





# CONTENTS

Acknowledgements	4
ODI FOREWORD	5
Executive Summary	6
1 Background & Methodology	8
1.1 Context	8
1.2 Approach	10
2 Economic Framework	19
2.1 Modelling trust in a data ecosystem	19
2.2 Logic model	23
3 Framework Calibration and Quantitative Evidence	30
3.1 Evidence used	30
3.2 Framework calibration	30
3.3 Aggregate impact of trust on data sharing	35
3.4 Impact of changes in trust in data ecosystems	42
3.5 Economic impact of data sharing	51
3.6 Evidence gaps and direction of future research	53
4 Conclusion	56
5 Bibliography	58
Annex A Additional Detail on Evidence Review and Analytic Strategy	64

## ACKNOWLEDGEMENTS

Frontier Economics were assisted throughout by Dr Kieron O'Hara. Staff at the ODI met and corresponded with Frontier Economics regularly in order to develop and steer this research project. However, the robustness of the findings are the responsibility of Frontier Economics alone. We are grateful to Liz Brandt from Ctrl-Shift who agreed to be interviewed as part of this project.

## ODI FOREWORD

Data flows can create value for our societies in a myriad of ways: when data is made open or shared, it enables better decisions, more effective policies, more innovative and useful services, and generally can make our lives better, while fueling economic growth and productivity. Some of this value creation can be measured in monetary terms, while in other ways it is almost intangible. Researching and documenting this rich variety of value creation has been at the heart of the work of the Open Data Institute (ODI), from our 2015 report *Open data means business*,<sup>1</sup> to our recent collaboration with the Bennett Institute for Public Policy.<sup>2</sup>

We also know from countless examples that data can cause harms, and that trust and trustworthiness is a key to enabling this creation of value. Fear of harms caused by the use and misuse of data can lead people to opt out of data collection or avoid using services; organisations may avoid collecting, using or sharing data to avoid risks.

One part of the equation has until now been elusive: can we quantify the importance of trust on the potential value of data? Can we get a better sense of how much trust – or lack thereof – impacts the value created by data flows?

This is a hard question to answer: trust is complex, and so is its impact on the value of data. But it is a question worth exploring: understanding the economic value of trust in data ecosystems would help policymakers and companies justify investment in activities that assess, build, and demonstrate trust and trustworthiness.

We are therefore delighted to introduce this work by economics consultancy Frontier Economics, commissioned by the ODI as part of the InnovateUK-funded 'Data innovation for the UK: research and development' programme,<sup>3</sup> and which took on the challenge to explore this topic with us, rigorously reviewing existing evidence and proposing an economic model of the impact of trust on data ecosystems.

*- Jeni Tennison, Vice President and Chief Strategy Advisor, ODI.*

---

<sup>1</sup> <https://theodi.org/article/open-data-means-business/>

<sup>2</sup> <https://theodi.org/article/the-value-of-data/>

<sup>3</sup> <https://theodi.org/project/data-innovation-for-uk-research-and-development/>

## EXECUTIVE SUMMARY

Data is abundant in the modern world. Many emerging technologies and ecosystems rely on high-quality information. If used effectively, data can drive productivity and inclusion. However, data is not always held by the party that can realise these benefits and data sharing does not always happen. Trust is important in this context. It needs to be maintained between parties who collect and use data. Mistrust reduces the positive economic and social value that can be generated.

The Open Data Institute (ODI) commissioned Frontier Economics to conduct an evaluation of the *economic impact of trust in data ecosystems*. In particular, we examined the impact of trust on data sharing, collection and use within a data ecosystem and the economic value that can be attributed to trust. Our work addressed an important gap in the current evidence base and complimented ongoing research exploring mechanisms to improve trust around data.

We designed an economic framework to describe the effect of increased trust on data sharing, collection and use and the subsequent economic and social benefits that can be realised. We have calibrated this framework based on an extensive review of academic evidence which quantifies the impact of trust on data sharing. Finally, we used key estimates from the literature on the economic value of data to assess, at a high level, the economic impact of trust in data ecosystems.

Our work examined both the impact of trust on organisational and individual data sharing. However, our research revealed that the volume of available quantifiable evidence tends to be more skewed towards the latter. Nonetheless, our results tentatively suggest there is currently no strong reason to expect any difference in the impact of organisational versus individual trust on data sharing. Our calibration of existing survey evidence revealed that, on average, a 1 point increase on a 5 point trust scale leads to a 0.27 point increase on a 5 point data-sharing scale.

This aggregate result suggests that even large increases in trust will only correspond to moderate impacts on willingness to share data. This emphasises that there are many factors, alongside trust, which affect data sharing. Increasing trust without addressing other factors is unlikely to be sufficient. To reach optimal levels of data sharing it is likely that data-sharing infrastructure would need to be improved and access mechanisms would have to be addressed.

Linking these estimates to the academic and policy literature of the social and economic impact of data access and sharing allows us to estimate the economic impact of trust in data ecosystems. This high-level estimation suggests that a 25% increase in data sharing could generate an additional 47.3 to 118.3 billion US\$ in the world's 20 largest economies. Clearly, this will affect certain ecosystems more than others as the importance of trust in determining data sharing varies. It is worth noting that most of the studies which examine the economic value of data sharing focus on organisations sharing and re-sharing data (rather than individual data sharing). This does lead to some additional uncertainty as the majority of our quantified evidence on the impact of trust on data sharing is at the individual level.

We explored in which circumstances trust is likely to have a larger or smaller effect on data flows by conducting a set of semi-structured interviews with actors active in the healthcare and financial services data ecosystems. Triangulation between

academic evidence and interview insights revealed that the impact of trust tends to be larger where initial levels of trust are low. In these cases, timely intervention to build trust is vital. Norms and unwritten attitudes play a key role in determining baseline levels of trust. These are likely to vary based on both individuals' and organisations' cultural background. In addition, contextual factors are crucial to understanding the role of trust in influencing data sharing. For example, sharing more sensitive forms of data would require higher levels of trust.

Finally, we examined shocks to trust in real-life settings ('natural experiments') by selecting instances where trust was eroded or regulations increased trust. This enabled us to test the dynamic impact of trust on data-sharing behaviour and get closer to identifying a causal relationship between trust and data sharing.

In all cases analysed, a breach of trust caused a decrease in the number of individuals willing to share data with the institutions involved in the breach. We explored what might be driving the persistence of impacts as part of our qualitative engagements. Stakeholders suggested that a loss of trust is likely to have longer-lasting consequences if the affected organisation is driven by commercial motives and if the scale of the problem is concealed at first. Negative impacts tend to be more transitory if the affected institution made an error rather than intentionally deceiving for commercial reasons, and can visibly and promptly rectify the issue. Conversely, policy interventions aimed at increasing trust (e.g. GDPR) can have positive impacts on the level of data sharing for both individuals and organisations, but these benefits tend to increase over time and can involve an initial cost for organisations which have to comply to prove their trustworthiness.

On the whole, this research suggests there is robust quantified evidence that greater trust is associated with increased data sharing. This confirms existing anecdotal evidence and justifies ongoing efforts to design mechanisms to boost trust. In some cases where there is scope to achieve significant improvements in trust, the relevant effect size will be large.

However, the average magnitude of the relationship between trust and data sharing suggests that boosting trust will have to be accompanied by a suite of other interventions (enabling greater discoverability of data, for example) in order for data sharing to reach its optimal level. Exploring wider determinants of data sharing and how they can complement increases in trust will be an important area for future research. Context is also extremely important in determining the impact of changes in trust. The specific trust linkages and their maturity matter.

Whilst there is robust evidence on the importance of trust in data ecosystems, our research identified key gaps in the existing evidence, which make it challenging to link our core findings on the impact of trust on data sharing to the wider literature on the economic impact of data sharing, collection and use. In particular, a relatively small set of papers in our sample examine the actual impact of trust on data sharing using a quantified approach, mostly focusing on individuals sharing data about themselves. Less quantified evidence is available on organisational trust linkages and data sharing. Conversely, the majority of the evidence which assesses the impact of data sharing on economic outcomes is focused on organisational data sharing. Adding to these existing evidence bases will allow future work to determine how the economic value of trust varies across different activities as well as different relationships.

# 1 BACKGROUND & METHODOLOGY

## 1.1 Context

### 1.1.1 Trust in data ecosystems

#### The role of trust

A successful end state for a data ecosystem is a ‘farmland for data’, where data is used in a way that creates positive impact: driving productivity, boosting research and innovation, and increasing inclusion and welfare.<sup>4</sup>

Achieving this desirable end state involves navigating a path between a ‘wasteland for data’ where data sharing and use is limited due to mistrust and there are substantial fears regarding ethics and equity in the use of data; and an ‘oil field for data’ where data is hoarded and used for competitive advantage, limiting the potential for innovation and capabilities to grow.

Trust is a critical enabler of a ‘farmland for data’ ecosystem – in order for data to flow optimally within an ecosystem allowing for its full potential to be realised. There are multiple actors within any single data ecosystem. This implies that trust needs to exist between several different ecosystem participants:

- those who contribute information to a data set need to place trust in the organisations they share data with;
- organisations who are responsible for collecting, managing or ensuring access to data need to prove their trustworthiness to carry out that role in a capable and ethical way;
- data users need to trust the validity of the data to derive value from it by creating products, services and insights; and
- the communities in and around the data ecosystem, such as the general public who need to understand the importance of trust in allowing flows of data and therefore extracting value from the data ecosystem.

In turn, mistrust between these different groups hinders the potential for data to flow within the ecosystem and the positive economic and social value of data as a result.

#### Prominence of trust

Evidence suggests that trust is a critical factor enabling an open and trustworthy data ecosystem. A recent survey conducted by ODI and YouGov found that nearly 9 in 10 people (87%) feel it is important or very important that organisations they interact with use data about them ethically.<sup>5</sup>

---

<sup>4</sup> <https://theodi.org/about-the-odi/our-vision-and-manifesto/our-theory-of-change/>

<sup>5</sup> <https://theodi.org/article/odi-survey-reveals-british-consumer-attitudes-to-sharing-personal-data/>  
<https://theodi.org/article/nearly-9-in-10-people-think-its-important-that-organisations-use-personal-data-ethically/>

However, this evidence also suggests that a low proportion of the general population trusts a number of organisational types with ethical data practices/collecting personal data. For example, social media sites are seen as being the least trustworthy and healthcare institutions the most trustworthy when handling data.<sup>6</sup> This suggests that current levels of trust and trustworthiness tend to be very context-dependent. Factors influencing trust in an ecosystem include:

- the maturity of the ecosystem – baseline trust tends to be lower for nascent ecosystems, and higher for established ones; and
- the understanding and awareness of data-sharing practices within a certain ecosystem.<sup>7</sup>

Ethics around data-sharing practices has become a prominent issue in recent years, as organisations increasingly rely on data to improve the way they work and consumers are able to access increasingly targeted and personalised services based on their data.

In parallel, people’s awareness about their data rights and the potential for misuse of personal data and data breaches has increased. This has in part been driven by the rollout of regulations, such as the General Data Protection Regulation (GDPR), the ongoing debate on the regulation of big tech and widespread media coverage of events like the Cambridge Analytica and Facebook incidents.<sup>8</sup>

### 1.1.2 Role of ODI

The mission of the Open Data Institute (ODI) is to work with companies and governments to build an open, trustworthy data ecosystem.

The ODI is currently engaged in research to develop the next generation of public and private services. This involves applying ethical considerations into how data is collected, managed and used.

One strand of the ODI’s research is focused on exploring which mechanisms are likely to have the most impact in improving trust between organisations around data. Some of the ODI work in this area investigates:

- the levers through which trust in data ecosystems can be enhanced;<sup>9</sup>
- the ways in which organisations can demonstrate trustworthiness when sharing data;<sup>10</sup> and
- the value of sharing data to build trust and trustworthiness.<sup>11</sup>

Another key strand of the ODI’s research focuses on the value of sharing data, both in the private and public sector.<sup>12</sup>

<sup>6</sup> <https://theodi.org/article/nearly-9-in-10-people-think-its-important-that-organisations-use-personal-data-ethically/>

<sup>7</sup> [https://ec.europa.eu/commission/commissioners/2019-2024/vestager/announcements/speech-executive-vice-president-margrethe-vestager-building-trust-technology\\_en](https://ec.europa.eu/commission/commissioners/2019-2024/vestager/announcements/speech-executive-vice-president-margrethe-vestager-building-trust-technology_en)

<sup>8</sup> <https://theodi.org/article/nearly-9-in-10-people-think-its-important-that-organisations-use-personal-data-ethically/>

<sup>9</sup> <https://theodi.org/project/building-trust-through-audit-and-certification/>

<sup>10</sup> <https://theodi.org/article/demonstrating-and-assessing-trustworthiness-when-sharing-data/>

<sup>11</sup> <https://theodi.org/article/case-study-the-value-of-sharing-data-to-build-trust/>

<sup>12</sup> <https://theodi.org/project/the-value-of-data-sharing-in-the-private-sector/#1586850452070-39243495-0d94>

### 1.1.3 Research question

Frontier were asked to assess the relationship between trust and the sharing, collection and use of data: ***what is the economic impact of increased trust in a data ecosystem?***

Beyond the work developed by the ODI and their collaborators, there is a substantial literature exploring the mechanisms through which trustworthiness and trust can be established and maintained within data ecosystems. Likewise, a large volume of academic and policy research has been recently devoted to understanding the relationship between data sharing, collection and use and economic impact.

**Figure 1 The relationship of interest**



Source: Frontier Economics

Our work has explored a key evidence gap connecting these two literatures: the ways in which trust and trustworthiness enable the sharing, collection and use of data in a given ecosystem. There is strong theoretical and anecdotal evidence supporting this relationship and disparate academic studies which have quantified this relationship in a number of different ways. In this study, we have rigorously synthesised available evidence in a consistent way to validate key hypotheses of interest.

After validating and testing the relationship between trust and data sharing, collection and use, we have linked these findings to existing work on the economic value of data access and sharing. This has allowed us to conduct a high-level assessment of the economic impact of trust in data ecosystems.

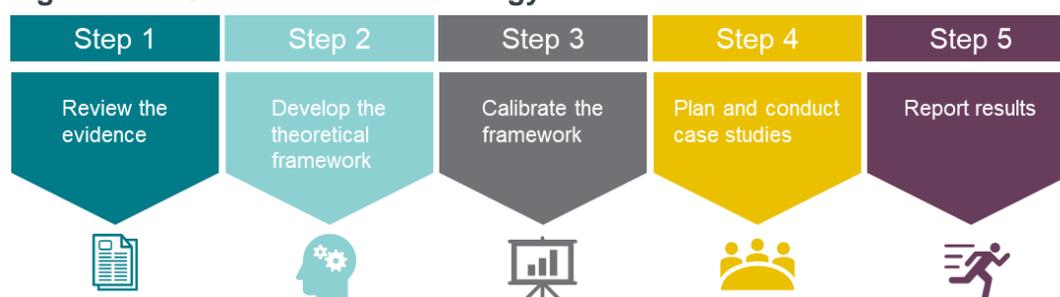
By conducting an in-depth analysis of existing strands of literature and comparing evidence on trust and the economic value of data, this research has made an important first step towards answering the research question originally set out by the ODI.

While our attempts to synthesise, compare and link existing evidence has enabled us to generate valuable preliminary results on the economic impact of trust in data ecosystems, our research has also highlighted some key gaps in the existing evidence base. These gaps make it challenging to draw definitive conclusions on the economic impact of trust in a data ecosystem, but serve as a guide to direct future research efforts.

The following section sets out the approach we adopted to answer the research question.

## 1.2 Approach

We have implemented a 5-step approach to address the ODI's questions. We have illustrated this methodology in Figure 2 below and provided further detail in the following sub-sections.

**Figure 2 Overview of methodology**

### 1.2.1 Stage 1. Literature review

At the core of our approach is a review of the evidence that uses empirical methods to quantify the relationship between trust in data ecosystems, and the amount of data sharing, collection and use that takes place in those ecosystems.

We implemented a rigorous search strategy to identify a wide pool of available studies which explored the relationship between trust and data sharing. We then ranked these studies based on relevance to our research question and robustness of the study's methodology.

In addition to exploring existing evidence on the aggregate effect of trust, we also investigated the contextual factors that make trust more or less important to the functioning of data ecosystems, such as the type of actors and audiences within a certain ecosystem. This has helped to highlight areas for further research.

#### The search strategy

Our comprehensive review of the available literature started with the development of a codified search strategy to identify a broad range of academic studies. This strategy set out the type and fields of literature that we were to review, the databases that we were to search and the search terms that we would apply, as well as any relevant exclusion criteria.

We wanted to include publications from multiple disciplines including economics, policy, computer science, social science and ethics/applied philosophy. We therefore used both discipline-specific databases (RePEc and EconLit for economic literature) and generic academic databases (JSTOR and ScienceDirect for multi-disciplinary literature) to identify existing studies.

Our search terms included multiple combinations of 'trust' (or synonyms) AND 'data' OR 'data sharing' OR 'data collection' OR 'data use' to identify academic studies as well as research papers and policy articles.

We augmented our rigorous review of publicly available academic evidence with additional insights from relevant grey literature. These articles and reports were identified based on domain knowledge within our team and the ODI's suggestions. We have included further detail on our search strategy in Annex A.

## Categories of relevant research identified

There are several ways in which the impact of trust on data sharing is quantified in the identified literature:

- **Empirical survey evidence.** The most common form of quantification is via surveys. In these surveys, participants value the importance of trust and describe their willingness to provide information using Likert scales.<sup>13</sup> Based on the survey data collected, the studies then quantify the impact of trust on data sharing using multivariate regression techniques (in an attempt to isolate the impact of trust on data sharing as separate to the impact of confounding factors).<sup>14</sup> For this type of evidence, observed correlations can be computed and compared in different contexts, but the direction of causality may not always be clear. In particular, while survey results are informative to evaluate the magnitude of the relationship between trust and data sharing, coefficient estimates might suffer from reverse causality: On the one hand, greater levels of trust might lead to more data sharing in a given ecosystem, on the other hand, it is possible that people who are more willing to share data are more inclined to display general attitudes of trust.
- **Natural experiments.** It is possible to exploit major shocks to trust and examine data sharing and use before and after. For example, the advent of GDPR or a major data incident could both impact trust. The resulting evidence is more likely to be causal in nature but may not always be applicable more widely.
- **Theoretical models.** The value of trust is also quantified using conceptual game theory models which include assumptions and parameter values.<sup>15</sup> These models use mathematical modelling to explore strategic interactions between different decision-making entities. For example, models can be developed which attempt to predict the behaviour of suppliers and retailers given different levels of trust in forecasted demand.

In general, the survey evidence relates to stated preferences with regards to data sharing and its drivers. The natural experiments provide an indication of actual behaviour following a shock to trust. Examining actions rather than intentions is usually more reliable. However, in this context there may be examples where additional information is contained in the attitudinal information. For example:

- just because you are a customer of one bank (and therefore share data with it) does not necessarily mean that you trust it more than alternatives. Your choice of bank could have been driven by a range of other factors; and/or
- you might state you do not trust a particular organisation (e.g. a social media network) but continue to use and share data with it as the benefits it offers outweigh the potential risks.

---

<sup>13</sup> Likert scales are psychometric scales which allow respondents to specify their level of agreement or disagreement on a symmetric agree-disagree scale.

<sup>14</sup> These techniques are typically more informative than simple bivariate correlation.

<sup>15</sup> For example, agents are assumed to behave rationally or seek to maximise their utility/profits.

To account for this, our quantification methodology relies on the use of both survey evidence and natural experiments. These two types of evidence are complementary (see further detail below).

### Volume of relevant research identified

In total, we identified 87 potentially relevant studies. Based on a more thorough review, we classified these papers into three categories: high, medium and low relevance. In Figure 3 below we explain in detail the criteria we used for this classification approach and set out the number of papers that fell in each category. In general, more robust, relevant and recent studies were classified as higher relevance.

**Figure 3 Evidence relevance ranking**

Relevance ranking	Number of studies	Criteria for ranking
High	21	<ul style="list-style-type: none"> <li>▪ the research question is highly relevant to inform the relationship between trust, data sharing/collection/use and subsequent economic impacts.</li> <li>▪ the methodology includes an assessment of the impact of trust on data sharing/collection/use, or the economic impact of data sharing/collection/use.</li> <li>▪ The assessment is quantitative in almost all cases. Where surveys are conducted, the sample size is large enough to draw meaningful statistical conclusions. If qualitative, the assessment is highly rigorous and is able to isolate the impact of changes in trust on data sharing.</li> <li>▪ the type of trust linkages and types of actors under analysis are easily identifiable.</li> <li>▪ all studies were conducted post 2010.</li> </ul>
Medium	16	<ul style="list-style-type: none"> <li>▪ the research question is highly relevant to inform the relationship between trust, data sharing/collection/use and subsequent economic impacts.</li> <li>▪ the methodology includes an assessment of the impact of trust on data sharing/collection/use, or the economic impact of data sharing/collection/use. The assessment is either qualitative or quantitative.</li> <li>▪ the study was conducted post 2010.</li> </ul>
Low	50	<ul style="list-style-type: none"> <li>▪ the research question is relevant to inform the relationship between trust, data sharing/collection/use and subsequent economic impacts.</li> <li>▪ however, the study does not always include an assessment of the impact of trust on data sharing/collection/use, or the economic impact of data sharing/collection/use.</li> <li>▪ some studies in this group were conducted pre-2010.</li> </ul>

Source: *Frontier Economics*

For the identified 21 'highly relevant' studies, we carried out a full, in-depth content review, considering the following dimensions:

- research question;
- methodology and evaluation of the study's robustness (internal validity);

- sample size (if applicable);
- estimate of the impact of trust on data sharing (if applicable);
- direction of trust movement; and
- type of trust linkage: individual to organisation, organisation to organisation or a combination of the two.

When conducting the review, we paid particular attention to which actor in the ecosystem each study focused on: whether it investigated the willingness to share data (or the actual data-sharing behaviour) of an individual, of an organisation, or possibly both.

Figure 4 below reports a more detailed split of the type of data-sharing attitude/behaviour analysed for each category of studies identified (high, medium and low relevance).

**Figure 4 Actor data-sharing behaviour analysed**

	Individual	Organisation	Both	Total
High relevance	10	7	4	21
Medium relevance	8	5	3	16
Low relevance	17	11	22	50
<b>Total</b>	35	23	29	87

Source: Frontier Economics

Note: The 'both' category refers to papers discussing data-sharing behaviour for multiple actors in the ecosystem (i.e. individuals and organisations).

This shows that most of the publicly available literature on the relationship between trust and data sharing analyses the behaviour of individuals sharing data about themselves (35 out of 87 studies). Organisation-to-organisation data sharing is also discussed in the literature. However, studies on organisational data sharing tend to be more qualitative in nature and do not include a quantified estimate of the impact of trust on data sharing.

### Assessment of reliability

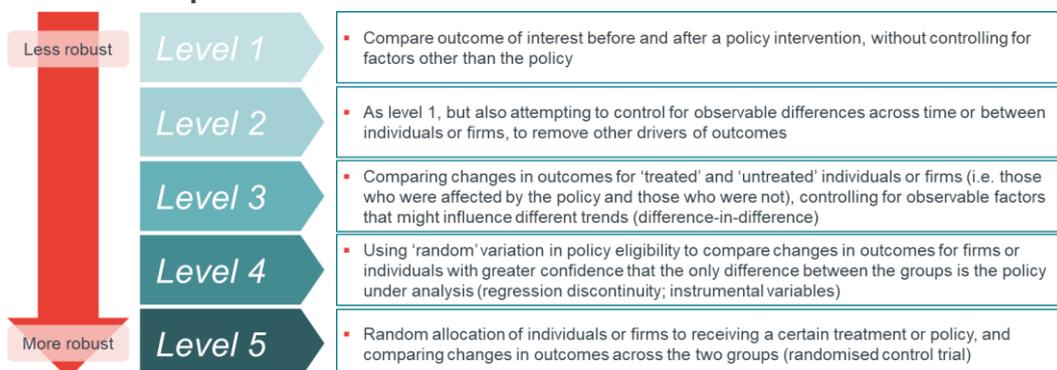
The value of our literature review depends not just on what the evidence says about the impact of trust on data sharing and use, but crucially on the quality of the methods used to draw those conclusions. We have placed higher weight on higher-quality evidence<sup>16</sup> that is relevant to the research question under consideration.

To consider the robustness of each quantitative study that we identified, we drew on an approach in keeping with the Maryland Scientific Methods Scale (SMS).<sup>17</sup> The SMS is widely used within government and includes a five-point (1 to 5) scale. Higher scores indicate that a study has used a relatively more reliable approach when developing a counterfactual. In Figure 5 below we present an example of the type of evidence that would fall under each category.

<sup>16</sup> Quality in this case is primarily based on internal validity, which refers to the extent to which research methods applied in a certain study and subsequent results provide an adequate reflection of the true relationship under analysis and are not due to methodological errors.

<sup>17</sup> The original SMS was developed by Sherman et al. (1998), <https://www.ncjrs.gov/pdffiles/171676.PDF>

**Figure 5 Maryland Scientific Methods Scale to review validity of quantitative evidence**



Source: Frontier Economics, adapted from What Works Centre for Local Growth, <http://www.whatworksgrowth.org/public/files/Scoring-Guide.pdf>

The specific language used in the Maryland SMS is not appropriate for all of the quantitative evidence we uncovered. However, we have been guided by the principles embedded within the scale in order to compare the robustness of different quantitative estimates. These principles and areas we considered include:

- data sources used;
- sample size;
- granularity of impact; and
- extent to which the study has accounted for other differences between the treatment and control groups.

To assess the relevance and robustness of qualitative and theoretical evidence, we considered the extent to which the study is able to isolate the impact of changes in trust on data sharing, either through theoretical modelling or through semi-structured qualitative interviews.

As outlined above, some of the academic literature identified focuses on analysing the impact that major shocks to trust (which we refer to as 'natural experiments') had on data sharing. The literature analyses the impact of both negative shocks to trust, e.g. the Cambridge Analytica incident, and positive shocks to trust driven by regulations, e.g. GDPR.

To test the robustness of the results arising from this evidence, we looked into a wider set of natural experiments (either negative or positive shocks to trust) in different sectors (for example, healthcare and banking), and any related evidence that suggests what the impact of the shock might have been. Further detail is provided in Annex A.

## 1.2.2 Stage 2. Develop a structured theoretical framework

During the second stage of the study we developed a structured theoretical framework which describes the channels through which trust drives data sharing, and the factors that make trust more or less important. This includes two high-level components:

- **Conceptualising trust:**

- definition of trust in the context of data ecosystems;
- setting out the main drivers of trustworthiness and of trust;
- exploring the contextual factors which determine where trust is more or less important in enabling data sharing.
- **Defining a theory of change** which links trust to data sharing and ultimately economic impact, using a logic model framework.

The conceptualisation of trust draws from an existing trust model designed by O'Hara (2012) and applied by the ODI in recently published research.<sup>18</sup> The theory of change has been designed using a logic model framework, a technique widely adopted to identify, describe and arrange the key aspects of an intervention and represent how the intervention produces change.<sup>19</sup>

Developing a robust economic framework to describe the channels through which trust drives data flows and in turn generates economic and social value ensures the evaluation is grounded in economic theory.

### 1.2.3 Stage 3. Calibrating the framework

We then assessed the sign, size and significance of the conceptual relationships articulated in Stage 2 by quantifying the aggregate impact of trust on data sharing. This validation was based entirely on the quantitative evidence that we identified as part of our literature review.

Our ability to quantify the nature and the strength of these conceptual relationships depended on the base of existing evidence. It was not possible to split out effects according to each channel and theory of change we articulated in the conceptual framework. For example, the existing evidence did not always discriminate between specific activities (e.g. data collection vs. data use) or different actors (e.g. data institution vs. data subject). We carried out the framework calibration exercise in three steps, each described below.

#### Step 1: assess the sign and magnitude of the relationship

First, we compared estimates from the academic evidence review to measure the aggregate relationship between trust and data sharing, collection and use. Available academic literature enabled us to assess:

- whether the relationship between trust and data sharing is positive (as would be expected based on our conceptual framework); and
- what is the order of magnitude of the relationship (how small or large the impact of trust on data sharing could be).

---

<sup>18</sup> <https://theodi.org/article/designing-trustworthy-data-institutions-report/>

<sup>19</sup> While our research objective is to describe the impact of trust rather than of a specific intervention, it still provides for a useful framework to distil the key ways in which trust can lead to greater data sharing and wider economic and social value.

### Step 2: test the direction of causality

While most evidence from existing academic studies on trust in data ecosystems has informed the sign and magnitude of the relationship, it cannot provide a causal estimate of trust on data sharing, collection and use. This is because most research to date tests the impact of trust on data sharing through surveys, i.e. measures respondents' self-reported levels of trust, willingness to share data, and then analyses the impact of trust on data sharing performing a regression or correlation analysis on survey data.

As such, surveys are unable to assess the impact of trust on actual data-sharing behaviour. Surveys also tend to record preferences at a specific point in time (i.e. the evidence is cross-sectional) without capturing changes in behaviour over time. This makes it challenging to disentangle cause and effect when comparing these type of studies.

Therefore, we adopted a different approach to test whether changes in trust have an impact on data-sharing behaviour. The main objective of this assessment was to test that the direction of causality is in line with our expectations, i.e. isolating the impact that trust has on data sharing from any reverse effects or feedback loops (e.g. in some ecosystems where there are greater levels of data sharing, actors might be more willing to trust each other with data).

We did so by looking into both academic (where available) and anecdotal evidence on a set of major shocks to trust in real-life settings (i.e. 'natural experiments'), by selecting a number of relevant instances of negative shocks to trust, i.e. where trust was suddenly eroded because of an incident or a data breach (e.g. the Cambridge Analytica and Care.data incidents), or, conversely, positive shocks to trust, i.e. where regulations were introduced to increase trustworthiness and trust (e.g. the GDPR).<sup>20</sup>

### Step 3: link to economic impacts

Finally, we linked the aggregate estimates on the impact of trust on data sharing to the wider literature on the economic impact of sharing and using data. This enabled us to provide an indicative estimate of the economic impact of trust in data ecosystems.

## 1.2.4 Stage 4. Case studies to validate the framework

We carried out five semi-structured interviews, organised in two case studies, to test and contextualise the empirical evidence.

Firstly, we selected case study sectors in conjunction with the ODI. We wanted to explore the role of trust in facilitating greater data sharing within two different contexts and ecosystems. We focused on:

- the role played by trust in facilitating the use of patient data as an input to healthcare research; and

---

<sup>20</sup> Academic evidence was available on some of these natural experiments. For other natural experiments where academic evidence was not available, we have gathered anecdotal evidence and ad-hoc statistics from news articles. We have placed more weight on the academic evidence.

- the impact of trust on data sharing in the context of Open Banking.

The type of information that is exchanged in these two ecosystems is very different as are the specific organisational types involved. This variation allowed us to consider trust from a variety of different perspectives.

We used the interviews to consider a number of specific issues in depth:

- validation of our conceptual trust framework and theory of change;
- exploration of the role of trust in different ecosystems;
- consideration of a possible distinction between creating trust in new ecosystems versus maintaining trust in more established ecosystems; and
- real-world examples of changes in trust and their direct and indirect ramifications on data sharing.

The insights gathered from this qualitative engagement are presented in Sections 2 and 3.

### 1.2.5 Stage 5. Reporting

This report summarises all of the work we have undertaken and the conclusions we have reached. Staff at the ODI met and corresponded with Frontier Economics regularly in order to develop and steer this research project and provided feedback on an earlier draft of this report.

The remainder of the report is structured as follows:

- in Section 2 we set out our conceptual framework;
- in Section 3 we present the results of our framework quantification and qualitative validation; and
- in Section 4 we conclude and summarise key results.

## 2 ECONOMIC FRAMEWORK

In this section we define trust in the context of data ecosystems. We also set out conceptually:

- how trust can impact data sharing, collection and use; and
- the economic and social value added associated with greater data sharing, collection and use.

### 2.1 Modelling trust in a data ecosystem

To determine how trust impacts data sharing, the first building block of our economic framework sets out what we mean by a data ecosystem and an appropriate definition of the role of trust in that context.

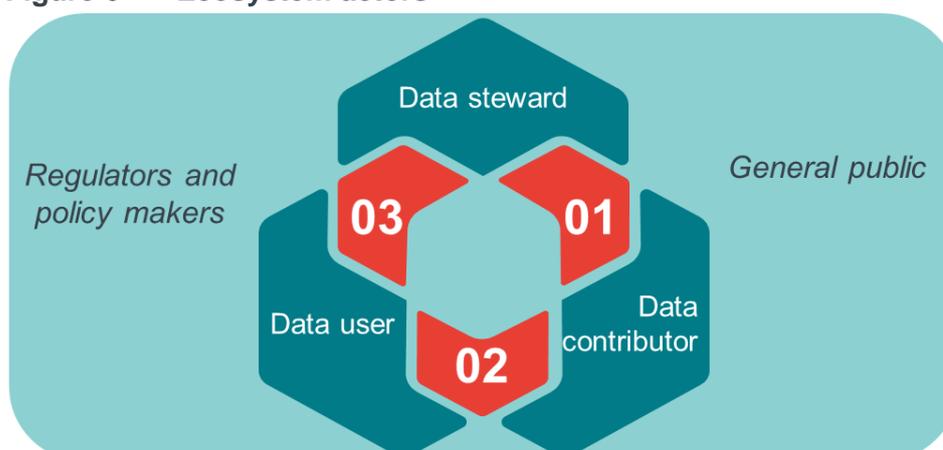
#### 2.1.1 The data ecosystem

To distil the key determinants of trust and trustworthiness in a data ecosystem, we need to consider what we refer to as a data ecosystem in the context of this research, and which actors are going to interact and form relationships within that ecosystem.

As defined by the ODI's Data Ecosystem Mapping methodology,<sup>21</sup> a data ecosystem consists of data infrastructure, and the people, communities and organisations that benefit from the value created by it. Data infrastructure in this context is made up of data assets, standards, technologies, policies and the organisations that steward and contribute to them.

Our work explores the trust linkages between a set of activities within a typical data ecosystem as we illustrate below (Figure 6).<sup>22</sup>

**Figure 6 Ecosystem actors**<sup>23</sup>



Source: Frontier Economics

<sup>21</sup> <http://theodi.org/wp-content/uploads/2019/07/ODI-Data-Ecosystem-Mapping-A1-fold-to-A5-2019-1.pdf>

<sup>22</sup> <https://theodi.org/article/mapping-data-ecosystems/>

<sup>23</sup> Additionally, the data ecosystem includes re-users, end users, capacity developers and technology providers.

- **Contributing to data:** data subjects contribute to the dataset, either knowingly or unknowingly through use of a service. Data contributors' trust towards other actors in the ecosystem ensures they feel comfortable sharing their data;
- **Data stewarding:** certain ecosystem participants collect, manage or ensure access to a dataset. These actors need to prove their trustworthiness to the parties they exchange data with;
- **Data use** involves the creation of things, products, services, analyses, insights, stories or visualisations. These actors want to determine that the data is trustworthy and accurate before engaging.

In addition, benefits from data ecosystems can accrue to actors who are able to make better decisions. They therefore benefit from a high level of trust within an ecosystem as it increases the confidence with which they can act.

## 2.1.2 Defining trust in data ecosystems

### Definition of trust

The academic literature offers various frameworks for describing trust, trustworthiness and what it means to be trusted in different settings. These frameworks break trust down into key components such as 'credibility' 'reliability' and 'competence'.<sup>24</sup>

Several academic studies attempt to pin down what trust means in the context of data ecosystems. Gupta (2015) defines trust as the "willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action".<sup>25</sup>

Therefore, trust can partially determine the degree to which individuals and organisations are willing to enter into agreements with each other. It also influences how actors behave towards other ecosystem participants. As a result, trust is going to impact on how data is exchanged within a given ecosystem.

To distil the key elements of trust in a data ecosystem, this study builds on the previous work by O'Hara (2012)<sup>26</sup> and the ODI (2020)<sup>27</sup> which has made a distinction between **trust** and **trustworthiness**. Being trustworthy is different from being trusted, and the two must be aligned to avoid a breakdown of trust.<sup>28</sup> Placing trust in an actor when it is not trustworthy can lead to a loss of trust following an unexpected event. For example, this can happen through *misrepresentation*, where a data institution provides data contributors with assurances about data

---

<sup>24</sup> Other components identified by academic studies are "intimacy and self-orientation" <https://trustedadvisor.com/why-trust-matters/understanding-trust/understanding-the-trust-equation> and "honesty, competency and reliability" or "rigorous logic, authenticity and empathy" <https://www.thebritishacademy.ac.uk/sites/default/files/Trust-Trustworthiness-Transparency.pdf>

<sup>25</sup> <https://link.springer.com/article/10.1057/dddmp.2015.32>

<sup>26</sup> [https://eprints.soton.ac.uk/341800/1/ohara\\_trust\\_working\\_paper\\_aug\\_2012.pdf](https://eprints.soton.ac.uk/341800/1/ohara_trust_working_paper_aug_2012.pdf)

<sup>27</sup> <https://theodi.org/article/designing-trustworthy-data-institutions-report/>

<sup>28</sup> [http://theodi.org/wp-content/uploads/2020/04/OPEN\\_Designing-trustworthy-data-institutions\\_ODI\\_2020.pdf](http://theodi.org/wp-content/uploads/2020/04/OPEN_Designing-trustworthy-data-institutions_ODI_2020.pdf)

protection, but does not have adequate security measures in place, which comes to light when a data breach occurs.

In the remainder of our work we generally consider trustworthiness as a precondition to build trust. At a high level, this means an actor in the ecosystem needs to prove their trustworthiness in order for another ecosystem participants to share their own data or use the information provided.

Despite this characterisation, we acknowledge the complexity and dynamic nature of the relationship between trust and trustworthiness. While trustworthiness is often a precondition to build trust, it can also be the case that some elements of trust are necessary to build trustworthiness (making the relationship between trust and trustworthiness bi-directional). For example, in some cases data contributors may decide to share their data with a new ecosystem player offering an innovative service, before the player is able to build a reputation of trustworthiness (due to a commercial urgency for example). If the data handling and corresponding service provision is considered satisfactory, then trust can grow over time.

### Components of trustworthiness

An actor's trustworthiness can be measured along multiple dimensions, including its ability, willingness and motivation to deliver on its promises.<sup>29</sup> These dimensions are going to depend on the specific circumstances in which claims, actions and commitments to trustworthiness are made, and also the counterparties with which an actor interacts within the ecosystem.

In addition to this, trustworthiness is going to be influenced both by *internal drivers* (e.g. reputation, presence of a pre-existing trust relationship) and *external drivers* (e.g. laws and regulations).<sup>30</sup> While factors such as regulations and codes of conduct are typically mandated externally, actors still have some agency on them (for example, they can choose to obtain an external certification or subject themselves to audit).

### Drivers of trust toward sharing and using data

The willingness and ability of individuals or organisations to place trust in another organisation varies based on the type of actor and data involved. (E.g. perceived privacy and security concerns will be greater when personal data is involved or one party's motives are commercial rather than altruistic. These factors may magnify the importance of trust.)

As articulated by Gupta (2015), the role of trust in determining data sharing is influenced by:

- **Perceived security.** Security corresponds to concerns about the protection of personal information with three specific goals: ensuring information is not altered during transit and storage; verification of a user's identity and eligibility for data access; and confidentiality requiring that data use be confined to authorised purposes by authorised people.

<sup>29</sup> [http://theodi.org/wp-content/uploads/2020/04/OPEN\\_Designing-trustworthy-data-institutions\\_ODI\\_2020.pdf](http://theodi.org/wp-content/uploads/2020/04/OPEN_Designing-trustworthy-data-institutions_ODI_2020.pdf)

<sup>30</sup> Although these drivers originate externally, there is also an internal element that determines how effective the driver is in increasing trust and trustworthiness (i.e. the way organisations respond to them).

- **Perceived privacy.** Privacy is defined as a process of anonymity preservation and is strongly connected to control over information about the self (i.e. in online environments, users who perceive higher threats to their privacy are less likely to disclose personal data, they feel less able to control it and protect themselves).<sup>31</sup>

Gupta (2015) suggests that both perceived security and privacy are relevant drivers of trust: if actors have control over their information flow and protection over their privacy, this will increase the likelihood they place trust in their counterparties and share more data within the ecosystem.

Other studies<sup>32</sup> point to perceived usefulness as another driving factor for trust in a data ecosystem. If an actor in the ecosystem feels that sharing their data will lead to a positive outcome for themselves (e.g. receiving a high-quality product in exchange for their data) or for society (e.g. medical data used for research), that is, if data sharing is perceived to be useful, this is also likely to increase the probability that the actor will place trust in their counterparty and share their data with them.

### 2.1.3 Where trust is more or less important in a data ecosystem

In addition to the above, the incentives to be trustworthy or to place trust in other actors are going to vary based on the context of a particular data ecosystem, as highlighted in recently published work by the ODI.<sup>33</sup>

For example, trustworthiness often assumes a different meaning for organisations working in different sectors. This is because different sectors have different rules governing ethics and trustworthiness around use of data, and different sectors value data use in different ways. There are also likely to be differences between more mature ecosystems, which are able to leverage pre-existing trust relationships, and newer ecosystems, where trust relationships have yet to be built.

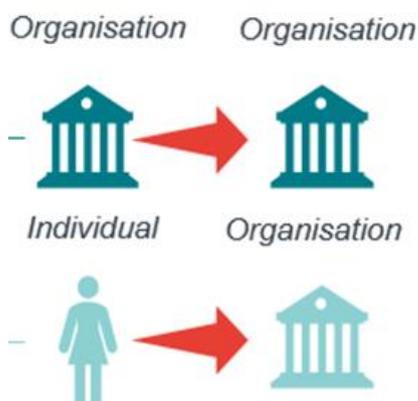
Therefore, in addition to exploring the aggregate effect of trust, we set out the factors that the literature suggests make trust more or less important to the functioning of data ecosystems.

- **Type of actor in the data ecosystem.** Trust might have a different role in enabling the sharing of data between different actors in the ecosystem, such as data institutions (or stewards), data contributors (or subjects), data users (or creators), and other stakeholders. The importance of trust in an ecosystem might also vary *when the data exchange is between an individual and an organisation relative to when it is between two organisations*. Finally, trust might be more difficult to generate if an actor is seen to be motivated by self-interest and commercial outcomes rather than the public good.

<sup>31</sup> <https://link.springer.com/article/10.1057/ddmp.2015.32>

<sup>32</sup> For example, Rissman (2019) <https://www.sciencedirect.com/science/article/abs/pii/S0264837719301000>

<sup>33</sup> <https://theodi.org/article/demonstrating-and-assessing-trustworthiness-when-sharing-data/>

**Figure 7 Data sharing behaviour**

Source: *Frontier Economics*

- **Type of data being exchanged.** The importance of trust in an ecosystem might depend on the type of data being exchanged, and especially whether it is personal or non-personal in nature.
- **Wider context.** The importance of trust might also depend on economic and social factors affecting the ecosystem (e.g. trust might be a more important enabler of data sharing in certain countries and cultural contexts than others, driven by differences in institutions, regulations and norms around sharing). The impact of trust on data sharing might also differ between emergent ecosystems and trust linkages versus previously established relationships.
- **Baseline level of trust.** Throughout this work we have not assumed that the effect of trust on data sharing, collection and use is linear. The effect of a given increase of trust may be larger where initial levels of trust are low than when there is already a significant amount of trust between actors in the ecosystem. There may also be feedback loops that reinforce or balance the data ecosystem after a change in the level of trust.

## 2.2 Logic model

After defining the key components of trust and trustworthiness in data ecosystems, we used a logic model, to articulate the different channels through which:

- trust impacts on data sharing, collection and use; and
- the sharing, collection and use of data impacts on the economy and society.

Logic models are frameworks used to set out the underlying theory for how interventions or actions are expected to generate particular outcomes and impacts. It is helpful to set out these stages explicitly and then examine the extent to which each linkage is validated by the underlying evidence. The focus of this research is not on a specific policy intervention but rather on the role of trust as a driver of data sharing, collection and use. Nonetheless, building a logic model provides for a useful framework to draw out a theory of change leading from trust to economic and social impact, and distil the main channels through which we expect the change to happen.

We present the full logic model in Figure 8. The model describes how activities including laws, regulations norms and standards facilitate an environment with a high degree of trust and trustworthiness.

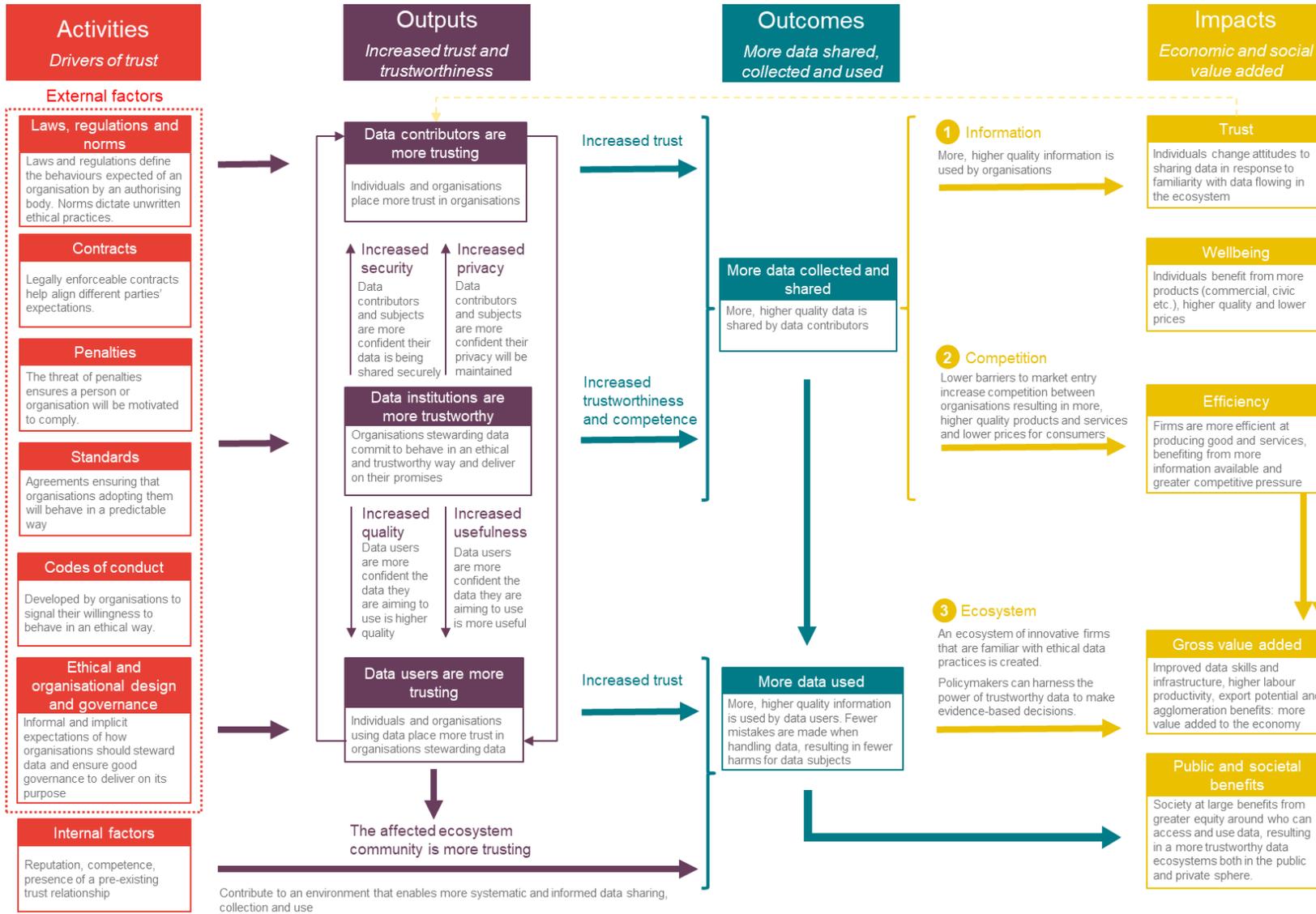
In turn, trust and trustworthiness enable greater data sharing and use. Greater data sharing and use then contribute to positive economic and social impacts through several mechanisms such as greater use of higher-quality information and more vibrant competition which is fostered as a result of lower entry barriers.<sup>34</sup>

Each of these mechanisms results in a range of social and economic impacts including individuals and consumers benefiting from higher-quality services, and firms increasing efficiency. We have also included a feedback loop which represents the possibility of economic impacts further boosting trust. The remainder of this section expands on each column of the logic model.

---

<sup>34</sup> These theories of change build on Frontier Economics research commissioned by the ODI in 2019, on the economic impact of open standards for data.

Figure 8 Impact of trust logic model



Source: Frontier

## 2.2.1 Logic model components and theories of change

### Actions: drivers of trust

System of rules exist to help ensure institutions are trustworthy when handling data, which is a precondition to build trust in a data ecosystem.<sup>35</sup> Some of these rules may be mandated by the environment present in the data ecosystem, others stated by an organisation stewarding data itself, or may be implicit because of social or industry norms.<sup>36</sup> Categories of rules include:

- *Laws and regulations (e.g. GDPR)*: they define the behaviours expected of an organisation by an authorising body;
- *Norms*: implicit or ‘unwritten rules’ of behaviour dictate unwritten ethical practices;
- *Contracts*: legally enforceable contracts help align different parties’ expectations;
- *Penalties*: the threat of penalties ensures a person or organisation will be motivated to comply;
- *Standards*: agreements ensuring that organisations adopting them will behave in a predictable way;
- *Codes of conduct*: developed by organisations to signal their willingness to behave in an ethical way;
- *Ethical design*: ethical principles and values outline how an organisation ideally wishes to behave, and what it is not willing to do;
- *Organisational design*: the process of aligning the structure of an organisation with its objectives, with the aim of improving efficiency and effectiveness. Both ethical and organisational design can be implemented and published to signal intrinsic motivation to be trustworthy.

We have defined these factors as ‘external’ as a single ecosystem player cannot define and implement these rules independently of others. However, it is often the case that individual organisations often have a high degree of autonomy and responsibility in adapting to these rules, and also deciding the degree of engagement they wish to have with external processes. For example this applies to codes of conduct and organisational design.

In addition to these rules, an organisation can boost their trustworthiness by developing a reputation of being competent<sup>37</sup> and having the required skills and expertise in-house. This type of reputation can be sustained through ongoing relationships within the ecosystem.

---

<sup>35</sup> <https://theodi.org/article/designing-trustworthy-data-institutions-report/>

<sup>36</sup> <https://theodi.org/article/designing-trustworthy-data-institutions-report/>

<sup>37</sup> Which will be highly dependent on the skills and expertise accumulated within an organisation over time.

### Outputs: increased trust and trustworthiness

A combination of the above activities in a specific data ecosystem can in turn foster an environment characterised by greater trust and trustworthiness.

At the outputs stage, the logic model then maps the trust relationship between the actors of a data ecosystem. This relationship will vary for each data ecosystem, based on the components of trust and trustworthiness we set out in Section 2.1.1.

In some ecosystems, an intermediary organisation stewards data on behalf of data contributors and shares it with data users. In these instances, activities result in organisations being more trustworthy and committing to behave in an ethical way and deliver on their promises.

As a result of organisations being more trustworthy:

- data contributors and subjects are more confident that their data is being shared securely, and their privacy is being respected. They will therefore place greater trust in an organisation and will be more willing to share data with them.
- data users are more confident that the data they will use is accurate and useful. They will therefore also place greater trust in the organisation and will be more willing to use data provided by them.

In other ecosystems, it might be the case that data flows directly from contributors/subjects to data users, without the presence of an intermediary. Even in these instances, activities aimed at ensuring greater trustworthiness will strengthen the trust linkage between them. This virtuous cycle strengthens trust overall in the community affected and enabled by the data ecosystem.

### Outcomes: more data shared, collected and used

A virtuous cycle of increased trustworthiness and trust enables data to flow more freely within the ecosystem (an end state consistent with a 'farmland for data' scenario). In particular:

- As a result of data institutions being more trustworthy and data contributors being more trusting, more data is collected and shared. As a result of greater volumes of data available, other things equal, more data is used.
- As a result of data institutions being more trustworthy and data users being more trusting, for a given volume of data shared, a greater portion of that data may now be used and its potential unlocked.

Incidents related to the mishandling of data are reduced as a result of increased trustworthiness and competence of data institutions, causing fewer breaches in trust in the data ecosystem.

## Impacts: economic and social value added

The last section in the logic model maps the economic and social value generated by greater data sharing, collection and use. Greater data sharing and use in turn generate economic and social impacts through three main theories of change:<sup>38</sup>

- *Information theory of change*: more, higher quality information is used by organisations.
- *Competition theory of change*: lower barriers to market entry increase competition between organisations resulting in more, higher-quality products and services and lower prices for consumers.
- *Ecosystem theory of change*: an ecosystem of innovative firms that are familiar with ethical data practices is created. Policymakers can harness the power of trustworthy data to make evidence-based decisions.

Each theory of change results in a range of social and economic impacts:

- First, as a result of greater information and competition unlocked by data sharing and use, individuals change attitudes towards sharing data and are more trusting of other actors in the ecosystem, reinforcing a positive trust-data sharing feedback loop;
- As a result of more competition between organisations, individuals benefit from higher-quality products and lower prices in commercial settings, and higher-quality services in commercial and civic settings (e.g. better public services from government);
- As a result of greater competitive pressure, firms are more efficient in producing goods and services and have greater incentive to innovate;
- A new and improved ecosystem of innovative firms generates improved data skills and infrastructure, higher labour productivity and agglomeration benefits, all leading to greater gross value added;
- Greater use and collection of data, in part facilitated by greater trust can also generate benefits associated with evidence-based policymaking. As a result, society at large benefits from a more trustworthy data ecosystem both in the public and private sphere. For example, increased trust in how personal data is handled in the UK could lead to more people willing to download the NHS COVID-19 app and share data about their movement. As a result, more data sharing and use can ultimately lead to a better understanding of how the virus spreads amongst the population.

As is further detailed in the following sections, we have not quantified each separate logical block in this model due to limitations in the underlying evidence. However, the logic model sets the foundations of our economic framework, helping us to contextualise and unpack the research question into its components.

---

<sup>38</sup> These theories of change build on Frontier Economics research commissioned by the ODI in 2019 on the economic impact of open standards for data.

## 2.2.2 Framework limitations

Our framework attempts to set out the key theoretical linkages between trust, data sharing and economic impacts. When investigating the economic impact of trust in data ecosystems, there is value in spelling out the channels through which we expect trust to bring positive economic and social impact.

The conceptual elements of the framework are generic in nature and not intended to capture the nuanced relationship between trust and data sharing within all possible data ecosystems. Context is going to impact the relationship between trust and data sharing differently in each specific data ecosystem.

It is also important to note how, by construction, the logic model embedded in our framework highlights how trust can bring positive economic and social impact, describing a virtuous cycle in which greater trust leads to more data sharing and associated economic benefits, which in turn are going to further strengthen trust in the ecosystem. This is a useful tool to aid conceptual understanding of such a complex topic, but does not necessarily capture what might happen if a vicious cycle were to materialise in an ecosystem, one where a lack of trust negatively impacts data sharing and generates potential economic losses. Nonetheless, the framework still serves as a useful starting point to step through the research question.

## 3 FRAMEWORK CALIBRATION AND QUANTITATIVE EVIDENCE

In this section we build on our theoretical framework which describes conceptually how trust can impact data sharing and wider economic impacts. We calibrated aspects of this framework using the evidence that we identified as part of our literature review. This allowed us to quantify relationships of interest as well as highlight gaps in the existing evidence base.

Below we articulate the key insights arising from the estimates of the impact of trust on data sharing, collection and use.

### 3.1 Evidence used

As outlined in Section 1.2.2, we identified a total of 87 academic and policy papers examining the impact of trust on data sharing, collection and use using three main techniques: empirical survey evidence, natural experiments and theoretical models.

As we outlined in the Approach section, we classified all identified relevant papers into high (21 papers), medium (16 papers) and low (50 papers) relevance. Highly relevant papers directly address our question of interest and include a quantitative assessment of the impact of trust on data sharing, collection and use.

Out of the highly relevant studies identified, 8 included a direct quantification of the impact of trust on data sharing, therefore forming the basis for our estimation.<sup>39</sup>

The medium- and low-relevance studies identified typically do not include a quantified impact estimate of trust. However, in many cases they still provide useful evidence to contextualise and provide greater nuance to our quantified aggregate impacts.

In addition we have analysed a wide set of “natural experiments” looking at dynamic changes to trust levels in real-life settings and their impact on data sharing behaviour. A detailed review of the evidence is provided in the Annex.

### 3.2 Framework calibration

#### 3.2.1 Triangulation of evidence

The core of our work focused on quantifying the impact of trust on data sharing, collection and use. Additionally, we linked our core estimates to existing evidence on the economic impact of data sharing and use. This second linkage has received extensive attention in previous work.

---

<sup>39</sup> The remaining quantified studies typically include a quantification on the levels of trust and data sharing measured via surveys. The reason for excluding them from the calibration exercise is that in those cases it is not possible to directly translate these into a reliable estimate of the impact of trust on data sharing.

In the context of this research, calibrating our economic framework has meant parameterising our framework to understand what proportional change in data use could be achieved by particular changes in trust within the ecosystem.

Therefore, our calibration has relied on existing estimates already available in the literature. Our ability to test the existence of these conceptual relationships depends on the base of existing evidence. As a result, it is not always possible to split out effects according to each relationship articulated in the economic framework, for example, the specific activity in question (e.g. data collection vs. data use) or different actors (e.g. data institution vs. data subject).

As described in the Approach section, based on academic evidence collected and on the review of natural experiments, we have been able to assess the aggregate impact of trust on data sharing, collection and use in two steps:

- We compared estimates from the academic evidence reviewed to measure the aggregate relationship between trust and data sharing, collection and use. The evidence we reviewed applies regression estimation techniques to assess the impact of trust, allowing us to draw statistically meaningful conclusions on the relationship between trust and data sharing.<sup>40</sup> However, as described above, the direction of causality may not always be clear in these studies.<sup>41</sup>
- To test the robustness of the main conclusions we draw from the calibration of academic estimates. We then assess whether changes in trust have an impact on data-sharing behaviour by looking into a set of major shocks to trust in real-life settings (i.e. 'natural experiments').

### 3.2.2 Standardisation of the impact of trust on data sharing

Seven studies amongst those we have reviewed provide a direct quantified estimate of the impact of trust on data sharing, and those studies formed the core of our evidence base.

To assess the aggregate impact of trust on data sharing, we transformed the available estimates to ensure they were directly comparable. Specifically, trust constructs and data-sharing constructs in each study needed to be converted to a common format.<sup>42</sup> That conversion allowed us to measure the estimates of the impact of trust on data sharing on a like-for-like basis and draw out more general conclusions on the aggregate impact of trust on data sharing.

#### Measurement of trust and data sharing

In many studies, trust is assessed via survey responses to questions which include set response options. In other cases, trust is measured through a proxy such as

---

<sup>40</sup> Typically, regression results provide an estimate for the impact of an independent variable (in this case trust) on a dependent variable (in this case data sharing, collection and use). However, a linear probability model may be unable to completely isolate the pure exogenous impact of trust on data sharing, collection and use, even when attempting to account for other trust determinants. More details on how to interpret regression results are presented in the Annex.

<sup>41</sup> This is because most estimates are obtained by applying regressions techniques to survey data, via linear probability models with or without controls and as such reverse causality could be an issue.

<sup>42</sup> For one out of the 8 quantified studies it was not possible to arrive at a standardised impact estimate, therefore that was excluded from the calculation of aggregate impact.

shared norms or common objectives. In the quantitative studies reviewed, willingness to share data or data sharing behaviour is usually measured:

- On a 5-point Likert scale, from 1 ('strongly disagree' with the statement: would you be willing to share data with an organisation in a specific ecosystem) to 5 ('strongly agree' with that statement)
- As a binary variable ('yes' or 'no' answer to the question "have you shared data/do you want to share data with a given intermediary or another organisation in a specific data ecosystem").

As explained above, most of the studies measure self-reported willingness to share data, rather than actual data-sharing behaviour over time.<sup>43</sup> Also, studies on the relationship between trust and data sharing tends to focus more on the data-sharing behaviour of individuals than of organisations.<sup>44</sup> Studies on individual data sharing also tend to be more quantitative in nature. Because of this, the majority of the quantified studies considered for the framework calibration refer to individual data sharing.

### Estimating the aggregate impact of trust on data sharing

In Figure 9 below we list all the quantified studies which formed part of our calibration exercise. In all of the studies reviewed, the sample is large enough to draw statistically meaningful conclusions. In the subset of 7 comparable quantified studies, 5 refer to individual data-sharing behaviour, while 2 focus on organisations sharing data.

---

<sup>43</sup> Most of the studies identified investigated the impact of trust on data sharing. Relatively less evidence is publicly available on the impact of trust on data collection or use.

<sup>44</sup> Of the 87 studies identified in total, 35 focus on individuals sharing data and 23 focus on organisations sharing data.

**Figure 9** Quantified studies

Author	Year	Title	Actor sharing data
Nasser	2020	<i>Impacts of Trust in Government and Privacy Risk Concern on Willingness to Provide Personal Information in Saudi Arabia</i>	Individual
Wiseman	2019	<i>Farmers and their data: An examination of farmers' reluctance to share their data through the lens of the laws impacting smart farming</i>	Organisation
Bauer	2019	<i>Trust and cooperative behaviour: Evidence from the realm of data-sharing</i>	Individual
Bijlsma	2020	<i>Consumer propensity to adopt PSD2 services: trust for sale?</i>	Individual
Gupta	2015	<i>Measuring the impact of security, trust and privacy in information sharing: A study on social networking sites</i>	Individual
Liao	2011	<i>Achieving mass customization through trust-driven information sharing: a supplier's perspective</i>	Organisation
Beldad	2019	<i>Here's my location, for your information: The impact of trust, benefits, and social influence on location sharing application use among Indonesian university students</i>	Individual

Source: *Frontier Economics review of evidence*

Note: While originally considered for the quantification, Rissman (2019) was excluded from the aggregate impact calibration for lack of comparability of coefficient estimates with the other studies.

All estimates in the quantified studies, except for Liao (2011) are obtained from regression coefficients which we have subsequently standardised. Generally, these regression coefficients can be classified in two categories based on how the independent variable (trust), and the dependent variable (data sharing) are measured in the regression:

- categorical variable (trust) on categorical variable (data sharing);
- categorical variable (trust) on binary variable (data sharing).

Figure 10 below provides more detail on the type of evidence used in the calibration as well as an illustrative example on how to interpret coefficient estimates. The 0.6 parameter value used here is purely illustrative and does not refer to a specific finding in the literature.

**Figure 10** Estimates standardisation

Category	Studies	Trust measurement	Data sharing measurement	Interpreting an example coefficient of 0.6
Categorical on categorical	Nasser (2020), Wiseman (2019), Gupta (2015), Liao (2011), Beldad (2015)	5 point Likert scale	5 point Likert scale	a 1 point increase in trust (i.e. a 20% increase in the five point trust scale) leads to a 0.6 point increase in data sharing also measured on a 5 point scale (i.e. a 12% increase in the data sharing scale).
Categorical on binary	Bauer (2019), Bijlsma (2020)	5 point Likert scale	Binary variable (would you share data: yes / no)	a 1 out of 5 point increase in trust (i.e. a 20% increase in the five point trust scale) leads to a 60% increase in data sharing

Source: *Frontier Economics*

Note: *the example coefficient indicates the impact of trust (independent variable) on data sharing (dependent variable) estimated using a linear probability model of the type  $y = a + bx + e$ . Note that  $b = 0.6$  is an illustrative example only and does not refer to any specific estimate used for the calibration.*

To make estimates comparable, we standardise each coefficient to align with the following relationship: a 1 point increase on a 5 point trust scale leads to a X point increase on a 5 point data sharing scale.<sup>45</sup> In this case, X is our coefficient of interest. The full details of the standardisation calculations undertaken for each study, and a list of other quantified evidence which was not directly included in the standardisation, can be found in Annex A.

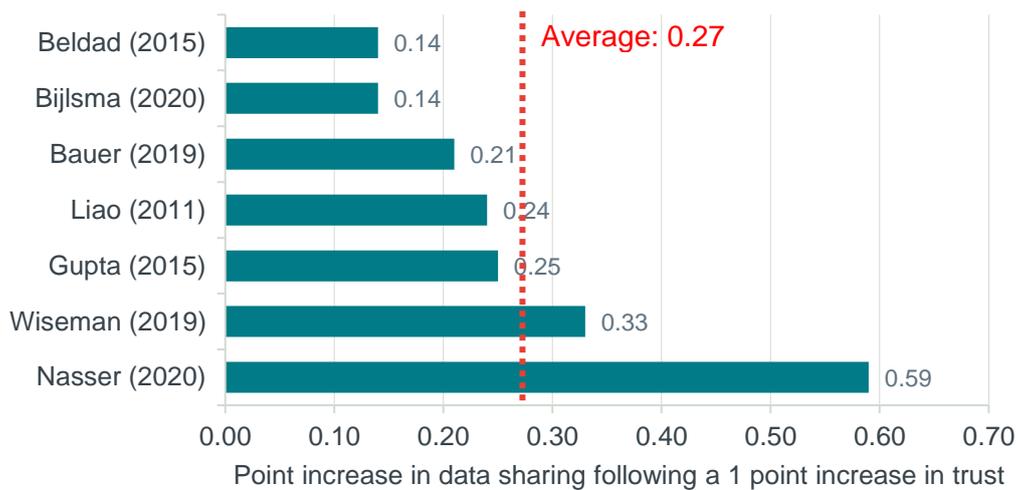
<sup>45</sup> Where Likert scales range from 1 (strongly disagree) to 5 (strongly agree) with a statement of the type: do you trust organisation x, would you share data with organisation x?

### 3.3 Aggregate impact of trust on data sharing

#### Point estimate of the linkage between trust and data sharing

We have calculated aggregate estimates of the impact on trust by computing an average of the quantified estimates available. As we have illustrated in the figure below, there is considerable variation in terms of trust’s impact on data sharing across different studies.

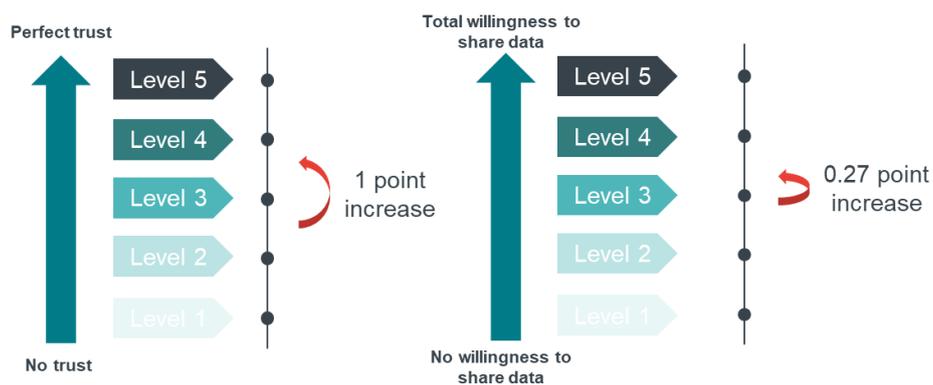
**Figure 11 Impact of a 1 point increase in trust on data sharing**



Source: Frontier Economics

The quantitative studies we have reviewed suggest that, on average, **a 1 point increase on a 5 point trust scale leads to a 0.27 point increase on a 5 point data sharing scale.**

**Figure 12 Average impact of trust on data sharing**



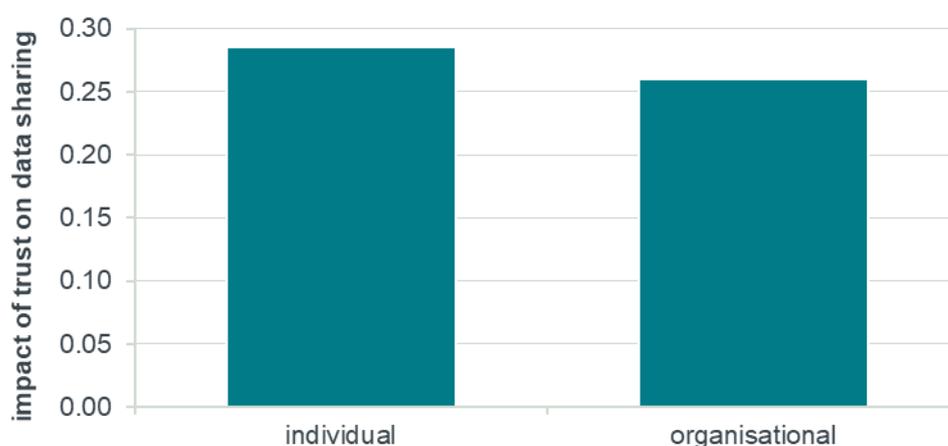
Source: Frontier Economics

The 0.27 point increase in the data sharing scale as a result of a 1 point increase in the trust scale does not refer to a specific level in either of the Likert scales. It indicates the effect of an increase in trust regardless of the baseline levels of trust and data sharing.

This simple extrapolation is illustrative as we do not have a strong reason to believe that the relationship between trust and data sharing will always be linear in nature. In fact, as we discuss below, the role of trust in determining data sharing within ecosystems is likely to exhibit non-linearities. It is nonetheless helpful to consider the possible impacts on data sharing associated with trust increases of different magnitudes.

Whilst most of the studies used for this average estimation focused on individual data sharing,<sup>46</sup> our results reveal there is no reason to believe there is any difference in the impact of organisational versus individual trust on data sharing. As shown in Figure 13 below, the difference in impact based on the reviewed evidence is less than 0.02 points.

**Figure 13** Impact of trust on individual versus organisational data sharing



Source: *Frontier Economics based on review of evidence*

As this result is based on the 7 quantified studies available in the literature, it is not possible to be definitive on the absence of any differences in individual and organisational data sharing. Further research to quantify this relationship is needed to test the robustness of this result.

### Illustration of the overall relationship between trust and data sharing

Based on the average estimate it is possible to extrapolate the effect of increases in trust of different magnitudes. In Figure 14 below we can see that larger jumps in trust correspond to bigger impacts on data sharing. For example, a **3 point** increase on a 5 point trust scale leads to a 0.81 point increase on a 5 point data-sharing scale. The equivalent impact for a **4 point** increase on a 5 point trust scale is a 1.08 increase on a 5 point data-sharing scale.

<sup>46</sup> 5 of the studies in the calibration focused on individuals sharing data, while 2 focused on organisations sharing data.

**Figure 14** Illustrative impacts for different increases in trust

### Interpretation of the results

As we noted in Section 1, currently a low proportion of the general population trusts a number of organisational types with ethical data practices/collecting personal data.<sup>47</sup> This suggests that there is scope to achieve significant improvements in trust in several settings. As a result data sharing could be significantly boosted by over 20% (in excess of one point on a five point willingness to share data scale) in some cases if average trust is increased from low levels to high levels.

In addition, the evidence we rely on to quantify this linkage of interest measures willingness to share data on standardised scales rather than actual data sharing. It may be that once a certain average level of willingness to share data is reached, actual data sharing rises by a greater amount. In other words, a one-point or 20% increase in willingness to share data (measured on a five-point scale) could lead to a far larger than 20% increase in actual data sharing in certain circumstances. This could occur if, for example, actors within a specific data ecosystem move from an average position of being neutral with regards to data sharing to being willing to share data. Crossing these key threshold points may only require a relatively small change in attitudes but still have a large impact on data sharing, collection and use.

These aggregate results show that even large increases in trust will correspond to moderate impacts on willingness to share data overall. This serves to emphasise that trust alone is unlikely to be sufficient to encourage greater sharing. Other factors such as the existence of suitable data-sharing mechanisms and the ability of ecosystem participants to locate and coordinate with each other will also play a key role. Exploring the determinants of data sharing in this context and how they can complement increases in trust will be an important area for future research.

<sup>47</sup> <https://theodi.org/article/nearly-9-in-10-people-think-its-important-that-organisations-use-personal-data-ethically/>

### 3.3.1 Potential drivers of observed variation

The impact estimates we have calculated and reported above are obtained by computing an average of the quantified estimates included in the calibration. The context of a specific trust relationship within a specific ecosystem is important in determining the specific impact of trust within a given context. When comparing the impact estimates included in the calibration, we have investigated whether any of the identified contextual factors that make trust more or less important to the functioning of data ecosystems or differences in the approach adopted by the researchers might be driving some of the observed variation in estimates.

Given that we are examining a relatively small number of studies, we cannot definitely determine what contextual factors might be driving all of the observed variation in quantitative estimates. Nonetheless, we can highlight some stylised facts which hold across the evidence base which can give some insights into the causes of heterogeneity in estimates.

#### Methodological differences

As we would expect, the methodological approach adopted by different researchers can affect the estimated magnitude of the impact of trust on data sharing. In particular, there is some tentative evidence that less internally robust studies (based on our assigned internal-validity scores) tend to overestimate the impact of trust on data sharing.

Lower internal-validity scores are due to small samples sizes or less rigorous methods and imply that the observed results are more likely to have been influenced by chance/randomness.

- For example, the largest impact estimate in the sample (Nasser 2020) might differ from other estimates due to the fact that the estimation was conducted on a small sample, using a simple linear regression model without attempting to control for confounding factors.
- On the other hand, estimates derived from the Bijlsma (2020) study are drawn from a more robust experimental survey design (i.e. participants are randomly allocated to different vignette scenarios to capture perceptions of trust in a specific scenario, and estimation is done through a range of regression techniques on a fairly large sample of respondents).

We can therefore infer that more internally valid studies approximate the relationship between trust and data sharing more closely.

#### Baseline levels of trust

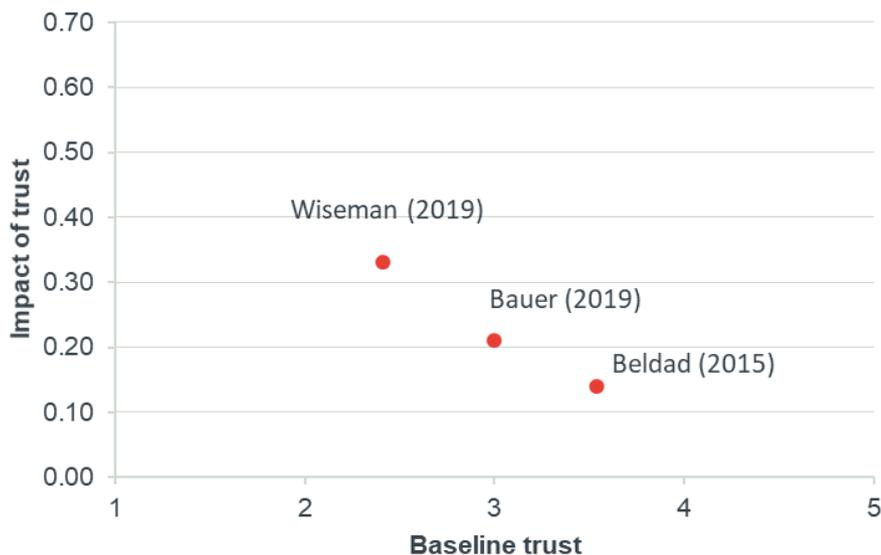
Once methodological differences are taken into account, there are a few potential factors which might explain where the impact of trust is more or less important. We have identified some tentative evidence that the impact of trust tends to be larger where initial levels of trust are low than when there is already a significant amount of trust between actors in the ecosystem. We illustrate this in Figure 15 below. In particular:

- the Wiseman (2019) study shows a relatively higher (0.33) point increase in data sharing as a result of a 1 unit increase in trust, starting from a relatively lower level of baseline trust (2.4/5).
- On the contrary, Beldad (2015) displays a relatively lower (0.14) increase in data sharing as result of a 1 unit increase in trust, starting from a relatively higher baseline level of trust (3.54).

This is in line with our assumption that the impact of trust on data sharing is not necessarily linear in nature and the impact of an increase in trust tends to be higher for lower starting levels of trust.

It is not possible to draw definitive conclusions based only on the relatively few studies which specify both a baseline level of trust and the relationship between trust and data sharing.

**Figure 15 The impact of trust on data sharing for different levels of baseline trust**



Source: Frontier Economics

Note: Only three studies out of the quantified evidence in the calibration included an indication of the baseline levels of trust for the ecosystem under analysis.

The different levels of baseline trust observed in these studies could in part reflect previous interventions deployed in different circumstances. They could also be driven by contextual factors which we discuss in greater detail below.

Our qualitative engagement confirmed that baseline levels of trust matter and that interventions to boost trust and encourage further data sharing need to be tailored to the specific ecosystem. For example, we were told by organisations active in the use of healthcare data for research that acting quickly is crucial when baseline trust levels are low. In these circumstances providing informative answers to questions raised by other actors or rectifying issues quickly becomes essential. Allowing uncertainty to persist can quickly lead to a significant loss of willingness to share data. Where trust is high, we were told that there may be less urgency as other stakeholders will be more likely to give data institutions the benefit of the doubt and need a relatively compelling reason to stop sharing data.

Other stakeholders noted that in newer ecosystems where baseline levels of trust are relatively low the importance of formal regulations and rules are enhanced. Stakeholders from both the health and financial services ecosystems emphasised that trust is also heavily dependent on informal linkages and establishing of relationships between institutions and individuals over time. This more organic development of trust is closely linked to commonalities in terms of culture and experience of repeated interactions. These drivers can complement the more formal rules and regulations. However, this type of informal trust development is usually less feasible initially in newer ecosystems where baseline levels of trust are lower (or between newer players in an established ecosystem).

Similarly, some interviewees did note that this type of informal trust development can in some circumstances make it difficult for a new entrant to successfully join an established ecosystem and access/share data even if they adhere to all the formal rules. This type of informal development of trust over time can also mean that attributing an increase in trust to a specific intervention or event can be challenging.

Stakeholders also made the point that ecosystems where trust has yet to develop are very different to cases where trust used to exist but has subsequently been lost. In these scenarios, stakeholders agreed that rebuilding trust is much more difficult and can require more radical interventions such as the establishment of new institutions and fundamental alterations to the way in which data is shared. We discuss examples of breaches of trust in greater detail below.

Other stakeholders noted that in order to encourage greater data sharing in new ecosystems development of trust has to be accompanied by a clear articulation of the benefits of sharing data to all parties. This will help to ensure that stakeholders will overcome the initial barriers associated with engaging with new counterparties.

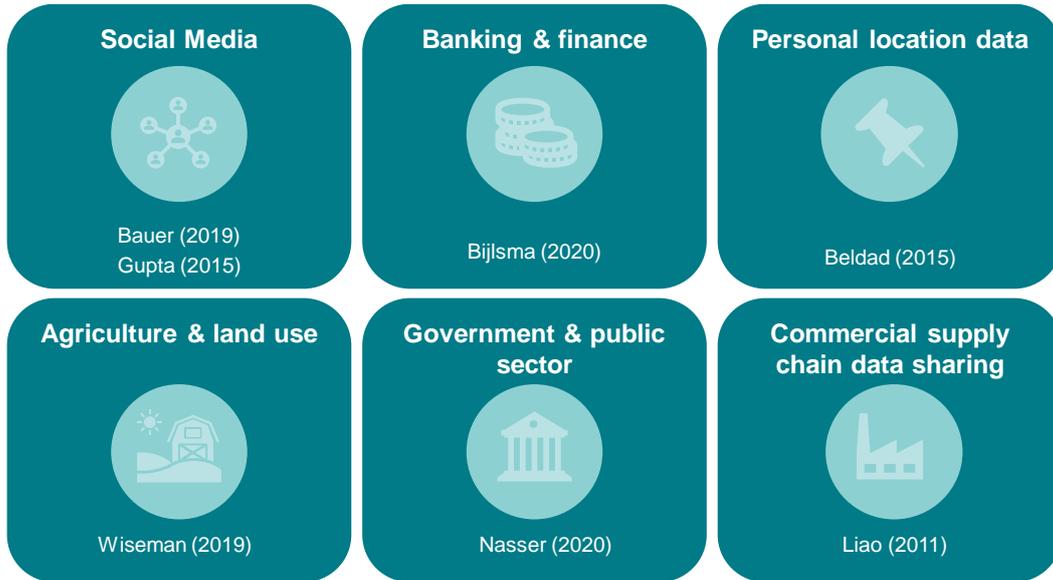
If this engagement involves organisations communicating with the wider public, stakeholders made the point that wherever possible this communication should be personalised. For example, this could include letting a patient know exactly how their data will be used and how it will contribute to high quality care for others.

### Ecosystem context

The studies we have used refer to a range of different sectors, geographies and data ecosystems. This in itself will influence the survey responses obtained and the resulting estimates of the impact of trust on data sharing. For example, in Figure 16 below we illustrate the sector diversity of the underlying estimates.

Most sectors only contain one study which provides applicable quantitative evidence. Therefore, we cannot draw any robust overarching conclusions regarding the impact of trust in different individual data ecosystems. However, we can be confident that the impact estimates linked to each of these studies will in part depend on the context and ecosystem in which they were carried out. As a result some of the observed variation will be linked to this context.

**Figure 16 Sectoral contexts of underlying evidence**

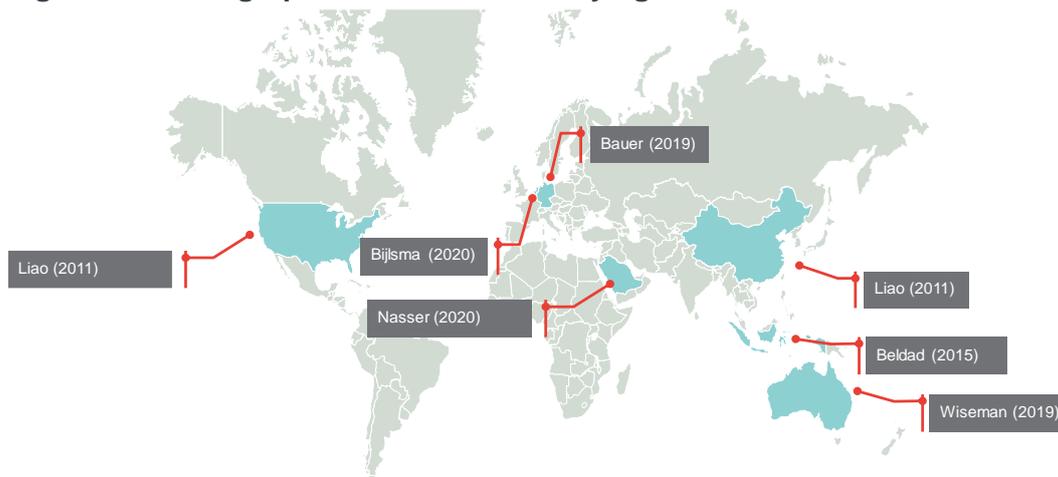


Source: Frontier Economics based on evidence review

Our qualitative engagement confirmed that contextual factors are crucial to understanding the role of trust in influencing data sharing. For example, all stakeholders agreed that sharing more sensitive forms of data such as personal financial information or health records would require higher levels of trust and perceived security in advance. The prevalence of these different types of data will clearly vary depending on the ecosystem or sector in question.

In addition, the studies we have used to parameterise our framework cover numerous geographic contexts within North America, Europe, Asia and Oceania (see Figure 17 below).

**Figure 17 Geographic context of underlying evidence**



Source: Frontier Economics review of evidence

Note: Geographic location was unavailable for Gupta (2015)

As we emphasised when describing the conceptual impact of trust via our logic model, norms and unwritten attitudes will play a role in determining baseline levels

of trust and how questions are intercepted. These norms will likely vary substantially both between individuals and also across different regions.

As such, a certain proportion of the underlying variation in estimates we have presented will likely be due to the geographic contents in which the studies were carried out. However, as mentioned above, we cannot robustly determine how large this proportion will be.

Importantly, interviewees emphasised that while different ecosystems are unique, efforts should be made to exhibit and build trust horizontally rather than purely on a sector-by-sector basis. This is because certain data ecosystems cannot be treated in isolation, particularly from the point of view of end users: Actors will simultaneously share data with different organisations in different sectors, and this data will be interconnected across ecosystems.

Achieving this level of consistency, while still accounting for differences in context, may involve setting out general principles and/or categories of intervention to build trust. These can then be adapted and tailored for flexible implementation.

## 3.4 Impact of changes in trust in data ecosystems

### 3.4.1 Assessing the impact of changes in trust in data ecosystems

Survey results can only measure the relationship between trust and data sharing in static terms and are not able to capture actual data-sharing behaviour over time. To address this, we have augmented the survey evidence with a wide set of “natural experiments” looking at shocks to trust levels in real-life settings and their impact on data-sharing behaviour.

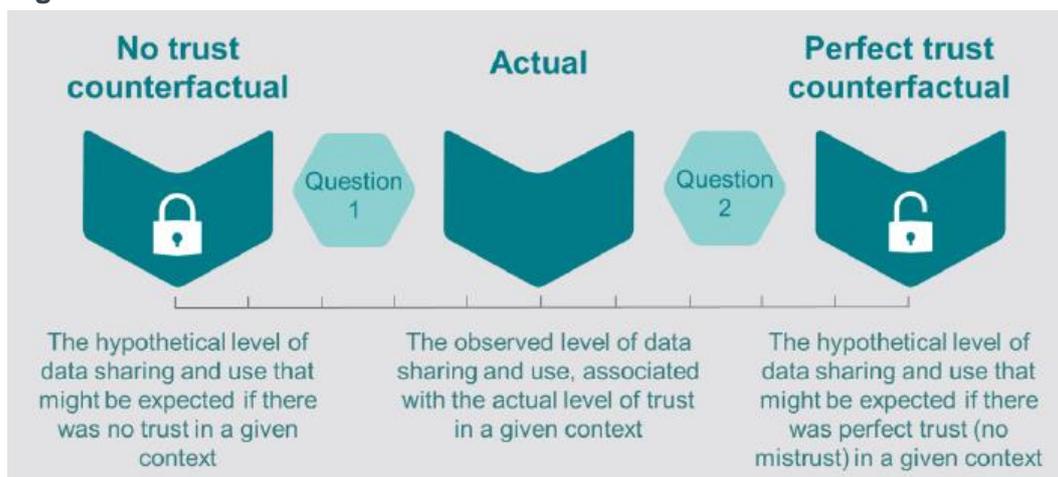
This analysis complements our assessment of the aggregate impact of trust on data sharing, to test whether changes in trust have a causal impact on data-sharing behaviour in real-life settings.

#### Defining the counterfactual

To evaluate the impact of trust shocks on data-sharing behaviour in these real-life examples, we have contextualised it via a theoretical framework. This framework defines two alternative counterfactual scenarios.

Specifically, we have considered the amount of data sharing, collection and use which is consistent with actual levels of trust and compared this to two counterfactual scenarios (see Figure 18 below for further detail):

- *No trust counterfactual*: the hypothetical level of data sharing and use that might be expected in there was no trust in a given context;
- *Perfect trust counterfactual*: the hypothetical level of data sharing and use that might be expected if there was perfect trust in a given context.

**Figure 18** Movements in trust

Source: Frontier Economics

Ideally, to examine the impact of trust at the global level, we would measure movements from actual levels of trust to a scenario where trust is completely eroded (Question 1), and vice versa to one where there was no mistrust at all (Question 2). However, the real-life examples we are looking to examine are only able to partially proxy these theoretical trust movements:

- Question 1: impact of moving from current levels of trust to somewhat “less” trust (but not possible to determine a real-life scenario where trust is fully eroded);
- Question 2: impact of moving from current levels of trust to somewhat “more” trust – not necessarily reflecting a perfect trust counterfactual (e.g. the introduction of GDPR increases trust by making data-sharing practices more secure, but there are other aspects of trust that GDPR doesn’t address, so post-GDPR we are still not in a perfect trust scenario).

### Calibrating the dynamic impact of changes in trust in data ecosystems using natural experiments

To consider the impact of exogenous shocks in trust on data sharing and thereby validate the robustness of the correlational estimates, we evaluate a set of natural experiments. As we described above, the main advantage of these natural experiments in relation to survey evidence is that they allow us to gauge data-sharing behaviour over time as a result of a shock in trust at a given point in time.

While natural experiments may provide an indication of the impact of an exogenous shock in trust, we cannot be certain that any impacts are entirely due to changes in trust. It may be, for example, that a data breach leads to a loss of trust and a greater understanding of the actual costs and benefits of sharing data. Both of these factors could then impact subsequent data sharing.

### 3.4.2 Moving from actual trust to no trust

#### How trust breaks down and real-life examples

There are several routes by which trust can be eroded. We have listed out possible channels in Figure 19 below which draws on O'Hara (2018).<sup>48</sup>

**Figure 19** How trust breaks down

Channel	Example
Misrepresentation	"An example of misrepresentation of a data institution's ability might be if it provides data contributors with assurances about data protection, but does not have good security measures in place, which comes to light when it suffers a data breach."
Misunderstanding	"A misunderstanding of the audience could arise when a data institution claims to have a social benefit purpose, such as improving healthcare, and furthers that purpose by sharing data with private sector organisations, such as pharmaceutical companies. The data institution may see their purpose at a system level, and the development of new drugs as ultimately contributing to better healthcare. But data contributors may have a narrower view, and see companies making profit through using data about them as unacceptable."
Inability to determine trustworthiness	"A stakeholder might be unable to determine how much to trust a data institution if it cannot get access to information about the sources of the data it stewards, or the process through which requests for access are granted."
Failure to communicate changes	"A failure to communicate a change in circumstances might arise if a data institution is absorbed into, or spun out of, another organisation and this is not communicated clearly; or if new data contributors or data users enter into agreements with the data institution, particularly if those organisations have poor reputations."

Source: O'Hara (2018)

The real-life examples of trust erosion we have considered mostly relate to breaches of trust due to ***misrepresentation***. In the past this has occurred because:

- The data institution was found sharing data with third parties without the consent of the data subject;
- The data institution suffered a data breach as a result of infrastructure failure/not following security protocols.

Figure 20 below outlines the examples of negative trust shocks we looked into to gauge the impact of a decrease in trust on data sharing. This is representative of a variety of industries and ecosystems but it is not intended to be exhaustive.

<sup>48</sup> [https://eprints.soton.ac.uk/341800/1/ohara\\_trust\\_working\\_paper\\_aug\\_2012.pdf](https://eprints.soton.ac.uk/341800/1/ohara_trust_working_paper_aug_2012.pdf)

**Figure 20 Breaches in trust through misrepresentation**

Trust breach	Year	Type of incident
Cambridge Analytica - Facebook data breach	2018	handing data to third party without consent
NHS care.data information scheme closed	2016	handing data to third party without consent
Pharmacy2U fined by ICO on selling customer information to marketing companies	2015	handing data to third party without consent
Google DeepMind/Royal Free NHS Foundation Trust partnership fails to comply with GDPR		internal mishandling of sensitive data
Talktalk data breach	2015	data breach due to infrastructure failure
Experian data breach	2020	data breach due to infrastructure failure
Ashley Madison data breach	2015	data breach due to security failure
TSB online banking failure	2018	online banking failure causing massive disruption to customers, and to some extent, revelation of personal data to other customers

Source: Frontier & ODI

### Impact of breaches in trust

To assess the impact of exogenous negative shocks to trust on data sharing, we have compared statistics and quantified commercial impacts on each of the selected incidents.

This comparison exercise enabled us to test the direction of causality (i.e. whether, as we would expect, a negative shock to trust has led to a decrease in data sharing over time), and draw out some key stylised facts on how this is likely to vary based on the context of the case at hand.

As expected, in all cases analysed, a breach in trust caused reputational damage to the data institution as well as a decrease in the number of individuals willing to share data with that particular institution. This is fully consistent with our parameterised trust model.

In addition, the effects were not homogenous. Rather, in some cases the negative shock to trust had a permanent impact, whereas in other cases the reduction in data sharing was temporary. For example:

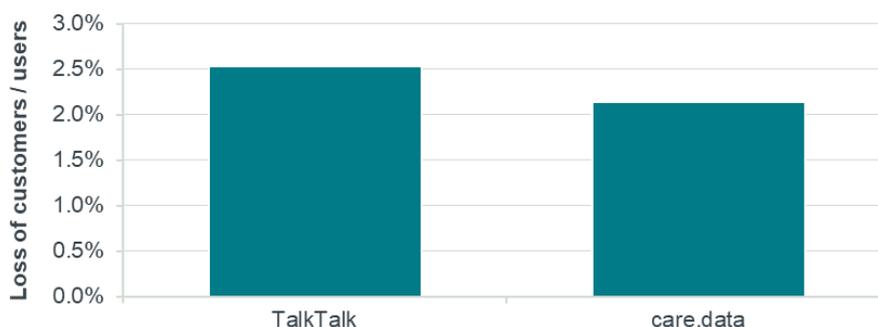
- In 2018 it came to light that millions of Facebook users' personal data was acquired without consent by Cambridge Analytica. This information was predominantly to be used for political advertising. After this incident, many users professed they did not trust Facebook to hold their data and temporarily

cancelled their accounts.<sup>49</sup> However, only months later, users had by and large resumed their activity on the platform.<sup>50</sup>

- care.data was an attempt by the then Health and Social Care Information Centre to extract data from GP surgeries into a central database. Patient data would be used in anonymised form by health care researchers, including those outside the NHS, such as academic institutions or commercial organisations. In 2016, concerns about the NHS care.data information scheme were raised in relation to the sharing of sensitive medical information with commercial companies without the explicit consent of patients. The scheme was closed after more than a million users opted out.<sup>51</sup>
- In 2015 over 150,000 of TalkTalk's customers had their personal details hacked in a cyber-attack. In the following months 101,000 accounts were closed, 95,000 of which were closed as a direct result of the data breach. In 2018 TSB customers were locked out of their accounts following a failed IT upgrade. TSB's online banking failure resulted in 80,000 customer switching accounts to another bank.<sup>52</sup>

We illustrate the number of NHS patients opting out of care.data and the number of customers shifting away from TalkTalk in proportional terms below (Figure 21).

**Figure 21** Examples of trust shocks' impact



Source: Frontier Economics analysis

[https://www.theregister.com/2016/04/20/one\\_million\\_patients\\_have\\_opted\\_out\\_of\\_caredata/](https://www.theregister.com/2016/04/20/one_million_patients_have_opted_out_of_caredata/),  
<https://www.theguardian.com/business/2016/feb/02/talktalk-cyberattack-costs-customers-leave>

<sup>49</sup> Studies have shown that users' awareness of data sharing practices increased right after the news became public. For example, Shipman et al (2020) found via a survey that participants with the greatest awareness of the news story's details have more polarized attitudes about reuse, especially the reuse of content as data. They express a heightened desire for data mobility, greater concern about networked privacy rights, increased scepticism of algorithmically targeted advertising and news, and more willingness for social media platforms to demand corrections of inaccurate or deceptive content. They found that 1/5th of the interviewees deleted their account as a result of the incident. Source: Shipman, F.M. and Marshall, C.C., 2020, April. *Ownership, Privacy, and Control in the Wake of Cambridge Analytica: The Relationship between Attitudes and Awareness*. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

<sup>50</sup> This is confirmed by a study looking at the impact of the Cambridge Analytica incident on user behaviour over time. Edwards (2019) tracked US Facebook user attitudes towards using the platforms over the course of 4 months after the incident became public. Results suggest that the decrease in Facebook's user base was so small that it was likely to make little to no noticeable damage to Facebook's profits or reputation. Source: Edwards, J.L., 2019. *An Examination of Consumers' Social Media Trust In the Wake of the Facebook and Cambridge Analytica Scandal*

<sup>51</sup> <https://www.wired.co.uk/article/care-data-nhs-england-closed>  
<https://www.bbc.co.uk/news/health-26259101>

<sup>52</sup> Porcedda, M.G. and Wall, D.S., 2019. Cascade and Chain Effects in Big Data Cybercrime: Lessons from the TalkTalk hack, proceedings of WACCO 2019: 1st Workshop on Attackers and Cyber-Crime Operations, IEEE EuroS&P 2019, Stockholm, Sweden, June 20, 2019.

We explored the persistence of shocks to trust in terms of impact on data sharing as part of our qualitative engagement. Stakeholders highlighted a number of factors which could lead to a bigger and longer-lasting impact following an initial loss of trust. We illustrate these drivers in Figure 22 below.

These qualitative insights serve to highlight that a number of contextual factors such as data subjects' outside options, the nature of the connection between data subjects and data stewards as well as the motives of the data steward all have a role to play.

**Figure 22 Factors influencing persistence of negative trust shocks**



Source: Frontier Economics based on qualitative engagement

Across all of the examples we have described above it is not possible to entirely attribute the loss in users and consumers and knock-on impacts on data sharing to a loss in trust. There might be other important contributing factors leading to a decrease in data sharing (which are likely to be somewhat related to trust in a data steward, for example perception of how competent the institution is in handling data or a greater understanding of the actual risks involved).

Our qualitative engagement also revealed that negative shocks to trust which are directly related to a single actor can have wider spill-over ramifications for an entire sector or ecosystem. In the case of healthcare research, we were told that the care.data episode meant that accessing data for legitimate research reasons became significantly more difficult for a sustained period of time. This occurred for two reasons:

- Patients were more inclined to opt out of other data collection exercises due to a loss of trust. This loss in trust was not spread equally across the population. Therefore, research datasets were more likely to be biased. As a result there was less appetite from policy-makers to act on the conclusions of any work that was carried out.

- Data stewards were increasingly conscious of the importance of maintaining patients trust and were less inclined to share the data that they held with other researchers.

Both of these factors meant that less socially beneficial research was carried out and effective policymaking and evaluation were hampered. However, interviewees also emphasised that in some ways this shock to trust may have led to some positive developments as it forced organisations to improve their data collection and sharing processes. Specifically, data users told us that the processes for applying for data are now more streamlined and efficient compared to the situation before.

Stakeholders from the financial services sector made similar observations around how a major loss of trust associated with a single provider can have negative impacts across the sector via a wider loss of trust with regards to an entire ecosystem.

Other stakeholders emphasised that these type of negative spill-over effects apply even more broadly and a loss of trust in one ecosystem can affect public willingness to share data in a completely separate sector. We were told that often members of the public only hear about this type of personal data sharing when something goes wrong. As a result they may default to overestimating the associated risks.

### 3.4.3 Moving from actual trust to perfect trust

#### How trust is built and real life examples

Symmetrically to the exercise above, to define movements from actual levels of trust to perfect trust, we have identified what the possible ways in which the status quo (i.e. actual levels of trust) can improve, by highlighting conceptual mechanisms through which trust is built.<sup>53</sup>

The main levers through which trust is built are:

- external factors:<sup>54</sup> Laws, regulations and norms, contracts, penalties, standards, codes of conduct, ethical and organisational design, organisational governance.
- internal factors: Reputation, competence and skills, presence of a pre-existing trust relationship.

Trust-building interventions may include several policy levers, put in place with the objective of increasing trust in a data ecosystem.

Some examples are listed below. This set of examples is representative of a variety of industries and contexts, but, as above, it is not intended to be exhaustive.

---

<sup>53</sup> As per our logic model in Section 2.

<sup>54</sup> As outlined in the logic model, organisations often have a degree of agency over these factors, e.g. how the organisation reacts to the external regulation, norms and its decisions on organisational governance and design.

**Figure 23 Trust-building interventions**

Intervention	Year	Purpose
GDPR	2018	The GDPR's primary aim is to give control to individuals over their personal data and to simplify the regulatory environment for international business by unifying the regulation within the EU.
SHARE		SHARE was created to establish a register of people interested in participating in health research. SHARE can then check whether individuals might be suitable for health research studies. It provides an easy way for people to get involved and clearly articulates benefits for research that donors might benefit from. Also, importantly the data is held securely with clear opt-out mechanisms in place to help build trust.
Open Banking		Open Banking is a secure way to give providers access to an individual's or business' financial information. Providers enrol in Open Banking and are subject to FCA regulation in order to provide services. This regulatory oversight is designed to build trust and encourage engagement. New providers can use a test environment to try out their service once they have completed enrolment.
Establishment of data trusts		Data trusts are an approach to looking after and making decisions about data. They involve one party authorising another to make decisions about data on their behalf, for the benefit of a wider group of stakeholders. For example, UK Biobank was set up in 2006 to steward genetic data and samples from 0.5m people and takes the form of a charitable company with trustees.

Source: *Frontier Economics*

### Impact of trust-building interventions

These interventions mostly focused on designing a set of rules that data intermediaries can conform to, to increase their trustworthiness. Their effect on data sharing may be non-linear over time:

- at the beginning, the intervention might pose an additional burden on a data institution, therefore deterring data sharing in the ecosystem; but
- over time, as data intermediaries become more comfortable with the regulation and data contributors are able to place greater trust in them, data sharing may gradually increase, possibly overtaking pre-intervention levels of data sharing.

The highest-quality quantitative information relates to the introduction of GDPR given its wide reach and high profile. We have summarised some of the existing evidence below.

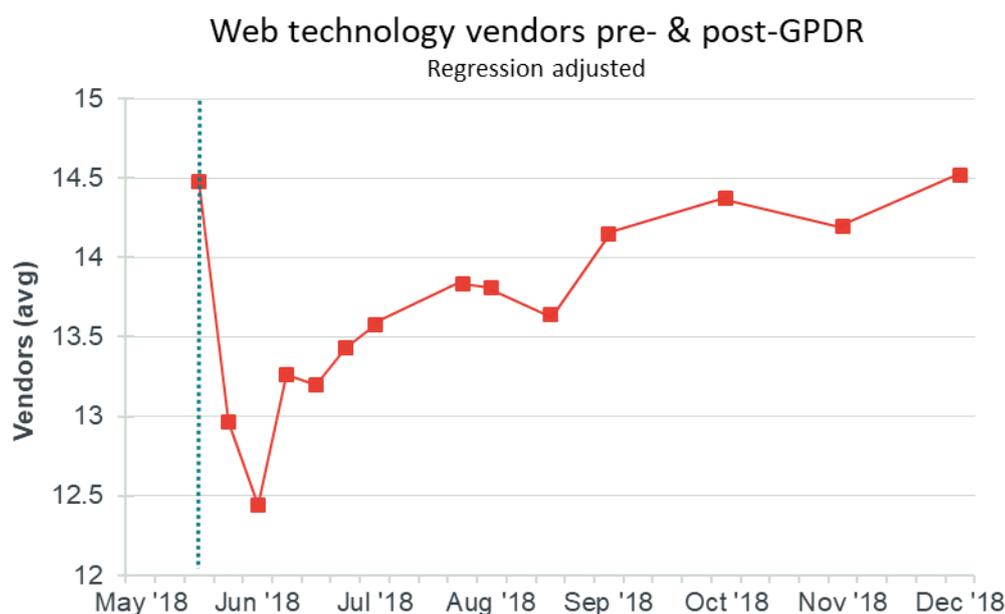
The non-linearities in terms of timing is confirmed by academic studies investigating the impact of the introduction of GDPR in 2018 on data sharing. Johnson et al. (2020) found that the week after the enforcement of GDPR, online use of web technology vendors fell by 15% in the EU.<sup>55</sup> Similarly, Goldberg (2019)

<sup>55</sup> This can be a proxy to measure data sharing, collection and use in the online space. Websites rely on inputs from specialized web technology vendors to provide various services. For instance, "audience

found that recorded page views and revenues fell by about 10% for EU users after the GDPR's enforcement deadline<sup>56</sup> and Urban (2020) concluded that the number of ID-syncing connections decreased by around 40% around the time GDPR went into effect.<sup>57</sup>

However, interestingly in several cases these effects reversed over time. Specifically, Johnson (2020) found that the short-run effect eroded over time and the level of data sharing return to levels observed before the introduction of the policy (see Figure 24 below). Likewise Urban's (2020) long-term analysis showed a slight rebound in ID-syncing connections.

**Figure 24 Evolution of average web technology vendors per EU website**



Source: Chart derived from Johnson (2020)

Similarly Zhang (2019) found that a company's voluntary adoption of GDPR led to positive effects on their customers' intention to disclose information to that company and increased customer trust.

Stakeholders from both the financial services and health ecosystems agreed that this type of formal intervention can boost trust and facilitate greater data sharing. Specifically, we were told that these type of rules can shorten approval times when accessing data and that standardisation can help to reduce the number of individual trust linkages that are needed.

---

measurement" vendors record user site visits and generate statistics on user characteristics, onsite activities, and referral channels (examples are Google Analytics, Adobe Audience Manager or smaller competitors). The authors argue that websites who do not have the resources to understand and comply with GDPR would stop using these services, implying lower levels of data sharing an use overall. Their analysis confirms that, the week after the GDPR's enforcement, website use of web technology vendors for EU users fell by 15%.

<sup>56</sup> They argue this is because privacy regulation increases the firm's cost of collecting consumer data which makes matching with users more costly. As such, the GDPR has the potential to reduce both the amount of traffic to a website as well as the amount and quality of web outcome data stored for analytics purposes.

<sup>57</sup> ID-syncing connections are a measure of the number of users who are being tracked (and their personal data collected) by websites to provide personalised advertising.

As we described above, stakeholders agreed that they needed to be implemented alongside more organic development of trust over time and that regulation alone was insufficient. Other interviewees noted that these type of interventions will have differential impacts on different types of actor within a specific ecosystem. For example, we were told that smaller organisations or individuals may not have the time or expertise to fully engage with regulatory systems and rules that are in place and may therefore feel that sharing data is more risky than it actually is. This serves to emphasise the importance of clearly communicating why certain rules are in place and how they operate.

In addition, some interviewees in the financial services ecosystem noted that just because an intervention of this type leads to a significant impact in data sharing, it does not necessarily imply that there has been an increase in trust.

For example, we were told that in some cases providers are forced to share data with organisations that they do not necessarily trust because the other organisations have been certified as a legitimate ecosystem participant by the central regulator. This clearly reduces the number of trust interactions required (the data holder now only needs to trust the regulator's capability to certify appropriately rather than trust all possible new participants). However, we were told that without having this choice providers feel as though trust is no longer a relevant concept.

In order, to mitigate these concerns we were told that having ex-ante rules and regulations needed to be accompanied by clear information on the sanctions that will be applied if something goes wrong and the redress mechanisms that will be enforced.

### 3.5 Economic impact of data sharing

The final stage of our calibration involved linking greater data sharing with economic outcomes.

As we set out in our logic model, there are several channels through which data sharing can benefit the economy and society, including:

- more higher-quality information circulates and more insights are generated within ecosystems;
- greater competition within and across sectors;
- new business opportunities, increased efficiency and innovation.

A detailed examination of the economic impact of data sharing, collection and use is beyond the scope of this study. However, we have reviewed the available academic and policy literature of the social and economic impact of data access and sharing. This has allowed us to provide an indicative, high-level estimate of the economic and social impact of trust.

The OECD estimates the economic value of data sharing as threefold (2019).<sup>58</sup> Evidence shows that greater levels of data access and sharing can generate positive social and economic benefits:

- increase the value of data to holders (direct impact);
- but also help create 10 to 20 times more value for data users (indirect impact);
- and 20 to 50 times more value for the wider economy (induced impact).

However, the OECD also notes that quantifying the overall economic and social value of data access and sharing is particularly challenging. This is because available studies tend to differ in terms of the scope of the sectors (e.g. public-sector versus private-sector data sharing), the types of data (e.g. personal, proprietary or public), and the degrees of data openness (and the arrangements included such as open data). Available studies also tend to differ in the methodologies employed.<sup>59</sup> This is fully consistent with the heterogeneity we have observed in the present study when reviewing evidence on the impact of trust on data sharing, collection and use.

Overall, the evidence gathered<sup>60</sup> suggests that data sharing can help generate social and economic benefits worth between 0.1% and 1.5% of gross domestic product (GDP) in the case of public-sector data, and between 1% and 2.5% of GDP (in a few studies up to 4% of GDP) when also including private-sector data.<sup>61</sup> It is worth noting that most of the studies which examine the economic value of data sharing focus on organisations sharing and re-sharing data (rather than individual data sharing).

Scaling to the GDP of the 20 largest economies in 2019, estimates suggest that data sharing could unlock between 700 billion and 1.75 trillion US\$ in value. Linking these wider results to our framework suggests that a 25% increase in trust could therefore generate an additional 47.3 to 118.3 billion US\$.

---

<sup>58</sup> <https://www.oecd-ilibrary.org/sites/90ebc73d-en/index.html?itemId=/content/component/90ebc73d-en#:~:text=Overall%2C%20these%20studies%20suggests%20that,also%20including%20private%2Dsector%20data.>

<sup>59</sup> OECD (2019). Enhancing access to and sharing of data: reconciling risks and benefits for data re-use across societies. <https://www.oecd-ilibrary.org/sites/90ebc73d-en/index.html?itemId=/content/component/90ebc73d-en#back-endnotea3z5>

<sup>60</sup> Evidence gathering on the economic impact of data access and sharing was mainly guided by the OECD's own review available at <https://www.oecd-ilibrary.org/sites/90ebc73d-en/index.html?itemId=/content/component/90ebc73d-en#back-endnotea3z5>

<sup>61</sup> Aggregate estimates are computed by drawing from a large pool of available studies on the impact of data sharing and access in the public and private sector. The result of the aggregation is likely to be one-off effect (or, alternatively an effect that develops over a number of years), rather than an effect that can be expected repeatedly year after year.

**Figure 25** Scaling the economic impact of trust through data sharing

Effects calibration	Impact	Unit
Estimated economic benefits from data access and sharing, for both public and private sector data	1%-3%	Lower bound-upper bound, as average % of GDP
2019 GDP for the 20 largest economies in the world	70.1	Trillion, US\$
Scaled economic impact of data sharing	1%-3% of 70.1 trillion = 701 – 1,753	Billion, US\$
Scaled economic impact of a 25% increase in trust through data sharing	6.75% of 701 – 1,753 = 47.3 – 118.3	Billion, US\$

Source: Frontier Economics calculation based on WEO and OECD data.

Note: A 25% increase in trust corresponds to a 1 point increase in the trust Likert Scale.

The economic impact of a 25% increase in trust is achieved through a simple scaling exercise which assumes the relationship between trust and data sharing, collection and use (and the subsequent relationship between data sharing, collection and use and economic impact) are linear. Therefore, these figures are able to provide only an indicative estimate of the economic impact of trust, without capturing nuances on how this is likely to vary based on circumstances and sectors.

### 3.6 Evidence gaps and direction of future research

By linking the results of our in-depth research on the impact of trust on data sharing and existing evidence on the economic value of data, this work has made an important first step towards assessing the economic impact of trust in data ecosystems.

While our research has generated valuable preliminary results on the economic impact of trust in data ecosystems, it has also revealed some key gaps in the existing evidence base. In particular, the study's original aim was to connect a key evidence gap between two key existing strands of literature:

- the mechanisms through which trustworthiness and trust can be established and maintained within data ecosystems;
- the economic impact of data sharing, collection and access/use.

**Figure 26 The research question**



Source: *Frontier Economics*

Our in-depth literature review revealed that:

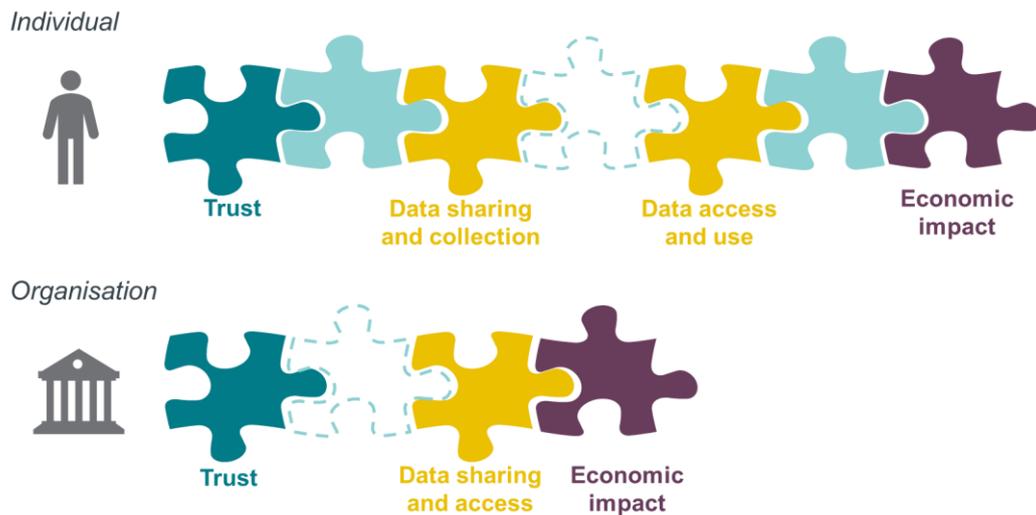
- the volume of available quantified evidence on the impact of trust has so far been skewed towards analyses of individuals sharing data about themselves.
- However, studies on the economic value of data have largely focused on the economic value arising from organisations re-sharing data or benefiting from greater access to data.

As a result of this, the research highlighted two key gaps in the existing literature, as shown in Figure 27 below:

- Evidence on the link between individuals sharing more data about themselves and organisations (and individuals) getting greater access to data, which in turn lead to economic impacts;
- Quantified evidence on the impact of trust between organisations on the amount of data sharing.<sup>62</sup>

<sup>62</sup> While our review revealed some evidence on this, the volume of studies is relatively low compared to the studies on individual data sharing, so further evidence in this area would be beneficial.

**Figure 27 Outstanding gaps in the evidence**



Source: Frontier Economics

These gaps make it challenging to draw definitive conclusions on the economic impact of trust in a data ecosystem, but serve as a guide to direct future research efforts. Three key areas to explore in future research are:

- whether increased data sharing by individuals leads to greater data access and use by organisations – if so, what are the drivers, if not, what are the barriers;
- whether increased data sharing has any direct economic impact; and finally
- additional quantitative research on whether increased trust between organisations might increase data sharing.

## 4 CONCLUSION

Overall this work has shown that there is robust quantified evidence that greater trust consistently leads to increased data sharing. This confirms existing anecdotal evidence and justifies ongoing efforts to design mechanisms to boost trust.

In particular, our qualitative engagements confirmed how in cases where there is scope to achieve significant improvements in trust, the relevant effect size will be large. This could occur if, for example, actors within a specific data ecosystem move from an average position of being neutral with regards to data sharing to being willing to share data. Crossing these key threshold points may only require a relatively small change in attitudes but still have a large impact on data sharing, collection and use.

While evidence gathered as part of this work suggests trust might have a larger impact on sharing in contexts where the baseline level of trust is low, our work suggests that the role of trust alone in fostering data sharing might be moderate on average.

A likely reason for this might be that the *elasticity* of data sharing to trust, i.e. how strong of an impact trust can have on data sharing, will depend not just on baseline trust alone, but also on baseline levels of other key determinants of data sharing. Other drivers of data sharing and access include building the foundations to enable data flows such as infrastructure and standards, mechanisms to mitigate commercial, regulatory and legal risks associated with data sharing, and addressing gaps in knowledge and incentives to share and access data.

For example, in ecosystems with more established infrastructure and standards for data sharing, an increase in trust has the potential to cause larger increases in sharing.

Our work aimed to analyse trust as one of the levers through which data sharing, collection and use can be improved. While focusing on trust alone enabled us to answer this question in-depth, our conclusions reflect a partial equilibrium exercise and are unable to reach conclusions on the interdependencies between trust and other determinants of sharing. Exploring wider determinants of data sharing and how they can complement increases in trust will be an important area for future research.

Both our qualitative engagement and our summary of quantitative research emphasises the importance of context when considering the role of trust in data ecosystems. The specific trust linkages that exist and their maturity will vary across and within certain sectors.

Our examination of changes in trust over time confirm that trust has a causal impact on data sharing. Our stakeholder engagement highlighted a range of factors that will play a role in determining the scale of the impact of trust on data sharing and the dynamic persistence of effects.

Further research in this area would be beneficial. As we have noted, a relatively small set of papers examine the actual impact of trust on data sharing using a quantified approach. Quantitative evidence is particularly limited on trust relationships and data-sharing behaviour between organisations. However, most

of the studies have been carried out recently, which highlights this area as an active branch of research.

Our research identified key gaps in the existing evidence, which make it challenging to link our core findings on the impact of trust on data sharing to the wider literature on the economic impact of data sharing, collection and use. Adding to the existing evidence base will allow future work to determine how the economic value of trust varies across different activities (e.g. data sharing, data collection and data use) as well as different relationships (e.g. between individuals and organisations or between organisations specifically).

## 5 BIBLIOGRAPHY

AGCOM, 2018. Propensity of online users to consent to the use of their personal data by third parties when using online services (*translated from Italian*).

Allen, D.W., Berg, A., Berg, C., Markey-Towler, B. and Potts, J., 2019. Some economic consequences of the GDPR. *Allen, DWE, Berg A, Berg C, Markey-Towler, B and Potts, J (2019) 'Some Economic Consequences of the GDPR', Economics Bulletin, 39(2), pp.785-797.*

Anderson, N. and Edwards, K., 2010, December. Building a chain of trust: using policy and practice to enhance trustworthy clinical data discovery and sharing. In *Proceedings of the 2010 workshop on governance of technology, information and policies* (pp. 15-20).

Anhalt-Depies, C., Stenglein, J.L., Zuckerberg, B., Townsend, P.A. and Rissman, A.R., 2019. Tradeoffs and tools for data quality, privacy, transparency, and trust in citizen science. *Biological Conservation, 238*, p.108195.

Ayaburi, E.W. and Treku, D.N., 2020. Effect of penitence on social media trust and privacy concerns: The case of Facebook. *International Journal of Information Management, 50*, pp.171-181.

Ba, S. and Pavlou, P.A., 2002. Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS quarterly*, pp.243-268.

Bansal, G. and Gefen, D., 2010. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision support systems, 49(2)*, pp.138-150.

Bauer, P.C., Keusch, F. and Kreuter, F., 2019. Trust and cooperative behavior: Evidence from the realm of data-sharing. *PloS one, 14(8)*, p.e0220115.

Beldad, A. and Kusumadewi, M.C., 2015. Here's my location, for your information: The impact of trust, benefits, and social influence on location sharing application use among Indonesian university students. *Computers in human behavior, 49*, pp.102-110.

Bijlsma, M., van der Crujisen, C. and Jonker, N., 2020. Consumer propensity to adopt PSD2 services: trust for sale?.

Blanco-Perez, C.B., 2019. Transparency in empirical economic research. *IZA World of Labor*.

Brown, A.J., 2020. "Should I Stay or Should I Leave?": Exploring (Dis) continued Facebook Use After the Cambridge Analytica Scandal. *Social Media+ Society, 6(1)*, p.2056305120913884.

Cao, Q.H., Giyyarpuram, M., Farahbakhsh, R. and Crespi, N., 2020. Policy-based usage control for a trustworthy data sharing platform in smart cities. *Future Generation Computer Systems, 107*, pp.998-1010.

Cao, Q.H., Khan, I., Farahbakhsh, R., Madhusudan, G., Lee, G.M. and Crespi, N., 2016, May. A trust model for data sharing in smart cities. In *2016 IEEE International Conference on Communications (ICC)* (pp. 1-7). IEEE.

Catoiu, I., Orzan, M., Macovei, O.I. and Iconaru, C., 2014. Modelling users' trust in online social networks. *Amfiteatru Economic Journal, 16(35)*, pp.289-302.

Centre for Data Ethics and Innovation, 2020. Addressing trust in public sector data use. Independent report prepared for the UK Government.

- Chavanne, D., 2018. Generalized trust, need for cognitive closure, and the perceived acceptability of personal data collection. *Games*, 9(2), p.18.
- Cole, A. and Towse, A., 2018. Legal Barriers to the Better Use of Health Data to Deliver Pharmaceutical Innovation. *London: Office of Health Economics*.
- Cottrill, C.D., 2020. MaaS surveillance: Privacy considerations in mobility as a service. *Transportation Research Part A: Policy and Practice*, 131, pp.50-57.
- Critchley, C., Nicol, D. and Otlowski, M., 2015. The impact of commercialisation and genetic data sharing arrangements on public trust and the intention to participate in biobank research. *Public Health Genomics*, 18(3), pp.160-172.
- Dahmm, H., 2020. Laying the Foundation for Effective Partnerships: An Examination of Data Sharing Agreements.
- Deloitte, 2013. Market assessment of public sector information. *A report prepared for the UK Department for Business Innovation and Skills*.
- Deloitte, 2017. Assessing the value of TfL's open data and digital partnerships.
- Diekmann, A., Jann, B. and Wyder, D., 2004. Trust and reputation in Internet auctions.
- DotEcon, 2015. Independent evaluation of the OFT's 2006 market study into the Commercial Use of Public Information (CUP). *A report prepared for the CMA*.
- Edwards, J.L., 2019. *An Examination of Consumers' Social Media Trust In the Wake of the Facebook and Cambridge Analytica Scandal* (Doctoral dissertation).
- European Commission, 2019. European legislation on open data and the re-use of public sector information. *Policy report*.
- Fecher, B., Friesike, S., Hebing, M., Linek, S. and Sauermaun, A., 2015. A reputation economy: results from an empirical survey on academic data sharing. *arXiv preprint arXiv:1503.00481*.
- Foster, D. and Arnold, Z., 2020. Antitrust and Artificial Intelligence: How Breaking Up Big Tech Could Affect the Pentagon's Access to AI.
- Freeman, J.L., Caldwell, P.H. and Scott, K.M., 2020. The Role of Trust When Adolescents Search for and Appraise Online Health Information. *The Journal of Pediatrics*, 221, pp.215-223.
- Gillespie, N., Dietz, G. and Lockey, S., 2014. Organizational reintegration and trust repair after an integrity violation: A case study. *Business Ethics Quarterly*.
- Goldberg, S., Johnson, G. and Shriver, S., 2019. Regulating Privacy Online: The Early Impact of the GDPR on European Web Traffic & E-Commerce Outcomes. *Available at SSRN 3421731*.
- Grund, C. and Gürtler, O., 2008. The Effect of Reputation on Selling Prices in Auctions. *Jahrbucher fur Nationalokonomie & Statistik*, 228(4).
- Grünwald, P. and Reisch, T., 2020. The trust gap-privacy perceptions of location data for energy services in the UK. *Energy Research & Social Science*, 68, p.101534.
- Gupta, A. and Dhami, A., 2015. Measuring the impact of security, trust and privacy in information sharing: A study on social networking sites. *Journal of Direct, Data and Digital Marketing Practice*, 17(1), pp.43-53.
- Hafen, E., 2019. Personal data cooperatives—a new data governance framework for data donations and precision health. In *The Ethics of Medical Data Donation* (pp. 141-149). Springer, Cham.

- Han, G. and Dong, M., 2017. Sustainable regulation of information sharing with electronic data interchange by a trust-embedded contract. *Sustainability*, 9(6), p.964.
- Hart, P. and Saunders, C., 1997. Power and trust: Critical factors in the adoption and use of electronic data interchange. *Organization science*, 8(1), pp.23-42.
- Hate, K., Meherally, S., Shah More, N., Jayaraman, A., Bull, S., Parker, M. and Osrin, D., 2015. Sweat, skepticism, and uncharted territory: a qualitative study of opinions on data sharing among public health researchers and research participants in Mumbai, India. *Journal of Empirical Research on Human Research Ethics*, 10(3), pp.239-250.
- Hawlitshchek, F., Stofberg, N., Teubner, T., Tu, P. and Weinhardt, C., 2018. How corporate sharewashing practices undermine consumer trust. *Sustainability*, 10(8), p.2638.
- IDC, and the Lisbon Council 2016. First Report on Facts and Figures Updating the European Data Market Study Monitoring Tool. *Study prepared for the European Commission*.
- Jakku, E., Taylor, B., Fleming, A., Mason, C., Fielke, S., Sounness, C. and Thorburn, P., 2019. "If they don't tell us what they do with it, why would we trust them?" Trust, transparency and benefit-sharing in Smart Farming. *NJAS-Wageningen Journal of Life Sciences*, 90, p.100285.
- Johnson, G., Shriver, S. and Goldberg, S., 2020. Privacy & market concentration: Intended & unintended consequences of the GDPR. *Available at SSRN 3477686*.
- Kalkman, S., van Delden, J., Banerjee, A., Tyl, B., Mostert, M. and van Thiel, G., 2019. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. *Journal of medical ethics*.
- Kamvar, S., Schlosser, M and Garcia-Molina, H., 2003. EigenRep: Reputation management in P2P networks.
- Layton, R., 2017. How the GDPR compares to best practices for privacy, accountability and trust. *Accountability and Trust (March 31, 2017)*.
- Lee, J.N., Huynh, M.Q. and Hirschheim, R., 2008. An integrative model of trust on IT outsourcing: Examining a bilateral perspective. *Information Systems Frontiers*, 10(2), pp.145-163.
- Liao, K., Ma, Z., Lee, J.J.Y. and Ke, K., 2011. Achieving mass customization through trust-driven information sharing: a supplier's perspective. *Management research review*.
- Linek, S.B., Fecher, B., Friesike, S. and Hebing, M., 2017. Data sharing as social dilemma: Influence of the researcher's personality. *PloS one*, 12(8), p.e0183216.
- Lippert, S.K. and Swiercz, P.M., 2007. Personal Data Collection via the Internet: The Role of Privacy Sensitivity and Technology Trust. *Journal of International Technology and Information Management*, 16(1), p.2.
- Lu, W., Xu, F. and Lv, J., 2010. An approach of software reliability evaluation in the open environment. *Chinese Journal of Computers*, 33(3), pp.452-462.
- Lu, Y., Wang, W., Bhargava, B. and Xu, D., 2006. Trust-based privacy preservation for peer-to-peer data sharing. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 36(3), pp.498-502.
- Lv, D. and Zhu, S., 2020. Achieving secure big data collection based on trust evaluation and true data discovery. *Computers & Security*, 96, p.101937.
- Mayer-Schonberger, V. and Ramge, T., 2018. A Big Choice for Big Tech: Share Data or Suffer the Consequences. *Foreign Aff.*, 97, p.48.

- McKinsey, 2013. Open data: Unlocking innovation and performance with liquid information.
- McKnight, D.H., Choudhury, V. and Kacmar, C., 2002. The impact of initial consumer trust on intentions to transact with a web site: a trust building model. *The journal of strategic information systems*, 11(3-4), pp.297-323.
- Meijer, R., Conradie, P. and Choenni, S., 2014. Reconciling contradictions of open data regarding transparency, privacy, security and trust. *Journal of theoretical and applied electronic commerce research*, 9(3), pp.32-44.
- Milne, R., Morley, K.I., Howard, H., Niemiec, E., Nicol, D., Critchley, C., Prainsack, B., Vears, D., Smith, J., Steed, C. and Bevan, P., 2019. Trust in genomic data sharing among members of the general public in the UK, USA, Canada and Australia. *Human genetics*, 138(11), pp.1237-1246.
- Moon, L.A., 2017. Factors influencing health data sharing preferences of consumers: A critical review. *Health policy and technology*, 6(2), pp.169-187.
- Mulder, T. and Broekstra, R.B., 2019. Does the GDPR Have Trust Issues?. Available at SSRN 3506807.
- Muravyeva, E., Janssen, J., Dirkx, K. and Specht, M., 2020, July. The Role of Trust in Personal Data Sharing in the Context of e-Assessment and the Moderating Effect of Special Educational Needs. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 328-332).
- Murphy, H., Keahey, L., Bennett, E., Drake, A., Brooks, S.K. and Rubin, G.J., 2020. Millennial attitudes towards sharing mobile phone location data with health agencies: a qualitative study. *Information, Communication & Society*, pp.1-14.
- Nasser, A.A.N., 2020. Impacts of Trust in Government and Privacy Risk Concern on Willingness to Provide Personal Information in Saudi Arabia. *International Journal of Management Science and Business Administration*, 6(2), pp.7-18.
- Neves, A.L., Poovendran, D., Freise, L., Ghafur, S., Flott, K., Darzi, A. and Mayer, E.K., 2019. Health Care Professionals' Perspectives on the Secondary Use of Health Records to Improve Quality and Safety of Care in England: Qualitative Study. *Journal of medical Internet research*, 21(9), p.e14135.
- Nicolaou, A.I. and McKnight, D.H., 2006. Perceived information quality in data exchanges: Effects on risk, trust, and intention to use. *Information systems research*, 17(4), pp.332-351.
- O'Hara, K., 2019. Data Trusts: Ethics, Architecture and Governance for Trustworthy Data Stewardship. Web Science Institute White Paper.
- OECD, 2015. Assessing government initiatives on public sector information. A review of the OECD Council Recommendation. *OECD Digital Economy Papers*.
- Open Data Institute and Nuffield Foundation, 2020. Valuing data: foundations for data policy.
- Open Data Institute, 2019. How we developed a model of data sharing in the economy.
- Ostherr, K., Borodina, S., Bracken, R.C., Lotterman, C., Storer, E. and Williams, B., 2017. Trust and privacy in the context of user-generated health data. *Big Data & Society*, 4(1), p.2053951717704673.
- Özer, Ö., Zheng, Y. and Chen, K.Y., 2011. Trust in forecast information sharing. *Management Science*, 57(6), pp.1111-1137.
- Patel, M.S., 2019. Merger Breakups. Available at SSRN 3469984.

- Peral, J., Gallego, E., Gil, D., Tanniru, M. and Khambekar, P., 2020. Using Visualization to Build Transparency in a Healthcare Blockchain Application. *Sustainability*, 12(17), p.6768.
- Porcedda, M.G. and Wall, D.S., 2019. Cascade and Chain Effects in Big Data Cybercrime: Lessons from the TalkTalk hack, proceedings of WACCO 2019: 1st Workshop on Attackers and Cyber-Crime Operations, IEEE EuroS&P 2019, Stockholm, Sweden, June 20, 2019.
- Przepiorka, W., 2013. Buyers pay for and sellers invest in a good reputation: More evidence from eBay. *The Journal of Socio-Economics*, 42, pp.31-42.
- Rissman, A.R., Morris, A.W., Kalinin, A., Kohl, P.A., Parker, D.P. and Selles, O., 2019. Private organizations, public data: Land trust choices about mapping conservation easements. *Land Use Policy*, 89, p.104221.
- Rumbold, J.M.M. and Pierscionek, B., 2017. The effect of the general data protection regulation on medical research. *Journal of medical Internet research*, 19(2), p.e47.
- Sanderson, S.C., Brothers, K.B., Mercaldo, N.D., Clayton, E.W., Antommaria, A.H.M., Aufox, S.A., Brilliant, M.H., Campos, D., Carrell, D.S., Connolly, J. and Conway, P., 2017. Public attitudes toward consent and data sharing in biobank research: a large multi-site experimental survey in the US. *The American Journal of Human Genetics*, 100(3), pp.414-427.
- Sandhaus, S., Kaufmann, D. and Ramirez-Andreotta, M., 2019. Public participation, trust and data sharing: gardens as hubs for citizen science and environmental health literacy efforts. *International Journal of Science Education, Part B*, 9(1), pp.54-71.
- Savona, M., 2020. The Saga of the Covid-19 Contact Tracing Apps: Lessons for Data Governance.
- Schneider, M.J., Jagpal, S., Gupta, S., Li, S. and Yu, Y., 2017. Protecting customer privacy when marketing with second-party data. *International Journal of Research in Marketing*, 34(3), pp.593-603.
- Shipman, F.M. and Marshall, C.C., 2020, April. Ownership, Privacy, and Control in the Wake of Cambridge Analytica: The Relationship between Attitudes and Awareness. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).
- Steedman, R., Kennedy, H. and Jones, R., 2020. Complex ecologies of trust in data practices and data-driven systems. *Information, Communication & Society*, pp.1-16.
- Stone, M.A., Redsell, S.A., Ling, J.T. and Hay, A.D., 2005. Sharing patient data: competing demands of privacy, trust and research in primary care. *British journal of general practice*, 55(519), pp.783-789.
- Tasman, A.C.I.L., 2009. Spatial information in the New Zealand economy. Realising productivity gains. *prepared for Land Information New Zealand*, pp.77-95.
- Urban, T., Tatang, D., Degeling, M., Holz, T. and Pohlmann, N., 2020. Measuring the Impact of the GDPR on Data Sharing in Ad Networks. In *Proceedings of the 15th ACM ASIA Conference on Computer and Communications Security, ASIA CCS* (Vol. 20).
- Walker, J., Simperl, E., Stalla-Bourdillon, S. and O'Hara, K., 2019. Decision making processes for data sharing: a framework for data trusts.
- Wallis, J.C., Rolando, E. and Borgman, C.L., 2013. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PloS one*, 8(7), p.e67332.
- Wang, Y., Cai, Z., Yin, G., Gao, Y., Tong, X. and Han, Q., 2016. A game theory-based trust measurement model for social networks. *Computational social networks*, 3(1), p.2.

- Welch, E.W., Feeney, M.K. and Park, C.H., 2016. Determinants of data sharing in US city governments. *Government Information Quarterly*, 33(3), pp.393-403.
- Williams, H., Spencer, K., Sanders, C., Lund, D., Whitley, E.A., Kaye, J. and Dixon, W.G., 2015. Dynamic consent: a possible solution to improve patient confidence and trust in how electronic patient records are used in medical research. *JMIR medical informatics*, 3(1), p.e3.
- Wiseman, L., Sanderson, J., Zhang, A. and Jakku, E., 2019. Farmers and their data: An examination of farmers' reluctance to share their data through the lens of the laws impacting smart farming. *NJAS-Wageningen Journal of Life Sciences*, 90, p.100301.
- Xiong, L. and Liu, L., 2004. Peertrust: Supporting reputation-based trust for peer-to-peer electronic communities. *IEEE transactions on Knowledge and Data Engineering*, 16(7), pp.843-857.
- Xuan, S., Zheng, L., Chung, I., Wang, W., Man, D., Du, X., Yang, W. and Guizani, M., 2020. An incentive mechanism for data sharing based on blockchain with smart contracts. *Computers & Electrical Engineering*, 83, p.106587.
- Yuan, W., Jian, L., Feng, X. and Lin, Z., 2006, April. A trust measurement and evolution model for internetware. In *Journal of software*.
- Zhang, J., Hassandoust, F. and Williams, J.E., 2020. Online Customer Trust in the Context of the General Data Protection Regulation (GDPR). *Pacific Asia Journal of the Association for Information Systems*, 12(1), p.4.
- Zhang, Y., Wang, T. and Hsu, C., 2019. The effects of voluntary GDPR adoption and the readability of privacy statements on customers' information disclosure intention and trust. *Journal of Intellectual Capital*.
- Zhao, J., Zhu, C., Peng, Z., Xu, X. and Liu, Y., 2018. User willingness toward knowledge sharing in social networks. *Sustainability*, 10(12), p.4680.

## ANNEX A ADDITIONAL DETAIL ON EVIDENCE REVIEW AND ANALYTIC STRATEGY

### A.1 Search strategy

We carried out a rigorous review of available publications through database searches. We explored some of the most comprehensive academic databases providing literature at the intersection between multiple disciplines, including economics, policy, computer science, social science and ethics/applied philosophy.

Key databases we explored include JSTOR, RePEc, EconLit, ScienceDirect and Google Scholar. Our search terms include multiple combinations of the following search terms:

- trust (or synonym) AND
- data OR
- data sharing AND
- impact (or synonym) AND
- health OR
- financial (OR banking) OR
- mobility OR
- data trust + impact

For each database/keyword combination, we have reviewed the top 40 search results (for queries returning more than 40 results) and identified the most relevant ones based on a review of the abstract. In doing so we have considered all academic papers, research and policy articles, and excluded book chapters and news articles. Amongst the research identified, we excluded all papers that:

- proposed new approaches to trustworthy data sharing without considering its impact on data sharing/collection/use;
- investigated the relationship between trust and knowledge sharing (based on our key words, the database search returned a high volume of publications about this in the management and organisation literature), as this was deemed not relevant to inform our research question.

As a result of this process, we identified a pool of 56 studies to be reviewed. A breakdown of our query results are shown in Figure 28 below.

**Figure 28 Database search results**

<b>Number of studies identified</b>	<b>Ideas Repec</b>	<b>Science Direct</b>	<b>Google Scholar</b>	<b>Jstor</b>	<b>Total</b>
<i>trust (or synonym) + data + impact (or synonym)</i>	4	4	#N/A	3	11
<i>trust (or synonym) + data sharing (or "data sharing") +impact (or synonym)</i>	11	4	14	0	29
<i>trust + data (or data sharing) + impact (or synonym) + health</i>	4	1	1	0	6
<i>trust + data (or data sharing) + impact (or synonym) + financial (or banking)</i>	1	1	2	0	4
<i>trust + data (or data sharing) + impact (or synonym) + mobility</i>	1	0	2	0	3
<i>trust + data (or data collection) + impact (or synonym) + health</i>	#N/A	#N/A	0	#N/A	0
<i>trust + data (or data use) + impact (or synonym) + health</i>	#N/A	1	1	#N/A	2
<i>data trust + impact</i>	#N/A	#N/A	1	#N/A	1
<b>Total</b>	21	11	21	3	<b>56</b>

Source: Frontier Economics research

Note: #NA means that we did not conduct a given key word search on that particular database

We have also carried out ad-hoc online searches informed by our domain knowledge and guided by ODI suggestions, which returned an additional 29 relevant studies, including studies published by government department and think tanks.

## A.2 Highly relevant papers identified

As set out in Section 1.2.1, we classified these papers into high, medium and low relevance. The criteria we used to identify highly relevant studies are the following:

- the research question is highly relevant to inform the relationship between trust, data sharing/collection/use and subsequent economic impacts.
- the methodology includes an assessment of the impact of trust on data sharing/collection/use, or the economic impact of data sharing/collection use.
- The assessment is quantitative in almost all cases. Where surveys are conducted, the sample size is large enough to draw meaningful statistical conclusions. If qualitative, the assessment is highly rigorous and is able to isolate the impact of changes in trust on data sharing.
- the type of trust linkages and types of actors under analysis are easily identifiable.
- The study was conducted post 2010.

The table below reports details for the 21 highly academic studies we classified as ‘highly relevant’, including sector, geography, methodology and research question.

**Figure 29** Highly relevant academic studies

Title	Author (year)	Geography	Sector	Methodology	Research question
Measuring the impact of security, trust and privacy in information sharing: A study on social networking sites	Gupta (2015)	NA	Social media	Quantitative survey	Understand the impact of security, trust and privacy concerns on willingness to share information on social networking sites. RQ1: What are the antecedents of trust in social networking sites? RQ2: What is the impact of privacy, security, and trust on the willingness of sharing information?
Here’s my location, for your information: The impact of trust, benefits, and social influence on location sharing application use among Indonesian university students	Beldad (2015)	Indonesia	location data	Quantitative survey	Investigating the impact of trust, benefits, and social influence on location sharing application use among Indonesian university students.
Trust and cooperative behavior: Evidence from the realm of data-sharing	Bauer (2019)	Germany	social media, academic research	Quantitative survey	Investigate the relationship between trust and cooperation (measured by data sharing)
Public Attitudes toward Consent and Data Sharing in Biobank Research: A Large Multi-site Experimental Survey in the US	Sanderson (2017)	US	Health	Quantitative survey	Assess willingness to participate in a biobank using different consent and data sharing models.

Title	Author (year)	Geography	Sector	Methodology	Research question
The trust gap: Social perceptions of privacy data for energy services in the United Kingdom	Grunewald and Reisch (2020)	UK	energy	Quantitative survey	Investigate attitudes towards location data sharing in the context of energy smart solutions (smart meters).
Consumer propensity to adopt PSD2 services: trust for sale?	Bijlsma et al (2020)	Netherlands	banking	Quantitative survey	Study consumers' attitudes towards sharing payments data with incumbent and new providers of payment and account information services, and using their services, to understand the impact of the revised Payment Services Directive (PSD2) on the functioning of the retail payments market.
Impacts of Trust in Government and Privacy Risk Concern on Willingness to Provide Personal Information in Saudi Arabia	Nasser (2020)	Saudi Arabia		Quantitative	Estimate the impacts of trust in government and privacy risk concern on willingness to provide personal information.
Building trust and sharing value: the twin challenges of health and care data	Mac Manus, (2019)	NA	health	Qualitative review	Review of several events involving breaches of trust or the relationship between trust and data sharing more generally, in the context of healthcare data.
Trust and privacy in the context of user-generated health data	Ostherr (2017)	US	Health	Qualitative, semi-structured interviews	Analysing concern about sharing health data with the companies that sold the devices or apps they used.
Reconciling Contradictions of Open Data Regarding Transparency, Privacy, Security and Trust	Meijer et al (2014)	Netherlands	government, research	Theoretical model and case study application	Analysing pre-commitment as a policy instrument whereby an organization imposes some restraints on its policy in order to restrict the extent to which values may conflict and the degree to which stakeholders should be concerned about the trustworthiness of that policy.
Private organizations, public data: Land trust choices about mapping conservation easements	Rissman et al (2019)	US	Land conservation	Quantitative survey	Focus on decisions by land conservation NGOs (land trusts) to share digital maps of conservation easements on private lands. Investigation of which land trusts were more likely to contribute digital maps to public databases, and what benefits and concerns with disclosure did land trust staff report.
Farmers and their data: An examination of farmers' reluctance to share their data through the lens of the laws impacting smart farming	Wiseman (2019)	Australia	Agriculture	Quantitative survey	Examining the attitudes of farmers to the collection, control, sharing and use of their farm data.

Title	Author (year)	Geography	Sector	Methodology	Research question
Sustainable Regulation of Information Sharing with Electronic Data Interchange by a Trust-Embedded Contract	Han and Dong (2017)	NA	Private sector	Theoretical model + experiments	Studying the risks in demand information sharing applications by electronic soft-orders using electronic data interchange (EDI) systems in e-commerce and aims to suggest a sustainable regulation mechanism with a trust-embedded contract.
Achieving mass customization through trust-driven information sharing: a supplier's perspective	Kun Liao et al (2011)	North America	private sector	Quantitative Survey	Testing empirically the relationship between trust and manufacturer-supplier information sharing to determine its (positive) effect on mass customization.
Determinants of data sharing in U.S. city governments	Welch (2016)	US	government	Qualitative review	Testing hypotheses predicting sharing behaviour of municipal government agencies with other agencies and with non-government organizations
Ownership, Privacy, and Control in the Wake of Cambridge Analytica: The Relationship between Attitudes and Awareness	Shipman (2020)	US	Social media	Quantitative survey	Investigating whether widespread news of abuse changed the public's perceptions of how user-contributed content from social networking sites like Facebook and LinkedIn can be used
Should I stay or should I leave? exploring the (dis)continued Facebook use after the Cambridge Analytica scandal	Brown (2020)	US	Social media	Qualitative semi-structured interviews	Analysing decisions to stay with or leave Facebook following the Cambridge Analytica case as such decisions intersect with privacy concerns.
Privacy & Market Concentration: Intended & Unintended Consequences of the GDPR	Johnson et al (2020)	EU	online websites/web technologies	Quantitative	Estimate the impact of GDPR enforcement on data sharing, by examining website choices of web technology vendors (platforms like FB and Google) in response to the European Union (EU) enforcing the GDPR.
Regulating Privacy Online: The Early Impact of the GDPR on European Web Traffic & E-Commerce Outcomes	Goldberg et al (2019)	EU	online traffic	Quantitative	Examining the impact of the GDPR on European web traffic and e-commerce sales.
Measuring the Impact of the GDPR on Data Sharing in Ad Networks	Urban et al (2020)	EU	online traffic	Quantitative	Analysing the underlying information sharing networks between online advertising companies in terms of client-side cookie syncing.

Title	Author (year)	Geography	Sector	Methodology	Research question
The effects of voluntary GDPR adoption and the readability of privacy statements on customers' information disclosure intention and trust	Zhang (2019)	EU		Quantitative	Examining the impacts of companies' voluntary adoption of the General Data Protection Regulation (GDPR) as well as the readability of privacy statements on US customers' intention to disclose information and their trust in a company.

Source: *Frontier Economics*

## A.3 Estimates calibration

As noted in Section 3.2.2, eight studies amongst those we have reviewed provide a direct quantified estimate of the impact of trust on data sharing, of which seven studies were directly comparable and therefore formed the core of our evidence base.

To make estimates comparable and draw aggregate impacts, we standardise each coefficient to align with the following relationship: a *1 point increase on a 5 point trust Likert scale leads to a X point increase on a 5 point data sharing Likert scale*.

The table below reports the details for each study which we used to complete the standardisation of estimates. As displayed in the table, most studies measure trust and data sharing on a 5 point Likert scale, therefore, the coefficient estimate will be of the type 'categorical variable (x) on categorical variable (y)'. These estimates are therefore already fully comparable. To bring them in line with the others, we have converted the 'categorical variable (x) on binary variable (y)' estimates from the Bauer (2019) and the Bijlsma (2020) studies by applying simple proportions.

**Figure 30 Estimates standardisation**

Paper	Estimation technique	Trust measurement	Data sharing measurement	Coefficient type	Baseline trust level	Impact of trust on data sharing (standardized)	Sample size	Internal validity score
<b>Nasser (2020)</b>	regression	5 point likert scale	5 point likert scale	categorical on categorical	NA	0.59	268	1
<b>Wiseman (2019)</b>	regression	5 point likert scale	5 point likert scale	categorical on categorical	2.41	0.33	1000	1
<b>Gupta (2015)</b>	regression	5 point likert scale	5 point likert scale	categorical on categorical	NA	0.25	250	4
<b>Liao (2011)</b>	pearson correlation	5 point likert scale	5 point likert scale	categorical on categorical	NA	0.24	208	3
<b>Bauer (2019)</b>	regression	3 point scale	binary	categorical on binary	3	0.21	2095	2
<b>Beldad (2015)</b>	regression	5 point likert scale	5 point likert scale	categorical on categorical	3.54	0.14	655	2
<b>Bijlsma (2020)</b>	regression	5 point likert scale	binary	categorical on binary	NA	0.11	2678	5

Source: *Frontier Economics*

Note: \*pps indicates percentage points

Almost all of the studies considered estimate the relationship between trust and data sharing through a regression coefficient. Instead, Liao (2011) estimates the relationship between trust and data sharing as a Pearson correlation coefficient, i.e. a coefficient bounded between -1 and 1, which gives an indication of how trust (x variable) and data sharing (y variable) move together, but is unable to measure the impact of a change in trust on data sharing. However, the methodology used in this study makes the correlation estimate comparable to the other regression coefficients.<sup>63</sup> In the table below we provide guidance on how to interpret econometric regression coefficients.

---

<sup>63</sup> To convert a correlation coefficient into a regression coefficient, one needs to know the variances of x and y. Since the Liao study measures both trust and data sharing on the same scale, it is plausible that their variances are equal, in which case the regression and correlation coefficients are equal.

## INTERPRETING REGRESSION RESULTS

To aid the interpretation of the coefficients resulting from the regression estimation techniques reported above, consider the following simplified example.

The below is a *linear probability model* estimating the impact of the independent variable (or regressor) of interest  $x_1$  (in our case, trust) on a dependent variable  $y$  (in our case, data sharing), where  $e$  is an error term capturing any residual 'noise' in the estimation.

$$y = a + bx_1 + e$$

In this case, the coefficient of the impact of trust on data sharing  $b$  will capture the impact of a 1 unit increase in trust on data sharing. However, this coefficient does not necessarily capture the *causal* impact of trust on data sharing.

The reason why this might be the case is that there are likely factors influencing data sharing other than trust which, if not taken into account as additional regressors, are likely to cause a bias in the impact estimate  $b$ . To mitigate for this, many of the studies reviewed estimated linear probability model regressions which also included other determinants of data sharing (or *control variables*). For example, by adding a control for the implementation of another determinant of data sharing  $x_2$  to the regression:

$$y = a + bx_1 + cx_2 + e$$

might make the estimation of  $b$  more precise ( $b$  will now capture the impact of  $x_1$  as separated from the impact of  $x_2$ ). This means that, in these cases, the impact estimate  $b$  approximates the true impact of trust more closely than those models without other controls (and for this reason we have considered studies with controls as more robust).

Nonetheless, linear probability models are still unlikely to capture the *causal* impact of trust on data sharing.

Another reason why a linear probability model might not capture the true impact of trust on data sharing is what is referred to as *reverse causality* in applied statistics and econometrics. In the case at hand, we expect that:

- In most cases, economic theory suggests that trust ( $x_1$ ) is a driver of impact for data sharing ( $y$ ) - if trust is increased, then this will lead to more data sharing, i.e. *causality*;
- It might also be the case that greater data sharing ( $y$ ) allows more trust to begin with ( $x_1$ ) – to some extent, data sharing could be a driver of trust, i.e. *reverse causality*.

Reverse causality cannot be fully accounted for by linear probability models. Only more sophisticated techniques exploiting exogenous and unanticipated shocks in trust ( $x_1$ , the independent variable) are able to capture the causal impact of trust on data sharing. There is extensive econometric literature employing these more sophisticated methods (for example, randomized control trials, difference-in-differences designs, regression discontinuity designs, or the use of instrumental variables). Some of the academic studies included in this review employ more sophisticated techniques and are better able to approximate the causal impact of trust - we have therefore assigned a higher robustness (or internal validity) score. In addition, as outlined in Section 3.4, by analysing a set of natural experiments (i.e. examples of breaches in trust like Cambridge Analytica and of increases in trust like the GDPR) we were able, although mostly based on ad hoc statistics and anecdotal evidence rather than on robust econometric studies, to test the direction of causality. That is, we have assessed whether changes in trust result in changes in data sharing behaviour over time.

