwise (e.g. public data, official statistics). It can include data your organisations collects as part of normal business operations (e.g. transaction data, customer data, sensor data from plants and machinery, HR data, etc.). It may be data over which your organisation has intellectual property rights or data covered by legal obligations (e.g. personal data).

For this survey we are not interested in sharing to comply with legal, regulatory and contractual obligations (such as official reporting duties, tax returns, commercial and employment contracts or similar).

A1.2   Survey

1. **Please provide your details below:**

Organisation name:

□ Prefer not to answer

2. **What is your role in the organisation?**

□ Chief Technical Officer
□ Project manager
□ IT decision maker
□ General manager
□ Small business owner
□ Other (please specify): _____

3. **Where are your organisation headquarters based?**

□ Northern Ireland
□ Scotland
□ Wales
□ East Anglia
□ East Midlands
□ West Midlands
□ London
□ North East
□ North West
□ Yorkshire and Borders
□ South East
□ South West

4. **What is the size of your organisation?**

□ Micro SME (Less than 10 employees and less than £2m turnover)
□ Small SME (Less than 50 employees and less than £10m turnover)
□ Medium-sized SME (Less than 250 employees and less than £50m turnover)
□ Large firm (More than 250 employees or more than £50m turnover)

5. **What sectors do you operate in? Select all that apply. (Answer using following SIC Code if possible, further details found here)**

□ A - Agriculture, forestry and fishing
□ B – Mining and quarrying
□ C – Manufacturing
□ D – Electricity, gas, steam and air conditioning supply
□ E – Waste supply; sewerage, waste management and remediation activities

☐ F – Construction
☐ G – Wholesale and retail trade; repair of motor vehicles and motorcycles
☐ I – Accommodation and food service activities
☐ H – Transport and Storage
☐ J – Information and Communication
☐ K – Financial and insurance activities
☐ L – Real Estate Activities
☐ M – Professional, Scientific and Technical Activities
☐ N – Administrative and Support Service Activities
☐ O – Public Administration and Defence; Compulsory Social Security
☐ P – Education
☐ Q – Human Health and Social Work Activities
☐ R – Arts, Entertainment and Recreation
☐ S – Other service activities
☐ T – Activities of households as employers
☐ U – Activities of extraterritorial organisations and bodies
☐ Sectors cannot be classified using the above categories
☐ Don't know

A1.3 Data usage and infrastructure

Most organisations produce digital data, which might include: sales data from electronic points of sale or online transactions; sensor data from machinery; environmental monitoring data; visitors to your website; and HR or payroll data.

6. **Does your organisation generate, collect, or hold any digital data?**

☐ Yes
☐ No
☐ Don't know
☐ Not applicable

7. **(IF Q6 ≠ Yes) Just to confirm, you stated that your organisation does not handle any digital data, or that you are unsure. By 'data', we mean any data on sales, marketing, customers, sensor data from machinery, monitoring data, payroll data and similar that is processed by your organisation. Is this correct?**

☐ Yes, correct – my business does not handle any digital data
☐ No, incorrect – my business does handle digital data

IF Q7== YES: terminate

8. **We want to understand better how much your organisation spends on your data infrastructure (all the things that allow data collection and manipulation in your business) per year.**

a. On *software* costs (i.e., for data collection, cleaning and sharing, including the cost of licences, as well as the cost of developing programmes in-house), can you estimate the cost using the following ranges?
   o Zero
   o Less than £1000
   o £1,000 to £4,999
   o £5,000 to £9,999
   o £10,000 to £49,999
   o £50,000 to £99,999
   o £100,000 to £499,999
   o £500,000 to £999,999
   o £1m to £4,999,999
   o £5m+
   o Unsure

b. On *storage* costs (i.e., cost of storing data on-premises or cloud), can you estimate the cost using the following ranges?
   o Zero
   o Less than £1000
   o £1,000 to £4,999
   o £5,000 to £9,999
   o £10,000 to £49,999
   o £50,000 to £99,999
   o £100,000 to £499,999
   o £500,000 to £999,999
   o £1m to £4,999,999
   o £5m+
   o Unsure

c. On *security* costs (i.e., administrative cost of ensuring that infrastructure is secure and the risk to your organisation is limited, including cybersecurity software and licensing costs), can you estimate the cost using the following ranges?
   ☐ If you can't separate the costs of security from the costs of storage and software, please tick here instead

   o Zero
   o Less than £1000
   o £1,000 to £4,999
   o £5,000 to £9,999
   o £10,000 to £49,999
   o £50,000 to £99,999
   o £100,000 to £499,999
   o £500,000 to £999,999
   o £1m to £4,999,999
   o £5m+
   o Unsure

9. How many full time equivalent (FTE) members of staff do you have employed whose primary role is to undertake activities related to operating and maintaining your organisation's data infrastructure (including all data software, storage and security)?

☐ None
☐ 1 to 2
☐ 3 to 4
☐ 5 to 6
☐ 7 to 10
☐ 11 or more
☐ Unsure

A1.4 Data Sharing

Your organisation may hold multiple types of data or datasets which could be valuable to external organisations if shared.

10. Does your organisation generate, collect or hold any digital data that you consider _valuable_ for external users or organisations (whether your organisation currently shares this or not)?
☐ Yes
☐ No
☐ Unsure
(If Q10 == YES , skip to Q12)

11. (IF Q10 == No | Unsure) Sharing data or datasets can be valuable if it has not been previously available to external users or if it can be used for the generation of further insights or analysis. Can you confirm that your organisation does not hold any digital data you consider valuable for external users or organisations?
☐ Yes – our organisation **does not** hold any valuable digital data
☐ No – our organisation **does** hold valuable digital data
☐ Unsure

(If Q11 == YES| Unsure, skip to Q28)

12. (IF Q10 == Yes OR Q11 == No) Which type of data that your organisation generates, collects or holds do you consider to be the _most_ potentially valuable for external users or organisations (whether your organisation currently shares this or not)? Tick the one response that best describes the data you are thinking of
☐ Environmental monitoring data (e.g., air quality and noise data)
☐ Satellite monitoring and imagery (i.e., for earth observation)
☐ Sales or transaction data (i.e., products sold)
☐ Stock and supply data (i.e., current stock, orders placed, inventory)
☐ Financial/accounting data (e.g., revenue, profit, accounts)
☐ Sensor data (e.g. from machinery)
☐ Website traffic data
☐ HR or payroll data
☐ Biometric data (e.g., face, fingerprint, voice etc.)
☐ Contact information (e.g., emails, names, addresses)
☐ Customer behaviour (e.g., buying habits, websites visited, links clicked etc.)
☐ Socio-demographic data (e.g., age, gender, political orientation)

- ☐ Social data (i.e., publicly shared data on social media profiles such as posts and check-ins)
- ☐ Location history
- ☐ Medical records
- ☐ Other (please specify): _____

The following questions all refer to the dataset you have identified as ***most potentially valuable for external users or organisations***

13. (IF Q10 == Yes OR Q11 == No): Do you share this data? [button for definition of data sharing]
- ☐ Yes
- ☐ No
- ☐ Unsure

(If Q13 == No | Unsure, skip to Q17)

14. (If Q13 == Yes): How do you share this data outside your organisation (Please tick all that apply)?

- ☐ Raw data (in the format in which it is generated / acquired)
- ☐ Cleaned data (ready for digital processing and analysis in a common, machine-readable format)
- ☐ Processed data (cleaned data with additional analysis/value added)
- ☐ Summary data (with reduced granularity compared with the raw data e.g., summary tables in a publication, interactive graphs on a website)
- ☐ Other (please specify): _____
- ☐ Don't know

A1.5   Costs of Data Sharing

15. (If Q13 == Yes): What does sharing this data cost your organisation (as a total rough £ estimate)? This includes infrastructure costs, staff costs, legal costs and any other related costs used explicitly for the sharing of this data.
- ☐ Zero
- ☐ Less than £1000
- ☐ £1,000 to £4,999
- ☐ £5,000 to £9,999
- ☐ £10,000 to £49,999
- ☐ £50,000 to £99,999
- ☐ £100,000 to £499,999
- ☐ £500,000 to £999,999
- ☐ £1m to £4,999,999
- ☐ £5m+
- ☐ Unsure

16. (IF Q13 == Yes): What costs affect your organisation's decision to share this data? Please select all that apply from the following list:

☐ Strategy and planning costs (e.g., costs of developing a strategy for data sharing, such as identifying the data needed to be shared, how to value this data and which users to share this with)

☐ Legal and regulatory costs (e.g., costs of ensuring that data sharing is compliant with regulation, administrative and legal cost of creating contracts)

☐ Technical and organisational costs (e.g., software, security, storage and data preparation costs)

☐ Operational costs (e.g., ongoing monitoring of data quality and ongoing maintenance of data sharing infrastructure)

☐ Reputational costs (e.g., sharing of sensitive data, sharing of data which is then misused, objections from stakeholders to the sharing of data)

☐ Competitive disadvantage (e.g., loss of competitive advantage from providing access to your data)

☐ Other – please state

☐ None

☐ Unsure

(If Q16 asked, skip to Q18)

17. (IF Q13 == No): What factors lead to your organisation decision not to share this data? Please select all that apply from the following list:

☐ Strategy and planning costs (e.g., costs of developing a strategy for data sharing, such as identifying the data needed to be shared, how to value this data and which users to share this with)

☐ Legal and regulatory costs (e.g., costs of ensuring that data sharing is compliant with regulation, administrative and legal cost of creating contracts)

☐ Technical and organisational costs (e.g., software, security, storage and data preparation costs)

☐ Operational costs (e.g., ongoing monitoring of data quality and ongoing maintenance of data sharing infrastructure)

☐ Reputational costs (e.g., sharing of sensitive data, sharing of data which is then misused, objections from stakeholders to the sharing of data)

☐ Competitive disadvantage (e.g., loss of competitive advantage from providing access to your data)

☐ Other – please state

☐ None

☐ Unsure

18. (If Q16/Q17 does not include None | Unsure) **Of the costs selected, please rank them in order of perceived importance to your organisation:**

   [List of costs selected in Q16 / Q17 (not including Other, None, Unsure)]

19. (If Q16/Q17 does not include None | Unsure) **Could you estimate the size of these costs from the following ranges?**
[Grid question]
[Rows: costs ticked in Q16 / Q17]

[Columns:
- ☐ Zero
- ☐ Less than £1000
- ☐ £1,000 to £4,999
- ☐ £5,000 to £9,999
- ☐ £10,000 to £49,999
- ☐ £50,000 to £99,999
- ☐ £100,000 to £499,999
- ☐ £500,000 to £999,999
- ☐ £1m to £4,999,999
- ☐ £5m+
- ☐ Unsure

Data use and sharing can be described using the following data lifecycle

■ *The following stages describe the data lifecycle, and where/how data is used.*
- ☐ *Plan – This stage is where an organisation (intending to collect, store and use data) must plan their processes and data storage. This stage involves determining business needs, identifying existing data and what needs to be collected.*
- ☐ *Collect, acquire, ingest – This stage involves the acquisition and collection of data.*
- ☐ *Prepare, store and maintain – This stage involves the preparation of data for storage, including formatting of data for later use, as well as storing this data appropriately.*
- ☐ *Use and process – This stage involves the use of data for your organisations business needs, this could include analysis of data or production of outputs.*
- ☐ *Share and publish – This stage involves preparing data for publication and sharing, as well as sharing data with users outside your organisation (whether this data is shared/ sold for monetary gain or otherwise)*

20. (IF Q13 == Yes): **Can you provide a rough % of where the <u>total costs</u> of sharing this data are incurred along the lifecycle i.e., what proportion of costs are incurred at which stage?**

[Grid question]

[Rows:
- ☐ Please provide a % of £ required:
- ☐ Unsure]

[Columns:
- ☐ Plan

□   Collect, acquire, ingest
□   Prepare, store and maintain
□   Use and process
□   Share and publish**]**

21. **(IF Q13 == Yes) (If Q16 does not include None | Unsure) : For each of the following costs, where along the lifecycle are these costs incurred? Please tick all that apply**

    [Grid question]
    [Rows: List of costs selected in Q16 (not including Other, None, Unsure)]

    [Columns:
    □   Plan
    □   Collect, acquire, ingest
    □   Prepare, store and maintain
    □   Use and process
    □   Share and publish**]**

22. **(IF Q13 == Yes) (*For each cost ticked in Q16) (*(If Q16 does not include None | Unsure) (If Q21 ≠ unsure) How are these costs incurred across the lifecycle? Please provide a % of the total cost for each stage.**

    [Grid question]

    [Rows:
    □   Please provide a % of £ required:
    □   Unsure]

    [Columns: Data lifecycle positions ticked in Q21 (i.e., Plan; Collect, Acquire, Ingest; etc.)]

23. **(IF any of Q21 / Q22 == Unsure) (For each cost from Q16) : Could you provide some detail on how these costs are incurred?**
    □   _____
    □   Unsure

XA. (IF Q13 == No): What does using this data cost your organisation (as a total rough £ estimate)? This includes infrastructure costs, staff costs, legal costs and any other related costs used explicitly for all uses of this data.
    □   Zero
    □   Less than £1000
    □   £1,000 to £4,999
    □   £5,000 to £9,999
    □   £10,000 to £49,999
    □   £50,000 to £99,999
    □   £100,000 to £499,999
    □   £500,000 to £999,999
    □   £1m to £4,999,999

□ £5m+

□ Unsure

24. **(IF Q13 == No): Can you provide a rough % of where the <u>total costs</u> of collecting, holding, processing and using this data are incurred along the lifecycle i.e., what proportion of costs are incurred at which stage?**

    [Grid question]
    [Rows:

    □ Please provide a % of £ required:_____

    □ Unsure]

    [Columns:

    □ Plan

    □ Collect, acquire, ingest

    □ Prepare, store and maintain

    □ Use and process**]**

25. **(IF Q13 == No) (If Q17 does not include None | Unsure) : For each of the following costs, where along the lifecycle are these costs incurred? Please tick all that apply**

    [Grid question]
    [Rows: List of costs selected in Q17 (not including Other, None, Unsure)]

    [Columns:

    □ Plan

    □ Collect, acquire, ingest

    □ Prepare, store and maintain

    □ Use and process **]**

26. **(IF Q13 == No) (*For each cost ticked in Q17) (*(If Q17 does not include None | Unsure) (If Q25 ≠ unsure) How are these costs incurred across the lifecycle? Please provide a % of the total cost for each stage.**

    [Grid question]

    [Rows:

    □ Please provide a % of £ required: _____

    □ Unsure]

    [Columns: Data lifecycle positions ticked in Q25 (i.e., Plan; Collect, Acquire, Ingest; etc.)]

27. **(IF any of Q25 / Q26 == Unsure) (For each cost Q17) : Could you provide some detail on how these costs are incurred?**

    □ _____

    □ Unsure

A1.6   Organisations which do not have valuable data

28. (IF Q11 == No | Unsure) : Why do you consider your organisation's digital data to not be valuable to external users or organisations?

- ☐ Unaware of external users or organisations who could use this data
- ☐ Data is unable to be shared due to legal or regulatory compliance requirements
- ☐ Data is not in a format ready for external users to generate value with
- ☐ The data's value is specific to our organisation's processes and uses
- ☐ Other (please specify):_____
- ☐ Unsure

XB (IF Q11 == No | Unsure) Which type of data that your organisation generates, collects or holds do you consider to be the most valuable to your organisation? Tick the one response that best describes the data you are thinking of

- ☐ Environmental monitoring data (e.g., air quality and noise data)
- ☐ Satellite monitoring and imagery (i.e., for earth observation)
- ☐ Sales or transaction data (i.e., products sold)
- ☐ Stock and supply data (i.e., current stock, orders placed, inventory)
- ☐ Financial/accounting data (e.g., revenue, profit, accounts)
- ☐ Sensor data (e.g. from machinery)
- ☐ Website traffic data
- ☐ HR or payroll data
- ☐ Biometric data (e.g., face, fingerprint, voice etc.)
- ☐ Contact information (e.g., emails, names, addresses)
- ☐ Customer behaviour (e.g., buying habits, websites visited, links clicked etc.)
- ☐ Socio-demographic data (e.g., age, gender, political orientation)
- ☐ Social data (i.e., publicly shared data on social media profiles such as posts and check-ins)
- ☐ Location history
- ☐ Medical records
- ☐ Other (please specify): _____

The following questions all refer to the dataset you have identified as most valuable to your organisation

XC. (IF Q11 == No | Unsure): What does using this data cost your organisation (as a total rough £ estimate)? This includes infrastructure costs, staff costs, legal costs and any other related costs used explicitly for all uses of this data.

- ☐ Zero
- ☐ Less than £1000
- ☐ £1,000 to £4,999
- ☐ £5,000 to £9,999
- ☐ £10,000 to £49,999
- ☐ £50,000 to £99,999
- ☐ £100,000 to £499,999
- ☐ £500,000 to £999,999

☐   £1m to £4,999,999
☐   £5m+
☐   Unsure

Data use and sharing can be described using the following data lifecycle

■   The following stages describe the data lifecycle, and where/how data is used.

☐ Plan – This stage is where an organisation (intending to collect, store and use data) must plan their processes and data storage. This stage involves determining business needs, identifying existing data and what needs to be collected.

☐ Collect, acquire, ingest – This stage involves the acquisition and collection of data.

☐ Prepare, store and maintain – This stage involves the preparation of data for storage, including formatting of data for later use, as well as storing this data appropriately.

☐ Use and process – This stage involves the use of data for your organisations business needs, this could include analysis of data or production of outputs.

☐ Share and publish – This stage involves preparing data for publication and sharing, as well as sharing data with users outside your organisation (whether this data is shared/ sold for monetary gain or otherwise)

XD. (IF Q11 == No | Unsure): Can you provide a rough % of where the total costs of collecting, holding, processing and using this data are incurred along the lifecycle i.e., what proportion of costs are incurred at which stage?

[Grid question]
[Rows:
☐   Please provide a % of £ required:_____
☐   Unsure]

[Columns:
☐   Plan
☐   Collect, acquire, ingest
☐   Prepare, store and maintain
☐   Use and process
☐   Unsure]

A1.7   Consent

29. Would you be interested in being contacted for a follow-up interview? This interview will be 30 minutes long and will ask questions related to your organisation's thinking regarding data

sharing. You will be paid a £50 voucher for your participation. If you are willing to be contacted, please provide your contact email below:

☐   _____

30. Please confirm that you consent to be contacted at this email address for a compensated follow-up interview:

☐   I confirm my consent to be contacted at this email address for a compensated follow-up interview

☐   I do not give my consent to be contacted, and I don't want my email address stored or used

A1.8   Termination / thank-you message

Thank you for taking the time to participate in this survey. Your responses will be valuable in enhancing the understanding of the costs facing organisations regarding holding and sharing digital data.

If you wish to find out more of the context of this research, you can find out about the UK National Data Strategy Policy Framework here.

Annex 2   Detailed sample overview

In terms of the positions that the respondents hold within their firms, 190 individuals interviewed identified as "IT decision makers", while 96 were general managers, 76 were Chief Technical Officers (CTOs), 64 were small business owners, and 46 project managers.

Figure 33   Overview of the roles of respondents
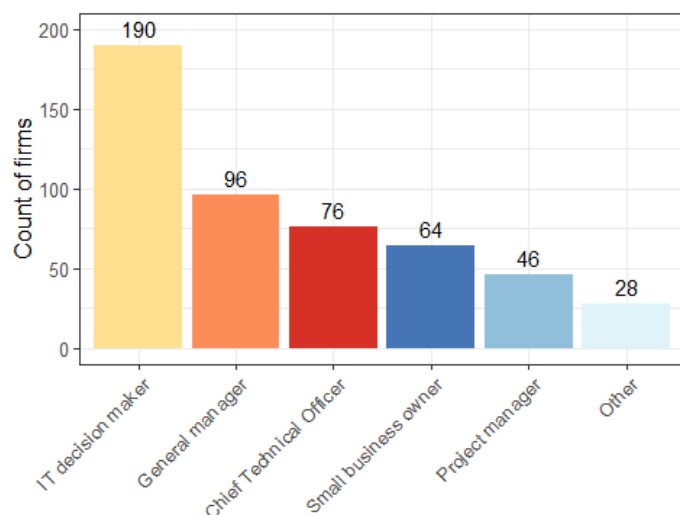
Source: LE analysis of survey data

The sample consists of 79 Micro businesses (fewer than 10 employees and less than £2m turnover), 64 small Businesses (fewer than 50 employees and less than £10m turnover), and 137 medium Businesses (fewer than 250 employees and less than £50m turnover). The largest portion of firms in the sample are large firms (220), defined as having over 250 employees or over £50 million turnover.

In comparison, the UK business population can be seen in **Table 19** below:

Table 19   Comparison of survey firm sizes to UK business population

Firm size	Sample[1]	Business population[2]
Micro	15.8%	82.1%
Small	12.8%	14.9%
Medium	27.4%	2.5%
Large	44.0%	0.5%

[1] *LE analysis of survey data*
[2] *Business population taken from BEIS (2021). Business population estimates for the UK and regions 2021: statistical release.*
Available at: https://www.gov.uk/government/statistics/business-population-estimates-2021/business-population-estimates-for-the-uk-and-regions-2021-statistical-release-html#composition-of-the-2021-business-population

Micro businesses comprise the majority of the UK business population, with 82.1% of firms in the UK being sole traders or micro-sized.[22] The sample is not representative of the business population especially due to an over-representation of large firms. All cost estimates in this report are weighted on firm size, unless results are disaggregated by size.

The average number of data employees[23] for each size of firm in the sample ranged from 3 to 9 (with the number of employees increasing with firm size). For Micro businesses, the average number of data employees represent a minimum of 30% of their total employment, for Small Businesses this represents between 10% and 30%, for Medium Businesses between 3% and 12%, and for large firms this proportion is maximally 4%.

It is unlikely that more than 30% of the employees in Micro businesses in the business population are data employees – this result therefore tells us more about the sample than it does about UK firms generally. Because the survey was aimed at businesses that collect and process digital data it is likely that the firms that were chosen to participate in this survey are firms such as tech start-ups, that do not represent the UK business population.

Figure 34   Average number of data employees

Note: Respondents could choose from several categorical ranges. The midpoints of these ranges were taken to produce an average. The lowest value of the highest range (11 or more) was taken as 12.5 - therefore these charts are potentially a conservative estimate.

Source: LE analysis of survey data

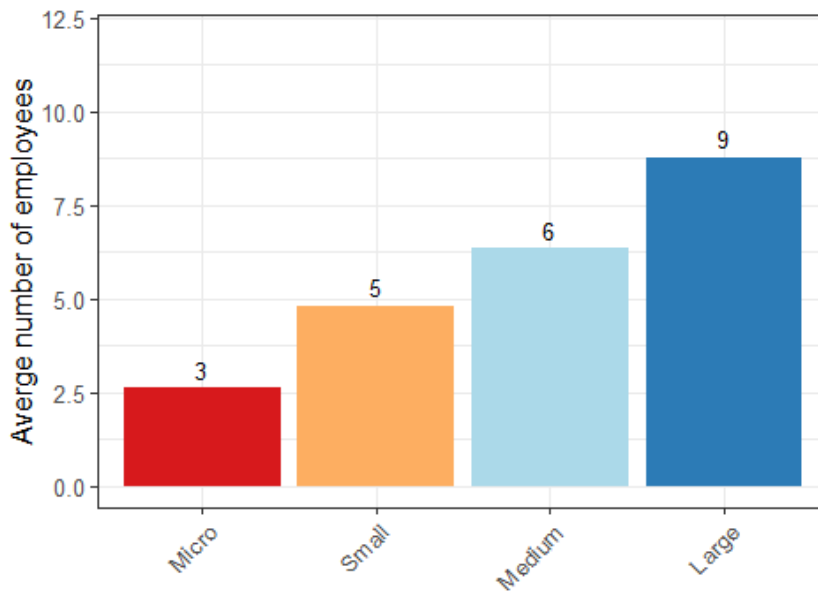The UK business population split can be seen below:

---

[22] Business population taken from BEIS (2021). *Business population estimates for the UK and regions 2021: statistical release.* Available at: https://www.gov.uk/government/statistics/business-population-estimates-2021/business-population-estimates-for-the-uk-and-regions-2021-statistical-release-html#composition-of-the-2021-business-population

[23] Respondents were asked how many members of staff they have employed whose 'primary role is to undertake activities related to operating and maintaining [their] organisation's data infrastructure (including all data software, storage and security).

Table 20   Sector comparison to UK business population

Sector	Sample	Business Population *
Agriculture, mining and utilities	1%	3%
Manufacturing	10%	5%
Construction	5%	16%
Retail	9%	10%
Transportation	4%	6%
Accommodation and food	4%	4%
Information and Communication	23%	6%
Financial and insurance	12%	2%
Real estate activities	3%	2%
Professional and scientific	11%	15%
Administrative and support services	3%	8%
Education	5%	5%
Health and social work	5%	6%
Arts and recreation	3%	5%
Other service activities	3%	6%

* Business population taken from BEIS (2021). Business population estimates for the UK and regions 2021: statistical release. Available at: https://www.gov.uk/government/statistics/business-population-estimates-2021/business-population-estimates-for-the-uk-and-regions-2021-statistical-release-html#composition-of-the-2021-business-population

The largest represented sector in the survey sample is the Information and Communication sector – which only comprises 6% of the total UK business population. The survey sample is seemingly biased heavily in favour of the Information and Communication, Financial and insurance activities, and Professional scientific and technical activities sectors.

Although the sector split is not representative (or closely aligned) to the UK business population, the imbalance towards these three sectors is less pronounced due to respondents selecting multiple sectors. Out of the 500 firms in the sample, 27 selected into two sectors, 16 into three, and 6 into four sectors or more. Over a fifth of companies that selected into the Information and Communication, Financial and Insurance, and Professional and Scientific sectors selected into multiple sectors. In particular, there is significant overlap between companies in the information and communication sector and companies in the Professional and Scientific sectors, as well as between Information and Communication and Financial and Insurance activities (see **Table 21**).

Table 21   Overlap between the 5 most represented sectors and other sectors

	Manufacturing	Retail	Information and Communication	Financial and insurance	Professional and scientific
Agriculture, mining, and utilities	1	0	1	0	2
Manufacturing	0	1	**4**	3	1
Construction	0	0	2	1	1
Retail	1	0	1	2	0

	Manufacturing	Retail	Information and Communication	Financial and insurance	Professional and scientific
Transportation	1	0	1	0	0
Accommodation and food service	1	1	2	0	0
Information and Communication	**4**	1	0	**7**	**9**
Financial and insurance	3	2	**7**	0	1
Real Estate Activities	1	0	1	1	0
Professional and Scientific	1	0	**9**	1	0
Administrative and Support Service	0	0	1	2	1
Education	1	0	1	0	1
Health and Social Work	0	0	0	0	0
Arts and Recreation	0	0	0	0	1
Other service activities	1	1	2	1	1

Restricting the sample to firms which share data, the sector representation across firms decreases further; as respondents could self-select into multiple sectors, a sector weighting on the survey data would result in heavily biased results.

Figure 35   Overview of sectors of survey respondents who share data
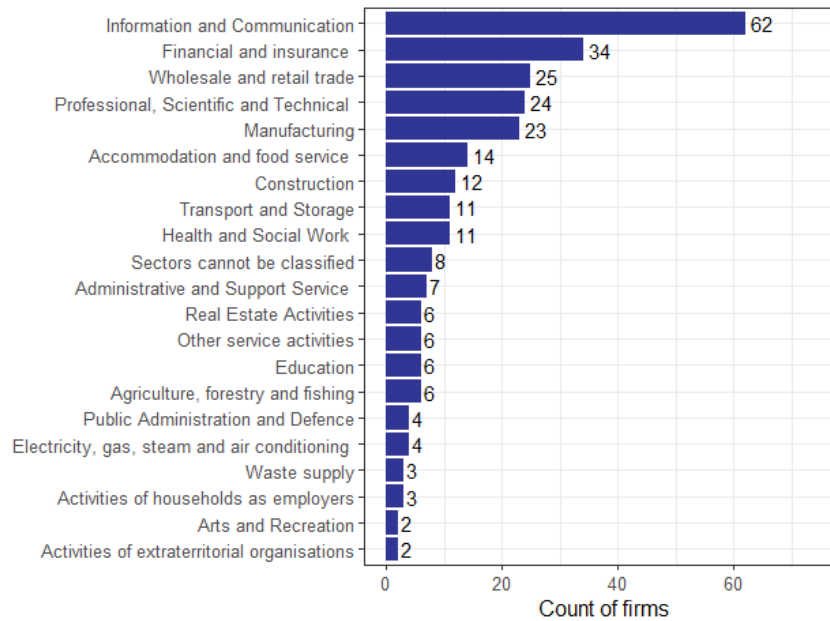
Almost a third of firms in the sample are headquartered in London (140 firms), followed by the South East (53), and the North West (50). Only a small number of firms that completed the survey are headquartered in Northern Ireland (8), however, respondents from other regions of the United Kingdom are fairly evenly distributed.

Figure 36   Overview of locations of firm headquarters

Compared to the UK business population, London and the South East are aligned in being the two most represented regions. However, this alignment diverges in instances such as the South West – the South West comprises 9.3% of the UK business population, but only 3.4% of the sample. Similarly, Wales is over-represented in the sample compared to the UK population, with 6.4% representation in the sample (versus 3.7% in the UK population).

Comparison of survey firm location to UK business population

Firm size	Sample	Business population *
London	28.0%	18.7%
South East	10.6%	15.7%
North West	10.0%	9.4%
East Midlands	9.2%	6.6%
Yorkshire and the Humber	7.0%	7.4%
East Anglia	6.8%	10.2%
West Midlands	6.4%	8.0%
Wales	6.4%	3.7%
Scotland	6.2%	6.1%
North East	4.4%	2.8%
South West	3.4%	9.3%
Northern Ireland	1.6%	2.2%

* Business population taken from BEIS (2021). Business population estimates for the UK and regions 2021: statistical release. Available at: https://www.gov.uk/government/statistics/business-population-estimates-2021/business-population-estimates-for-the-uk-and-regions-2021-statistical-release-html#composition-of-the-2021-business-population

Annex 3   Interviews and guide

A3.1   Recruitment process

Several sources were targeted in efforts to recruit respondents for the interviews. Offering a £50 incentive, the following process was followed:

1)   35 UK private organisations were identified from participation in previous data-related projects London Economics has completed or participated in.

    a) From this list, 5 interviews were completed; 15 contacts either refused, or were no longer working at this target organisation; 14 contacts did not reply.

2)   Our connections with DataIQ[24] were then leveraged to disseminate invitations to their panel. From this, an additional 1 interview was scheduled and completed.

3)   A survey with 500 businesses was conducted with IRBureau (details in Annex 2) in which respondents were asked for their consent to be contacted for a follow-on interview. From the panel of 500, 75 respondents consented for a follow-on interview.

    a) From this list of 75, 3 respondents scheduled and completed an interview.

    b) For the rest of this list, these respondents were repeatedly nudged by OMB Research as well as IRBureau and London Economics via email, however, this did not result in any additional responses. However, due to the requirements on participants in the panel, LE did not have access to phone numbers or organisation details from the survey results to enable further communication beyond email.

4)   In efforts to increase the number of successful interviews, LE has also contacted their partners at the Open Data Institute to identify further participants. From this strand, there has not been any additional respondents identified and interviewed.

Finally, DCMS recruited additional participants through their own network – from these efforts, 2 additional respondents were recruited; the incentive was also increased to £100.

A3.2   Sample

The sample consists of respondents from firms of differing sizes, sectors and interacting with different data types. These details can be seen in the following table:

Table 22   Interview panel details

Interview	Size	Sector(s)	Data type(s)
1	Small	Information and Communication; Arts, entertainment and recreation; Other service activities	Contact information; Sales or transaction data; Website traffic data

---

[24] DataIQ is a membership group connecting and supporting analytics and data professionals in the UK. More details: https://www.dataiq.co.uk/home

Interview	Size	Sector(s)	Data type(s)
2	Small	Information and Communication; Arts, entertainment and recreation; Other service activities	Customer behaviour; Activity data
3	Medium	Arts, entertainment and recreation; Environment; Human health and social work activities	Grant data
4	Large	Manufacturing; Professional, scientific and technical activities	Customer behaviour
5	Small	Arts, entertainment and recreation; Information and Communication; Other service activities	Customer behaviour; Activity data
6	Large	Retail; Professional, scientific and technical activities	Contact information; Customer behaviour; Stock and supply data; Website traffic data
7	Large	Professional, scientific and technical activities; Administrative and support service activities; Education	Socio-demographic data; Other (research/survey data)
8	Small	Accommodation and food service activities	Customer behaviour; Sales or transaction data
9	Large	Professional, scientific and technical activities; Education	Contact information; Customer behaviour; Sales or transaction data
10	Large	Professional, scientific and technical activities; Information and Communication services; Administrative and support service activities	Biometric data
11	Large	Professional, scientific and technical activities; Information and Communication services; Administrative and support service activities	All data types (stakeholder is a cloud provider)

A3.3   Preamble

The Department for Digital, Culture, Media and Sport (DCMS) has engaged London Economics (LE) to produce an understanding of costs facing businesses when sharing data. This research will be used to add to DCMS' understanding of the costs associated with data use, and will assist DCMS in supporting innovation from organisations making data available to users outside their organisations.

The results of this interview will be used to provide additional understanding on costs and barriers facing your organisation and the wider sector regarding data sharing – your input to this study is really appreciated, and will help inform any future policy interventions.

Whilst this interview will be recorded, your responses will be anonymised and unattributable in the final report – unless permission is granted otherwise.

Consent

Do you consent for this interview to be recorded? These recordings will only be used internally by London Economics to assist with their research and analysis for this project.

London Economics' privacy policy can be found: https://londoneconomics.co.uk/privacy-policy/

A3.4   Topics to explore

Background and organisation

- ■ Can you provide some background on yourself and the organisation? What are your business' core activities?
- ■ What is the size of your organisation? (i.e., *employment / revenue figures*)
  - ☐ Micro SME (Less than 10 employees and less than £2m turnover)
  - ☐ Small SME (Less than 50 employees and less than £10m turnover)
  - ☐ Medium-sized SME (Less than 250 employees and less than £50m turnover)
  - ☐ Large firm (More than 250 employees or more than £50m turnover)
- ■ Where are your headquarters based / what markets do you serve?
  - ☐ *Interviewer prompts:*
    - *North East*
    - *North West*
    - *Yorkshire and the Humber*
    - *East Midlands*
    - *West Midlands*
    - *East of England*
    - *London*
    - *South East*
    - *South West*
- ■ What sectors do you operate in?
  - ☐ *Interviewer prompts:*
    - *(Answer using following SIC Code if possible, further details found here)*

Data use and lifecycle position

- ■ What types of data do you hold that you consider potentially valuable to other parties?
  - ☐ Data on customers
  - ☐ Sales data
  - ☐ Monitoring data
  - ☐ Sensor data
  - ☐ HR + payroll data
  - ☐ Other
- ■ Where is this data sourced from?
  - ☐ *Interviewer prompts:*
    - – *Gained through consent or purchased from databases or lists used for outreach/ direct marketing*
    - – *Observed in-house (i.e., on your organisation's website)*
    - – *Procured from third party data providers (i.e., data processors/ online platforms)*
    - – *Other (please specify)*
    - – *Don't know*
- ■ Describe this data's characteristics
  - ☐ *(Use the following as prompts to ascertain data characteristics, but do not go through each entry in order. Aim of question is to get a qualitative understanding of the data) :*
    - – *Type (e.g., image, text, transaction records, machine sensor data)*
    - – *Category (personal, commercial, medical etc.)*
    - – *Volume (size of the data)*
    - – *Periodicity (frequency of data compilation)*
    - – *Latency (time taken for data to be available/ transferred)*
    - – *Any other attributes*
- ■ Does your organisation make this data available to users outside your organisation *(i.e. data sharing)?*
- ■ Who uses this data outside your organisation? *What types of users are your data shared with?*
- ■ What benefits does this data deliver? (*i.e, sell or license raw/aggregate data, development and selling of new products, improvement of existing products, improvement of existing business processes, reputational/PR benefits, marketing outreach)*
  - ☐ For your organisation
  - ☐ For users outside your organisation (external users)
  - ☐ For third parties / non-users of data
- ■ Why does your organisation make this data available/ does not make this data available? *(The motivation for sharing/ not sharing data)*
- ■ Have you considered making your data (more) available to users outside your organisation?
  - ☐ *Interviewer prompt: The ODI data spectrum illustrating data from closed to open is attached to the end of this guide; this can be shown to participants on screen. (Section 1.3.1)*
- ■ Are you aware of existing external demand or use cases for data your organisation has?

■ Do you know what types of users could create/gain value from access to your data?

Data sharing barriers and costs

Regarding costs of data sharing, we currently group together costs in four buckets (see also Section 1.3.3):
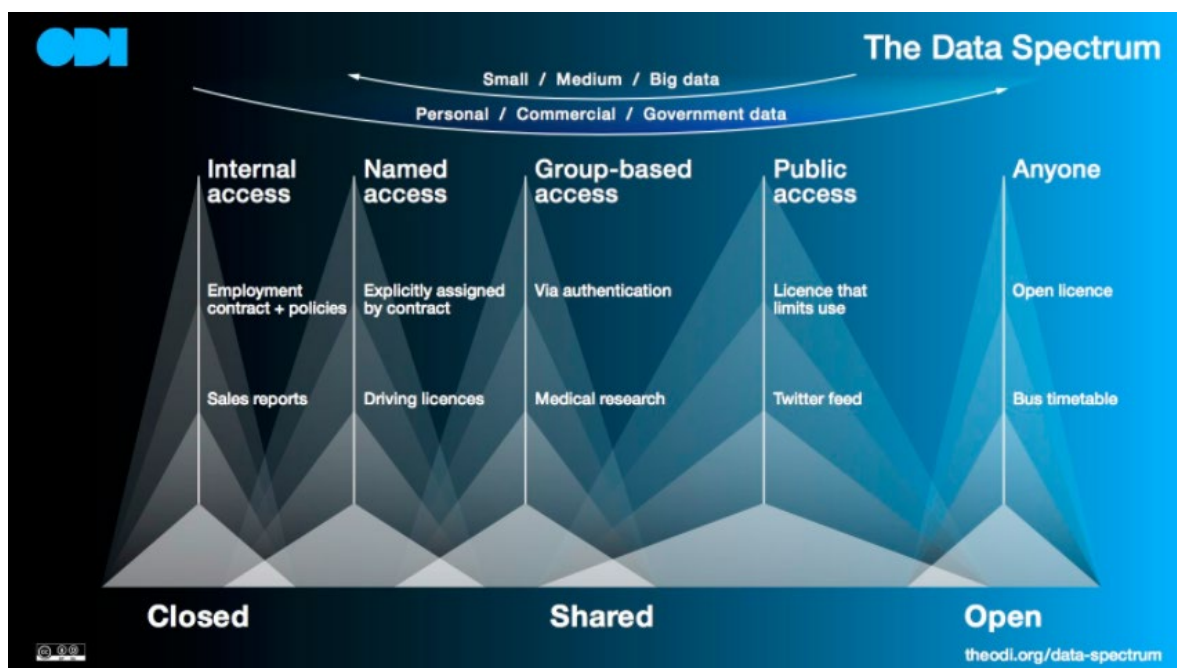
1. Data sharing strategy – organisations need to develop a strategy for data sharing, including what data will be useful to be shared, how it can be valued and the various structures that can be used to share it
2. Legal and regulatory considerations – organisations need to understand compliance requirements for the type of data shared, and how to structure the sharing agreements
3. Technical and organisational considerations – any technical or organisational costs needed to share data, such as software and data preparation costs
4. Operational costs – additional operational and ongoing considerations such as costs to maintain data quality
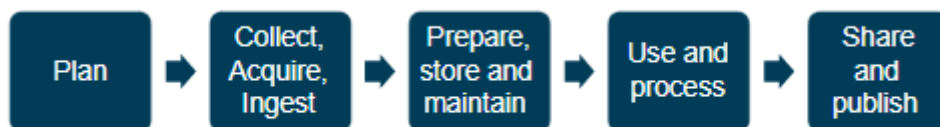
Keeping these costs in mind:

■ Are you/ your organisation aware of - or familiar - with the types of costs described above?
   ☐ *Interviewer prompt: If these organisations are unfamiliar with these costs, use taxonomy prompts to add additional detail. (Section 1.3.4)*
■ *For each bucket of costs identified*: How do these costs interact with your data sharing processes?
   ☐ *Interviewer prompts – below is a summary of our fuller taxonomy. The taxonomy will not be given to respondents but will be kept for interviewer for prompts*
      – For technical considerations – have you had to hire or train additional staff?
      – For legal and regulatory considerations – have you had to pay lawyer fees to help understand consideration?
      – For strategy – have you had a cost for data collection or coordination?
      – For operational costs – do you have an ongoing cost for ensuring data quality is maintained?
■ Thinking back to the data lifecycle (pictured here), can you describe where along this lifecycle these costs are incurred?
   ☐ *Interviewer prompts – Section 1.3.2*
■ *For each of these costs*: How have these costs materialised in your process? For example, additional staff hired or lawyer fees.
   ☐ *From a quantitative perspective*: Can you estimate the £ value of these costs? For example, salaries of additional staff or fees paid?
   ☐ Can you portray these costs as a relative magnitude of your overall business activities? For example, what % of your workforce are new potential staff, what % of revenue do these £ costs represent?
■ *To help understand the relative magnitude of these costs*: Can you comment on which of these costs are most significant in your data sharing processes? Could you rank these costs?
■ What would encourage your organisation to share (more) data with a third party?

A3.5　　Interviewer materials

A3.5.1　　ODI Data Spectrum

A3.5.2　　Data Lifecycle

- The following stages describe the data lifecycle, and where/how data is used.
  - ☐ Plan – This stage is where an organisation (intending to collect, store and use data) must plan their processes and data storage. This stage involves determining business needs, identifying existing data and what needs to be collected.
  - ☐ Collect, acquire, ingest – This stage involves the acquisition and collection of data.
  - ☐ Prepare, store and maintain – This stage involves the preparation of data for storage, including formatting of data for later use, as well as storing this data appropriately.
  - ☐ Use and process – This stage involves the use of data for your organisations business needs, this could include analysis of data or production of outputs.
  - ☐ Share and publish – This stage involves preparing data for publication and sharing, as well as sharing data with users outside your organisation (whether this data is shared/ sold for monetary gain or otherwise)

　　　　　　　　　　　　　　　　　Research into cost considerations of data sharing

A3.5.3 Taxonomy framework
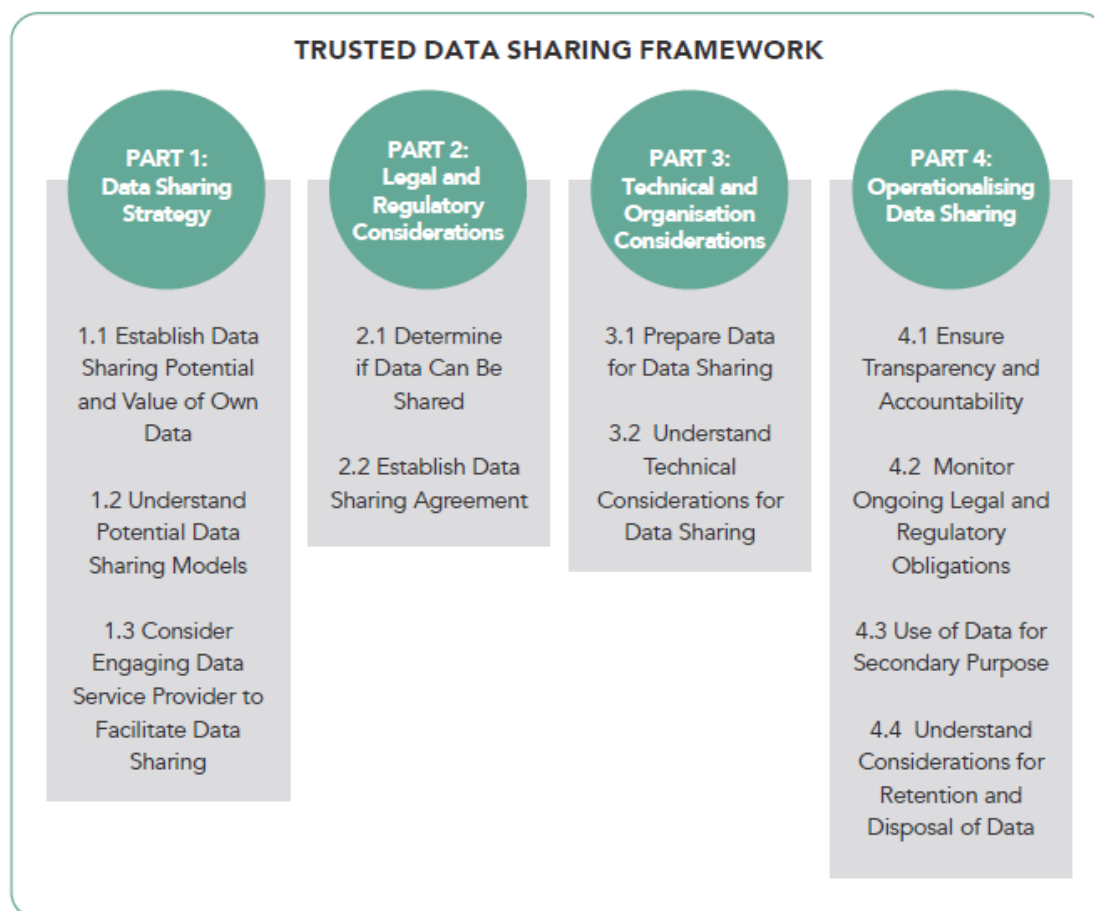
Figure 1: Trusted Data Sharing Framework and key outcomes

A3.5.4 Taxonomy

Explicit costs are costs that organisations experience directly, and are likely to be present in an organisations balance sheet. An overview of these costs can be found in Table 23.

Table 23   Overview of the explicit costs in data sharing

Cost	Description
Data sharing strategy	Part 1 in data sharing framework
Planning and coordination costs	Cost of identifying data to be shared and coordinating among various internal and external stakeholders
Data collection costs	Cost involved in collecting data that is to be shared. This applies in cases where additional data must be collected for the purpose of sharing, and excludes data collection required for business processes.
Procurement costs	Cost of purchasing data from a third party/vendor.
Legal and Regulatory considerations	Part 2 in data sharing framework
Compliance costs - collection	Ensuring that the collection of data to be shared is compliant with regulation.

Cost	Description
Compliance costs - sharing	Ensuring that data can be shared, and that it will be shared in a way that adheres to GDPR and other legislation.
Data sharing agreements	Administrative and legal cost of creating contracts between data sharers and data receivers.
Technical and organisational considerations	Part 3 in data sharing framework
Software costs	Software that may be required for data collection, cleaning and sharing, including the cost of licenses, as well as the cost of developing programmes in-house.
Security costs	Administrative cost of ensuring that the interface used for sharing is secure and the risk to the company is limited.
Storage costs	Cost of storing the data to be shared (on-premise or cloud).
Data preparation costs	Cost of cleaning data, preparing metadata and formatting data to adhere to standards.
Operationalising data sharing	Part 4 in data sharing framework
Data quality monitoring	Ongoing monitoring of data quality.
Data sharing interface maintenance	Ongoing maintenance of interface used for sharing.

Source: London Economics

The **implicit costs** of data sharing are related to the inexplicit costs experienced by organisations (or external stakeholders) as a result of data sharing. These costs are discussed in the economic literature on data sharing, but do not explicitly appear on the balance sheets of organisations.

Table 24   Overview of implicit costs related to data sharing

Cost	Description
Competitive disadvantage	Loss to a business if data leads to a loss in competitive advantage.
Negative externalities	Negative externalities occur when data is shared (for example, if data sharing leads to undesirable consequences for the individuals whose data is being shared).
Reputational cost	Firms that share data may experience a reputational cost if the nature of the data is sensitive, or if there are other objections from customers/partners/other stakeholders.
Cost of cyber security breaches	Firms may incur high losses if they are subjected to a hacking attack facilitated or enabled by the sharing arrangement.

Source: London Economics

Annex 4   Key findings from stakeholder interviews

This annex provides additional identified costs and detail from the stakeholder interviews with firms that hold digital data

A4.1   Data strategy costs identified from the interviews

- For a small firm, dealing and sharing customer behaviour, events and booking data and website search data: their time costs involved with understanding potential for sharing and models to use; **3 – 5 days** of management time.

- For a small firm within the arts, entertainment, and recreation sector, also sharing customer behaviour and activity data, data sharing strategy was embedded in their business model and seen as a business-as-usual cost.

- For a medium firm sharing data on research grant availability and outcomes, development of a data strategy takes place as an annual activity with 10 days of staff time used per year. According to this interviewee, this is equivalent to approximately **£300/day** - half of this cost is related specifically to strategies around sharing data.
  - ☐ There are no external costs in the development of this strategy, but ad-hoc projects are undertaken to look at specific elements of data sharing strategy. These projects approximate **£30,000** in expenditure per year (this cost is split between the planning and sharing stages in the lifecycle).

- For a large firm collecting customer behaviour data, strategic thinking is related to potential value of the individual projects. Time spent on understanding and strategizing data sharing implications is negligible as existing procedures and policies are already in place (costs sit at plan end, but also at use and process).
  - ☐ These costs are mostly incurred in the 'Plan' stage of the data lifecycle, but also at the 'Use and process' stage.
  - ☐ Main financial cost is the purchase of data from other third parties such as brokers.

- A small firm sharing customer behaviour and events and booking data commented that they spent the first 6 months of its set-up establishing data sharing potential and value using a grant of **£40,000 - £50,000**
  - ☐ An additional **€100,000** grant was used to explore commercialisation options for their data sharing activities.
  - ☐ **£50,000** in staff costs are spent per year on engaging with data sharing partners.
  - ☐ These costs are mostly incurred in the 'Plan' stage, but also in the 'Prepare', 'Use' and 'Share' stages due to the integrated nature of their business model.

- A large marketing firm attached to a UK retailer stated that time was spent planning specific data sharing activities and considering the data sharing implications of their wider data strategy. However, this was not something they could measure as these costs were built into their broader operating costs.
  - ☐ The key data types shared included loyalty scheme customer data as well as retail purchases.

- A large research institution working with survey/research data and customer data did not share data, but commented that the administrative burden associated with gathering data into a shareable format would amount to **2 months of a half-time role @ £20 per hour.**

■ The owner of a small restaurant chain commented that their strategy cost is their biggest cost area. This is due to fees paid to a third party provider to inform and aid this organisation with commercialising their data.

■ A large training provider and IT consultancy sharing learner data, course data and client data (of firms they consult for) commented that these costs were difficult to quantify, but suggested that **30% of the time** spent by internal analysts were related to dealing with requests and enquiries related to strategy and planning costs.

■ A large software provider working with biometric data said that if they were to share data, there would be **significant costs** inherent with convincing customers to share personal data, as well as solving coordination problems between these data controllers (as many controllers would not want to be the first to agree to share this sensitive data).

A4.2   Legal and regulatory costs identified from the interviews

■ For a small firm, dealing and sharing customer behaviour, events and booking data and website search data: drawing up data sharing agreements involved **£3,000** of legal fees, **£2,000** of internal time amending internal privacy policy and terms and conditions

■ For a small firm within the arts, entertainment and recreation sector, also sharing customer behaviour and activity data, data sharing has a considerable overhead and ad-hoc costs associated with maintaining compliance (in terms of staff time).

■ For a medium firm sharing data on research grant availability and outcomes, Legal and regulatory costs are very limited due to the lack of personal data involved.

   ☐ Main cost was updating policies and procedures to be GDPR compliant: internal work required costed at **20 days @ £300** per day

■ For a large firm collecting customer behaviour data, they commented that their organisation considers the implicit cost/risk of what the receiving party might use the data for, which impacts their choice of strategic partners (unless they can exert considerable legal controls on the use of data, they will not share) – these costs are incurred in the planning stages.

   ☐ For individual agreements to exchange data, it can take months to finalise legal agreements. These are one-off costs, which are small related to the overall budget.

■ A small firm sharing customer behaviour and events and booking data commented that they paid **£25,000** to external legal services to set up data sharing terms

   ☐ They also had on-going ad-hoc costs to update agreements in response to partner requests but these costs were considered marginal.

■ A large marketing firm attached to a UK retailer commented that legal costs were a relatively small proportion of total costs associated with data sharing. Their primary costs are staff related. They have a data governance team - which evaluate these costs on a case-by-case basis – costing **£1-2 million a year**.

   ☐ External lawyers and advisors are used, but these costs are considered lower and more difficult to quantify.

   ☐ These costs sit within the 'Prepare', 'Use' and 'Share' stages of the lifecycle.

■ A large research institution working with survey/research data and customer data did not share data, but if they did share data, they would need to re-contact participants to gain permission to share data and consult with in-house legal advisors.

   ☐ This would take roughly **4 hours** of time.

■ The owner of a small restaurant chain commented that **legal costs are their biggest cost area with regards to data sharing**. They remarked there is an important up-front cost (in terms of ensuring compliance and finalising data sharing agreements), but that this cost diminishes over time.

■ A large training provider and IT consultancy sharing learner data, course data and client data (of firms they consult for) commented that their legal counsel spends time scrutinising contracts and data sharing agreements - this cost representing **4 FTE** of staff time. This work is carried out for all data collection and usage aspects and is required even if data was not being shared.

  ☐ **10% of these legal costs** could be specifically attributed to sharing data.

■ A large software provider working with biometric data said that if they were to share data, there would be **significant legal costs** (especially at the outset of deciding to share data). They state that reputable companies will want to do a full legal due diligence exercise before embarking on a data sharing process.

  ☐ These fees can be **very significant** and can continue to be accrued as additional products roll out

A4.3 Technical and organisational costs identified from the interviews

■ For a small firm, dealing and sharing customer behaviour, events and booking data and website search data they built APIs for sharing data in real time. This cost them 1 week's work (**£1,000 - £2,000**).

  ☐ They also had an additional cost incurred from paying for external services to compress event's data for sharing and firm activities. This cost **£800** per month but was also needed for internal operation of business so was not strictly an additional cost.

■ For a small firm within the arts, entertainment, and recreation sector, also sharing customer behaviour and activity data, they had made significant investment in open standard technology (**£30,000 – £50,000** per year for 6 years). This infrastructure also requires regular maintenance to ensure continued data sharing.

■ For a medium firm sharing data on research grant availability and outcomes, they had a number of technical costs incurred from paying for external management systems, software and solutions. Their data management system used to collect data costs **£8** per user. With 40 employees, this equated to just under **£4,000** per year – these costs are incurred in the 'Collect, Acquire, Ingest' stage

  ☐ Their data warehouse solution also costs them **£100** per month (Prepare, Store and Maintain).

  ☐ **£15,000** per year is spent on a business intelligence tool (Use and process).

  ☐ Additional ad-hoc project work is also sometimes carried out to enhance data sets for sharing – for example, they paid **£20,000** last year for a project which uses ML to classify charity activities.

  ☐ There are no additional staff employed to specifically deal with data sharing. However, around 40% of the time spent by their staff is said to be related to data sharing specifically.

■ For a large firm collecting customer behaviour data, they commented that technical and organisational costs are involved to enrich their data before sharing.

- A small firm sharing customer behaviour and events and booking data commented that they had to pay a developer to develop prototypes for their data sharing platform using part of a **€100,000** grant.

  ☐ Software licences are also required to run the overall platform which costs an estimated **£3,000 - £5,000** per month.

  ☐ Some of these software costs also cover data security - this is supplemented by an additional cost of employing a third-party provider to manage data security, although the cost of this was not known.

- A large marketing firm attached to a UK retailer stated that the technical costs were their most significant cost overall – accounting for 70% of the total costs of data sharing.

  ☐ This business has invested significantly in new infrastructure and platforms to share data. These costs include staff costs and software licensing costs.

  ☐ The financial costs of this investment are "in the area of" tens of millions of pounds - with about 10% of this investment attributed to costs associated with sharing.

  ☐ However, these high levels of investment would not be maintained long term and is a function of the business' growth and development plans.

- A large research institution working with survey/research data and customer data did not share data but commented that data infrastructure would need to be established to share data – specifically costs for data storage.

- The owner of a small restaurant chain commented that these costs are the least significant of the four classifications of costs for them. Their business is small with relatively simple IT processes, which are largely managed by a third party.

- A large training provider and IT consultancy sharing learner data, course data and client data (of firms they consult for) commented that there are considerable costs with storing and preparing data – including software licences, data analyst staff costs and external consultants. However, these costs are built into the business model and not considered as additional costs.

  ☐ Sharing data is an increasingly important part of their business model, with additional investments of roughly £500,000 – £1 million being made for staff time, encryption, storage, and external consultants.

  ☐ These costs are incurred in the 'Plan', 'Collect' and 'Use' stages of the lifecycle.

- A large software provider working with biometric data said that if they were to share data, technical and organisational costs would continue to be their highest costs due to the nature of tech development. Software engineer salaries are very expensive and for their data sharing idea to happen, they would ned to increase their team by **20-25%** (although they are willing to incur this cost if the data sharing materialised due to this becoming a significant new focus for the company).

  ☐ For reference, their current product and tech team is roughly 250-260 staff out of total employment of 600.

A4.4   Operational costs identified from the interviews

- For a small firm, dealing and sharing customer behaviour, events and booking data and website search data: there were some staff costs associated with maintaining APIs to ensure quality of data shared in real time is accurate and of high quality. They estimated this at **£1,000** per year

- ☐ Cost of maintaining database of business customers' facilities also equated to around 1 day of staff time per month

- ■ For a small firm within the arts, entertainment and recreation sector, also sharing customer behaviour and activity data, new technology is sometimes developed around new programmes, and part of that development cost is spent on ensuring data standard compliance. This respondent estimates this at 10% to 20% of the total cost of data sharing.

- ■ For a medium firm sharing data on research grant availability and outcomes, they fund a third-party provider to the value of **£100,000** per year, which enables the sharing of data between (grant) funders and other parties to take place. On the data lifecycle, this cost was said to sit within 'Share and Publish'

  - ☐ The technical back-end work to maintain the data sharing system is carried out by a third-party provider. This firm estimated that **£1,000** per year is spent on maintenance of systems directly in relation to data sharing. On the data lifecycle, this sits in the 'Share and Publish' stage. It was noted that this is a small proportion of wider spend on maintaining the data platforms overall (which is **£100,000** per year).

- ■ For a large firm collecting customer behaviour data, the operational costs involve ensuring the shared data is handled correctly by the recipient – but these are considered operational to the firm's activities and could not be split out from other non-sharing related costs.

- ■ A small firm sharing customer behaviour and events and booking data commented that the cost of paying their development team is **£350,000** per year. This covers the costs associated with keeping their data sharing platform running, monitoring its efficiency, and maintaining the quality of data.

  - ☐ This cost sits in 'Collect, Acquire, Ingest', 'Prepare, Store and Maintain', 'Use and Process' and 'Share and Publish', due to the integration of data sharing as part of their business model.

- ■ A large marketing firm attached to a UK retailer stated some costs related to technical considerations could be considered operational costs. Some staff costs and software costs could be allocated to this bucket, but it was not feasible to put separate costs against these activities.

  - ☐ These costs sit in the 'Collect', 'Prepare', 'Use' and 'Share' stages of the data lifecycle.

- ■ The owner of a small restaurant chain commented that this was their **second most significant cost related to data sharing** – this cost is incurred from ensuring that data is high quality, stored securely and compliant with changing regulations.

- ■ A large training provider and IT consultancy sharing learner data, course data and client data (of firms they consult for) commented that the most significant operational cost was associated with ensuring the quality and maintenance of data sharing interfaces, as well as maintaining quality of data shared (**5% of time could be attributed directly to ensuring data is suitable for sharing**).

- ■ A large software provider working with biometric data said that if they were to share data, resources would shift away from strategy-related costs and towards operational-related costs.

Annex 5   Literature review methodology

We used a **systematic literature review approach** as part of the literature identification, evidence review and subsequent taxonomy creation. This process involved the following steps:

1) Catalogued papers identified from initial trawl of relevant papers detailed in the proposal.

2) Catalogued papers and sources shared by DCMS.

3) Conducted a Google Scholar search for academic papers, and a Google search for policy papers.

   a) Key words used for these searches: costs, barriers, challenges, incentives, value, open access, data sharing, open data.

   b) Catalogued papers deemed relevant following scan of abstract.

4) Scanned citations lists of catalogued papers to identify additional relevant papers and sources.

5) After the list of papers had become exhaustive, and the above steps failed to find further relevant papers, the catalogued papers were reviewed and their relevance to the research questions was assessed.

6) This list of relevant sources was then further analysed and synthesised to form the taxonomy of costs.

In a spreadsheet, the following characteristics were identified for each paper/report:

- Summary of the paper:
  - ☐ Short abstract
  - ☐ Nature of the paper (empirical, statistical, qualitative etc.)
  - ☐ Methodology of the paper (where relevant)
- Key themes / findings:
  - ☐ The identified barriers and challenges preventing data sharing
  - ☐ The types of data sharing costs discussed
  - ☐ Where along the lifecycle the cost is incurred?
  - ☐ Factors affecting the cost/barrier
  - ☐ Any sector or geographical information about where the research was conducted
  - ☐ Size of the organisation affected by these costs
- Assessment of the relevance of the paper to the study's research questions

These findings were synthesised to produce a taxonomy, detailing the categorisation of the cost, where in the lifecycle the cost is accrued, factors influencing this cost, any previous estimates of this cost, and the sources substantiating this identified cost.

This spreadsheet (containing the above information) was provided alongside this report.

Research into cost considerations of data sharing

Annex 6 Taxonomy and literature review

This section details the findings of the identified literature and the subsequent resulting taxonomy of costs.

A6.1 Market failures in the market for data

Many of the barriers to data sharing for firms stem from market failures in the market for data. This section will describe the economic characteristics of data, the market failures that arise due to these characteristics and how these may create barriers to data sharing.

Market failures occur when goods and services are inefficiently distributed in the free market, and can lead to over- or under consumption of resources. The presence of market failures generally provides a rationale for policy intervention. A range of market failures may be present within the market for data, which could deter organisations from sharing, or reduce the benefits that are generated through sharing.

Data is non-rival, which means that it can be used by multiple parties without depleting or losing its value, and it can be **excludable**, if it is possible to prevent individuals from accessing or collecting it (Coyle et al, 2020). If data is both non-rival and non-excludable can be classified as a **public good**. This is true for environmental data for example. Data on air or water quality can be collected by anyone with the right equipment. If it is possible to exclude people from accessing or collecting data, it can be classified as a **club good**. This is the case for administrative data which is only collected by those delivering a service.

Barriers to entry arise when firms who want to enter the market struggle to do so because they lack information about customers. Entrants into a market cannot compete effectively if consumer data is a significant factor. This barrier may be amplified when there are high up-front costs to data collection but low marginal costs (Coyle et al, 2020). Organisations are then deterred from entering that market and competition is limited.

Market failures can arise when there is **asymmetric information** between organisations, or when information asymmetries exist between organisations and their consumers (Ctrl Shift, 2018). Where the value of data is better understood by data driven businesses than by "analogue" businesses, the data driven businesses can prevent others from accessing it and creating value from the market opportunity. Information asymmetries also exist between businesses and the consumers from which data are obtained. Consumers may be unaware of how much data is being gathered, who it is being shared with and what it is being used for. In either case, asymmetries lead to worse outcomes for the market participants that have a poorer understanding of the data.

Principal-agent problems occur in data sharing when the incentives of the organisations that collect data, the third parties who access the data and the consumers from whom data is gathered are not in alignment (Coyle et al, 2020). Related to the principal-agent problem, frictions arise from 'incomplete contracts' when the agreements between data sharers and data users don't specify how data can be used in every possible contingency (London Economics, 2019). This presents a hold up for firms.

Coordination failures occur where market participants fail to align on issues, for example on data standards, and are therefore unable to share data (Ctrl Shift, 2018). Coordination failures can also occur when organisations are unaware of other parties who may be interested in the data that they collect, or when businesses who are interested in using data do not have information about who collects it. Furthermore, businesses are used to treating data as a resource to be protected, not shared, which may prevent them from seeking or engaging with potential data sharing partners.

Data may have positive or negative externalities[25] (Coyle et al, 2020; Ctrl Shift, 2018). Positive externalities arise when different datasets are linked together and thereby enhance decision making and lead to new insights. Positive externalities in this case can arise from the social or environmental benefits that occur due to the additional insights or improved decision making that can occur. Negative externalities arise when harms come about through data being collected or used, or through the amount of data being collected and stored.

When data sharing leads to positive externalities there is a danger that organisations will not share enough. This is because the benefits of sharing data may accrue to the individuals/firms that access the shared data, and not necessarily to the organisation sharing. The sharing organisation might not get enough benefits from sharing to cover the cost of collecting and storing the data in the first place (Coyle et al, 2020).

When negative externalities arise, there is a chance that organisations will share too much data, as the costs of the harms generated through sharing are not felt by the sharing organisation. In either case the market may not provide the best overall outcomes.

A6.1.1   Other barriers to data sharing

In addition to the market failures described above, there may be other barriers to data sharing that are not directly related to the economic characteristics of data.

Firms often don't see a **clear business case for data sharing** (Ctrl-Shift, 2018). This is either because the returns to sharing data for the firm are uncertain (Frontier Economics, 2020), or because firms do not see what value their data may have to an outsider.

Especially in smaller organisations the preconditions for data sharing, such as the hardware, software and skills required, are often lacking (London Economics, 2019). In these companies and ones that do not rely heavily on data, it may be more difficult to use and share data for value creation (London Economics, 2019). In particular, the **lack of technical knowledge** about the uses of data or what data is available may hamper data sharing (London Economics, 2019; Frontier Economics, 2021).

The costs of data sharing are not linear throughout the sharing journey. There are **high upfront costs** to data sharing, which may deter companies if there is no clear business case for sharing (Frontier Economics, 2021). These costs may be related to the technical set-up costs, to the internal coordination required such as identifying relevant datasets, the stakeholders involved and negotiating access, which can be difficult and resource-intensive (London Economics, 2020).

---

[25] In economics, externalities refer to the consequences of an industrial or commercial activity that are not reflected in market price.

Firms experience additional barriers to sharing data when the data relates to individuals. This is both because of the risks of breaching data protection or IP rights (Frontier Economics, 2021), and because of consumers reactions to their information being shared. Organisations face concerns about privacy from their customers (Acxiom and Foresight factory,2019), and collecting and sharing their data may lead to a lack of trust and engagement from customers (Data and Marketing Association, Acxiom and Foresight Factory, 2018).

A6.1.2 Expert interviews

The expert interviews, conducted with four data sharing experts, revealed information on additional barriers and costs facing businesses who wish to share data. Due to the low number of interviews, these findings are not intended to represent the views or experiences of all companies, but to corroborate the points found in the literature. As such, the findings from the expert interviews should be considered holistically and together with other evidence and findings. Where possible, these findings are signposted with relevant literature supporting these.

Respondents in the expert interviews commented that the barriers preventing data sharing include:

- Lack of digital literacy[26]
- Lack of awareness of the value of data sharing[27]
- Lack of knowledge about strategic reasons for data sharing[28]
- Perception that data sharing lowers competitive advantage[29]
- Concern over the risks of data breaches[30]

The expert interviews indicated that organisations making the decision to share data usually approach the process by establishing a business case, or by determining a clear benefit or value from this data sharing; therefore, organisations often do not look at data sharing primarily through the lens of cost consideration. If the business case can be justified, costs are unlikely to be sufficiently high to be an impediment preventing data sharing. The main problems preventing data sharing include establishing a business case, structuring data sharing relationships, and regulatory uncertainty;[31] regulatory uncertainty and failure to establish data sharing relationships between market participants may be an example of **coordination (market) failures**. Concern about misuse of data may also be evidence of **principal-agent problems** occurring from 'incomplete contracts', as well as an awareness of potential negative **externalities** arising from misuse and data-breaches.

Organisations choosing to share data do so for the following reasons:

- Organisational responsibility, sharing data to help improve innovation within the sector, or to gain trust.
- Commercial, transactional, or financial reasons.

---

[26] Problems related to lack of hardware, software and skills detailed in London Economics, 2019; Frontier Economics, 2021

[27] Coordination failures detailed in Ctrl Shift, 2018

[28] Asymmetric information failures detailed in Ctrl Shift, 2018. Lack of knowledge of reasons to share data, as well as demand for data can prevent organisations from sharing this data.

[29] Fear of loss of competitive advantage detailed in London Economics, 2019

[30] Cost of cyber security breaches and reputation costs detailed in Acxiom and Foresight factory, 2019; Frontier Economics, 2021

[31] Corroborated in Ctrl-Shift, 2018; Frontier Economics, 2020

■   Required due to regulation and licence conditions.

Costs which face businesses data sharing include:[32]

■   Legal and regulatory costs such as the time costs of compliance and ensuring data architecture is secure and compliant.[33]

■   Technical costs, such as setting up infrastructure, ensuring data is clean and ensuring that this data makes sense for users outside the organisation.[34]

■   Strategic and planning costs such as identifying data sources, planning data sharing strategy and validating business cases.[35]

■   Operational costs such as ensuring ongoing data quality and maintenance of sharing interface and architecture.[36]

Data sharing agreements often fail to materialise from the data sharer having low awareness of the full range of costs involved with data sharing. Organisations with lower levels of digital literacy, skills, processes, and infrastructure often agree to share data, but then pull-out of agreements after realising the full scope of costs. For example, data collected for internal usage may require additional licensing agreements if this data is to be shared or transformed for commercial use. This **information asymmetry** between data sharing participants may be a cause for these agreements – and subsequent data sharing – failing to come to fruition.

Factors which affect these costs include:

■   The size of the company – smaller businesses will often not have the skill set in-house necessary to set up and maintain data sharing infrastructure.

■   Size and complexity of the data shared – including characteristics such as whether the data is personal.

■   How 'digital' a company is, a company that is tech savvy – with a lot of data-related skills – will not struggle with much of the technical costs related to data sharing.

■   Nature of access of this data, such as full or partial access, or a necessity for real time data.

■   How this data is shared – data shared through an API may be expensive.

Raising the awareness of the value of data sharing – along with testimonials – would help increase management buy-in and increase awareness of the associated costs of data sharing. The lack of awareness of costs is also sometimes a barrier preventing data sharing – this low awareness affects businesses effectively forming the business case to share data. These factors are an example of continued **information asymmetry** market failures – a lack of awareness of the value of data between data controllers and users which can attain value from this data hampers the rate of data sharing.

---

[32] The costs below reflect the taxonomy classifications, further sources and details on the underlying literature can be found in the project taxonomy spreadsheet.

[33] Related costs detailed in Coyle et al, 2020; McKinsey, 2013

[34] Related costs detailed in OECD, 2019; OECD, 2020; Bald, 2020; Figueiredo, 2017

[35] Related costs detailed in London Economics, 2019; Frontier Economics, 2020

[36] Related costs detailed in Figueiredo, 2017

A6.1.3   Data lifecycle

All data used for productive purposes will go through a similar journey through organisations.[37] This journey is described by the data lifecycle, depicted in Figure 36. While the actual lifecycle of data will be specific to an organisation, the framework is a useful tool to identify points where data quality issues may arise.

The first stage in the lifecycle is the **planning stage**, which involves determining business needs, identifying existing datasets and what needs to be collected. During the second stage, data is **collected or acquired, and ingested**. Data then needs to be **prepared, stored and maintained** so that it can be useful to the organisation. During this stage it is important that consistent standards are applied and that data is cleaned and linked with other records where applicable. The fourth stage of the data lifecycle is related to the **use and processing** of the data, which includes exploration and analysis, and production of outputs. Data can also be **shared and published** where appropriate for secondary purposes. Finally, when data is no longer in active use, the data owner should determine whether it should be **archived or destroyed**.

Figure 37   The data lifecycle

Source: Government data quality hub (2020) Government data quality framework, available at: https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework#The-Data-Lifecycle

This research intends to explore the costs that are incurred as a result of the fifth stage of the data lifecycle: share and publish. While the data lifecycle can be mapped out into separate stages and presented in a linear form, the costs that arise from data sharing are often present in multiple stages of its lifetime, and therefore cannot be grouped this way. For example, software may be needed to collect data, to prepare and maintain it, as well as to use, process and share it. The licensing costs of software can therefore not be attributed to a specific stage in the data lifecycle.

A6.2   Taxonomy of data sharing costs

The Infocomm Media Development Authority (IMDA) has developed a useful framework for data sharing which may be used to group/categorise sharing costs. The framework describes the considerations for firms that wish to share data (IMDA, 2019).[38] According to the framework, there are four parts to the data sharing journey:

---

[37]   Government data quality hub (2020) Government data quality framework, available at: https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework#The-Data-Lifecycle

[38] The framework was developed as a planning tool for organisations and has therefore needed to be modified slightly to be appropriate for the cost considerations.

- ■ **Data sharing strategy**: organisations will need to develop a strategy for data sharing, which should include what data will be useful to be shared, how it can be valued and the various structures/interfaces that can be used to share it.
- ■ **Legal and regulatory considerations**: organisations need to understand the compliance requirements for the type of data that is to be shared, and how to structure the sharing agreements in order to facilitate trusted data sharing between parties (if data is not shared openly).
- ■ **Technical and organisational considerations**: organisations need to understand the technical considerations and mechanisms required for sharing data with external stakeholders.
- ■ **Operationalising data sharing**: there may be additional considerations for the operationalisation of data sharing and after data sharing has taken place.

A6.2.1   Summary of the costs of data sharing

The costs of data sharing, as identified in the literature and through expert interviews with stakeholders, have been split into explicit, and implicit costs. The explicit costs have been grouped based loosely on the trusted data sharing framework described above.

Explicit costs are costs that organisations experience directly, and are likely to be present in an organisations balance sheet. An overview of these costs can be found in Table 25.

Table 25   Overview of the explicit costs in data sharing

Cost	Description
Data sharing strategy	Part 1 in data sharing framework
Planning and coordination costs	Cost of identifying data to be shared, and coordinating among various internal and external stakeholders
Data collection costs	Cost involved in collecting data that is to be shared. This applies in cases where additional data must be collected for the purpose of sharing, and excludes data collection required for business processes.
Procurement costs	Cost of purchasing data from a third party/vendor.
Legal and Regulatory considerations	Part 2 in data sharing framework
Compliance costs - collection	Ensuring that the collection of data to be shared is compliant with regulation.
Compliance costs - sharing	Ensuring that data can be shared, and that it will be shared in a way that adheres to GDPR and other legislation.
Data sharing agreements	Administrative and legal cost of creating contracts between data sharers and data receivers.
Technical and organisational considerations	Part 3 in data sharing framework
Software costs	Software that may be required for data collection, cleaning and sharing, including the cost of licenses, as well as the cost of developing programmes in-house.
Security costs	Administrative cost of ensuring that the interface used for sharing is secure and the risk to the company is limited.
Storage costs	Cost of storing the data to be shared (on-premise or cloud).

Cost	Description
Data preparation costs	Cost of cleaning data, preparing metadata and formatting data to adhere to standards. This includes preparing data for machine learning, such as labelling.
Operationalising data sharing	Part 4 in data sharing framework
Data quality monitoring	Ongoing monitoring of data quality.
Data sharing interface maintenance	Ongoing maintenance of interface used for sharing.

Source: London Economics

The **implicit costs** of data sharing are related to the inexplicit costs experienced by organisations as a result of data sharing (see Table 26). These costs are discussed in the economic literature on data sharing, but do not explicitly appear on the balance sheets of organisations.

Table 26   Overview of implicit costs related to data sharing

Cost	Description
Competitive disadvantage	Loss to a business if data leads to a loss in competitive advantage.
Reputational cost	Firms that share data may experience a reputational cost if the nature of the data is sensitive, or if there are other objections from customers/partners/other stakeholders.
Cost of cyber security breaches	Firms may incur high losses if they are subjected to a hacking attack facilitated or enabled by the sharing arrangement.

Source: London Economics

A6.2.2   Developing a data sharing strategy

In this part of the data sharing framework, the firms need to evaluate their current and potential data sharing capabilities and establish the value of their data. The firms must also understand the attributes of different data sharing models, and engage with various internal stakeholders. The firm can also consider engaging with a data service provider to facilitate data sharing.

The identified costs associated with this part of the framework are costs associated with planning the data sharing process, coordinating between various internal and potentially external stakeholders, and the cost of collecting or procuring the data that will be shared.

Identifying valuable datasets and coordinating with other market participants reduces the potential market failures of **information asymmetries**, **principal-agent problems,** and **coordination failures**. Harmonisation of incentives between parties participating in data sharing resulting from effective strategizing enables data sharing by reducing the barriers associated with these failures.

Planning and coordination costs

The first costs incurred in the data sharing process are related to planning and coordination. This includes the time and resource cost of identifying the data to be shared and coordinating among various internal stakeholders. Before data can be shared, it may need to be anonymised, or linked with other data. Identifying relevant data and planning how data will need to be transformed, may require coordinating with individuals that work across various stages in the data lifecycle.

During this phase firms may want to consider engaging a **data service provider**, which is an organisation that provides services and generally supports the data sharing process. The services

that can be provided by a data service provider can include, among others (i) providing technical means (e.g. platform) to facilitate data exchange, (ii) data preparation, data management and technical/ risk/ governance advisory, and (iii) acquiring data from providers, processing it to enhance its value and supplying data to the data consumers or marketplaces (IMDA, 2019).

During this part of the data sharing journey, stakeholders will also need to decide whether they will charge for access to data, and if so, what the price will be. Because of the economic characteristics of data, and specifically the fact that data is non-rivalrous, pricing data can be difficult. This is especially true in cases where there are thin markets, i.e. markets where there are a limited number of buyers and sellers, and normal price setting mechanisms do not apply (Frontier Economics, 2020). Firms may need to spend time and resources on establishing a methodology for pricing their data.

Factors that influence this cost:

- In large firms and firms with complex organisational structures, the planning and coordination costs may be higher when inputs from multiple stakeholders from different departments/workstreams are required.
- The planning process may be made more complicated when data is held in different "silos" within the firm.
- When there are clear upsides for firms to share data (for example if sharing data leads to reputational or efficiency benefits), they will not have to spend resources on data valuation.[39]

Data collection costs

In cases where firms are required to augment their data with additional data, they will need to incur a cost to collect it. This cost applies only to the additional data that is collected for the purpose of sharing, and excludes data collection costs incurred from internal data-capture processes.

Factors that influence this cost:

- Costs may be higher for the collection of personal data as individuals often have strong feelings towards the collection of personal data, and consumers might not be willing to share data with them (London Economics, 2019).
- Cost may be higher if specialist equipment (such as sensors, cameras, scanners etc.) is required for data collection

Procurement costs

In some cases, businesses may wish to buy data in order to transform it or link it with other datasets and sell it to other firms. By procuring additional data from third parties/vendors, firms can enrich data to be more valuable for others. These costs may include licensing agreements to collect data from other firms, or the cost of purchasing datasets from vendors.

---

[39] Due to lack of technical knowledge about the data economy, companies may initially struggle with the idea that the benefits of sharing open data may exceed the revenue that could be generated from sharing data. Hogge (2016) describes the challenges that arose when TfL was developing their data sharing strategy, as many internal stakeholders had strong preferences to charge data consumers for TfL data.

In addition, the time and resource costs of identifying relevant data sources, their owners and negotiating access should be taken into consideration. In cases where data is openly shared there are usually security checks in place and verification may take time. For example, most public APIs require users to apply for API keys before they can access underlying data.

Factors that influence this cost:

■ Procurement costs will vary depending on how difficult it is to collect the data.

A6.2.3 Legal and regulatory considerations

The legal and regulatory considerations include determining if the data can be shared, and the regulations that need to be adhered to in the sharing process. The size of the costs involved in this stage will depend on the type of data that is being shared. Personal and sensitive data will have significantly higher compliance costs than other types of data.

If data is going to be shared bilaterally, or only with specific individuals/firms, then data sharing agreements will need to be drawn up in this stage as well.

Regulation and laws reduce misuse of data and subsequently reduces potential **negative externalities**. Failing to comply with these considerations may lead to greater risk of these externalities – the importance of satisfying these considerations is twofold: fulfilling legislative requirements and reducing these **externalities**.

Compliance costs – collection

Organisations that collect personal data need to adhere to data protection legislation. This includes the additional costs relating to the implementation of technical aspects and processes that are followed when data is collected from individuals (Coyle et al, 2020).

Factors that influence this cost:

■ Compliance costs for collecting personal or sensitive data will be higher than if data does not relate to individuals.

Compliance costs – sharing

Compliance costs may also arise when data is shared. For data on individuals additional consent may be required and there may be additional technical considerations if data is of a sensitive nature. Where data sharing is mandatory, companies may face additional costs of uploading data at regular intervals, or following strict standards for data formats and methods of sharing.

Factors that influence this cost:

■ Compliance costs for sharing sensitive data are higher due to data protection and privacy legislation.
■ Compliance costs may be higher in industries where data sharing is mandatory and rigorous protocols are in place that require businesses to spend additional time and resources on formatting and preparing data.

Data sharing agreements

Companies that share data with one another need to create clear contractual agreements regarding ownership and use of data. Data sharing agreements need to address a broad set of issues including (i) data quality, timeliness and lifecycle, (ii) compliance with the governance rules and (iii) enforcement of usage rights (McKinsey, 2013).

Factors that influence this cost:

- The costs of contractual data agreements may be higher if data is sensitive, due to the additional legislation associated with that type of data, and the higher risk for the sharing organisation.
- If data is shared between data partners, the number of partners may affect the coordination costs, as more players will have demands on the data (standards, usage, rights etc).

A6.2.4   Technical and organisational considerations

This part of the data sharing framework is concerned with the technical considerations of data sharing, such as preparing data for sharing, and the broader organisational considerations such as the security of the interface that will be used to share data.

Preparing data and ensuring data in appropriate forms are shared can help assuage potential **principal-agent problems.** Sufficiently prepared and secured datasets can help to reduce the risk of data being used for 'unspecified contingencies' as part of 'incomplete contracts', where data sharers and data users fail to specify exactly how data can be used (London Economics, 2019). Reducing these contingencies (and associated potential misuse) can also reduce potential **negative externalities**.

Software costs

Organisations that work with data will invariably need to allocate a budget to software, whether it be developed in house for specific data applications, or through software licenses. Software costs will accrue to data in almost every stage of its lifecycle, from the planning stage to the final stage where data is archived or destroyed. These costs include license fees for software for data collection, cleaning and transforming for sharing purposes as well as the cost of developing a data sharing interface.

Factors that influence this cost:

- Software costs will depend on the volume and complexity of the data being shared. Organisations with complex data streams may need to pay for specialist database management software and programmers to implement the software.

Security costs

Opening up data for sharing exposes the organisation to potential cyber threats (OECD, 2019). It is important for a security expert to be involved in the data sharing process from the planning stage to ensure that data sharing infrastructure does not result in data breaches. Security costs will naturally be higher for organisations that collect and use information relating to individuals. In order

to minimise the financial downside of a potential breach, organisations may buy cyber security insurance (OECD, 2020).

Factors that influence this cost:

■   Firms with a large volume of personal data should spend more on cyber security, especially when data is sensitive.

Storage costs

Data that is intended to be shared must be stored securely after it has been collected or procured. This could entail investing in hardware or cloud-based storage services, or a hybrid model. While the costs of data storage have decreased significantly in recent years, organisations that have terabytes or petabytes of data may still spend a sizable amount (Bald, 2020).

Factors that influence this cost:

■   The volume of data being shared will impact the size of storage costs

Data preparation costs

Data sharing will only generate value in the economy if it is of high quality, is well-structured and is accompanied by documentation, or metadata. Data may need to be transformed, cleaned, anonymised, aggregated or linked with other data before it can be shared with data partners or with the market (IMDA, 2019). Preparing data and metadata for sharing can be time-consuming and costly (Figueiredo, 2017).

In some industries data standards exist that prescribe how data and metadata should be prepared (Figueiredo, 2017). These standards may lower or increase the cost of preparing data depending on how rigorous they are and the quality of the raw data.

When data is shared with specific partners, such as the firms' suppliers or partners, data will also need to be transformed to be compatible with the applications used by those data consumers. For example, if a firm wishes to share data with their suppliers, and the suppliers use an application to manage their warehouses, then the data will have to be formatted accordingly.

Factors that influence this cost:

■   Industry-wide data sharing standards may increase/decrease the costs of preparing data.

■   Where no data sharing standards exist and data is shared openly, firms may need to spend less time and resources on formatting data. On the other hand, if firms are keen for their data to generate the maximum amount of value, they may need to create their own data and metadata standards which could be costly and time-consuming.

A6.2.5   Operationalising data sharing

Once the legal, technical and organisational factors associated with data sharing have been decided, the data can be shared. In this part of the framework, data sharing is operationalised – this stage includes ensuring transparency and accountability, continually monitoring the quality of the data being shared and maintaining the interface used to share.

A focus on continued transparency and accountability of this data use can allow for a reduction of potential **principal-agent problems** – ensuring incentives remain aligned between data sharers and third parties using this data. This transparency and accountability also assuage the impact of **information asymmetries** between these data users and consumers which may be affected by this data use.

Data quality monitoring

In order for shared data to be valuable to external economic actors it needs to be of high quality (Figueiredo, 2017). As businesses evolve and expand the data that is collected and shared may change. Firms may have to spend time and resources to ensure that data and metadata remains findable, accessible, interoperable and reusable (Figueiredo, 2017).[40]

Factors that influence this cost:

- The cost of monitoring data and metadata quality will depend on the volume of data being shared and how often it is uploaded.

Maintenance of data sharing interface

Organisations that share data will also need to spend time and resources on maintaining the sharing interface. This cost includes solving software bugs when they crop up, answering questions from data consumers and updating data structures as data collection methods and technologies evolve over time.

Factors that influence this cost:

- When data is shared openly, there may be a larger number of queries that firms must deal with corresponding to the economic actors that access the data.
- This cost will only apply when data sharing is not once-off, but continuous.
- In industries where data and metadata standards are established the cost of responding to data queries may be lower as data is shared in a widely accepted format and structure.

A6.2.6   Implicit costs of data sharing

Implicit costs refer to external costs which may be incurred on an organisation sharing data. For example, reputational costs, loss of competitive advantage and potential cyber breaches. These costs are not directly incurred on a firm but may be included in an organisation's decision to share data or not.

Negative externalities inflicted on consumers and external parties by an organisation's data sharing activities would result in a potentially higher reputational cost whereas positive externalities generated from data sharing may result in a loss of competitive advantage – which may lead to these organisations choosing to share less data.

---

[40] More information on the FAIR principles can be found here: Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data (2016) 3:160018. doi:10.1038/sdata.2016.18

Competitive disadvantage

An organisations' competitive advantage is an attribute or asset that the organisation possesses which allows it to outperform competitors. For some types of organisations, the data that they collect may give them an advantage over their competitors. Sharing this data may benefit other organisations, or even society as a whole, but could lessen that organisation's competitive advantage (London Economics, 2019).

When proprietary access to data provides organisations with a competitive advantage, then sharing it would impose a cost to that organisation. However, it may be the case that the perceived competitive advantage afforded by data is higher than the actual competitive advantage to the firms. This perception creates a barrier to data sharing.

Factors that influence this cost:

■   Firms with a lack of technical knowledge about the economics of data may be more likely to perceive data sharing as a competitive disadvantage.

Reputational cost

Organisations that share open data may have little control over who can use that data, but still bear the burden of liability and reputational impacts if data is misused, or if individuals can be identified (particularly important with sensitive data such as medical and financial).

Even in cases where data is not misused, organisations and individuals can be hostile towards data collection and sharing (London Economics, 2019). Just over half of consumers are happy to share data, but only if they can see a clear benefit for doing so, such as a free product or service or for personalised products and services (Acxiom and Foresight Factory, 2019; Data and Marketing Association, Acxiom and Foresight Factory, 2018). Firms may experience a loss of engagement from the remaining consumers that do not want to share data (Acxiom and Foresight Factory, 2018).

Factors that influence this cost:

■   The reputational risk is higher when the data in question is sensitive, such as with financial or health data.

Cost of cyber security breaches

When organisations develop interfaces to share data, they increase the risk that further parts of their systems will be exposed. (OECD, 2019). This is particularly true when data is shared more openly, or to a larger group of data consumers. Threats to cyber security can lead to data breaches where information that was not intended to be shared is leaked to malicious actors, as well as to disruptions to business activities and ransom fees. These attacks can severely implicate an organisations' assets, reputation and physical activities. In addition, where firms share data with suppliers, these incidents may have a negative impact on all along the supply chain (OECD, 2019).

The cost of a cyber security breach can be financial, psychological and reputational (University of Portsmouth, 2020). On average, the total cost of a data breach was estimated to be $4.24 million[41] in 2021, having increased by nearly 10% since 2020 (IBM Security, 2021). The largest share of security breach costs results from lost business, which includes increased customer turnover, lost revenue due to system downtime and increasing cost of acquiring new business due to reputational damages (IBM Security, 2021).

Factors that influence this cost:

- The total cost of a data breach depends on the sector in which the firm operates, with healthcare organisations experiencing the highest cost (IBM Security, 2021).
- Data breaches where personally identifiable information was leaked are associated with higher total costs (IBM Security, 2021).

Data breach costs are higher in organisations with less developed security infrastructure

---

[41] Approximately £3.13 million.