Deloitte.

Realising the economic potential of machine-generated, non-personal data in the EU

Report for Vodafone Group

July 2018

Contents

Important Notice from Deloitte

This final report (the "Final Report") has been prepared by Deloitte LLP ("Deloitte") for Vodafone Group Services Limited in accordance with the contract with them dated 21 November 2017 ("the Contract") and on the basis of the scope and limitations set out below.

The Final Report has been prepared solely for the purposes of contributing to the evidence base needed for the design of the new coordinated and pan-European approach to promoting the sharing of machine generated, non-personal data among European policymakers, as set out in the Contract. It should not be used for any other purpose or in any other context, and Deloitte accepts no responsibility for its use in either regard including its use by Vodafone Group Services Limited for decision making or reporting to third parties.

The Final Report is provided exclusively for Vodafone Group Services Limited's use under the terms of the Contract. No party other than Vodafone Group Services Limited is entitled to rely on the Final Report for any purpose whatsoever and Deloitte accepts no responsibility or liability or duty of care to any party other than Vodafone Group Services Limited in respect of the Final Report or any of its contents.

The information contained in the Final Report has been obtained from third party sources that are clearly referenced in the appropriate sections of the Final Report. Deloitte has neither sought to corroborate this information nor to review its overall reasonableness. Further, any results from the analysis contained in the Final Report are reliant on the information available at the time of writing the Final Report and should not be relied upon in subsequent periods.

Any decision to invest, conduct business, enter or exit the markets considered in the Final Report should be made solely on independent advice and no information in the Final Report should be relied upon in any way by any third party. This Final Report and its contents do not constitute financial or other professional advice, and specific advice should be sought about your specific circumstances. In particular, the Final Report does not constitute a recommendation or endorsement by Deloitte to invest or participate in, exit, or otherwise use any of the markets or companies referred to in it. To the fullest extent possible, both Deloitte and Vodafone Group Services Limited disclaim any liability arising out of the use (or non-use) of the Final Report and its contents, including any action or decision taken as a result of such use (or non-use).

Glossary

Term	Definition
Anonymised data	Information where personal data is rendered anonymous in such a manner that the data subject is not or no longer identifiable (with no key link back to a natural living person).
Application Programming Interface (API)	A set of functions and procedures that allow the creation of applications that access the features or data of an operating system, application, or other service.
Closed data	Data that can only be accessed by its subject, owner or holder – whether that is an individual person or single organisation.
Data	Qualitative or quantitative statements or numbers that are assumed to be factual and not the product of analysis or interpretation. Data can be structured (i.e. organised in a recognisable manner) or unstructured.
Data Infrastructure	An umbrella term to capture the technologies, processes, guidance and organisations involved in using and managing data, as well as data itself.
General Data Protection Regulation (GDPR)	EU regulation 2016/679 on the handling of personal data in the EU by public and private sector organisations.
Global Positioning System (GPS)	A global navigation satellite system that provides location and time specific information to a GPS receiver anywhere on Earth.
Internet of Things (IoT)	Describes the interconnection of a network of physical objects, which are embedded with unique identifiers and software that allows the collection and exchange of data.
Machine-generated data	Data recorded, collected or produced by connected devices, assets or networks independent of any human intervention.
Machine Learning	An artificial intelligence capability where computer systems improve their performance by exposure to data without the need to follow explicit instructions.
Mosaic effect	The process of combining anonymised data with auxiliary data in order to reconstruct identifiers linking data to the individual it relates to.
Non-personal data	Information that does not relate to an identified or identifiable natural person. Data which does not allow for the identification of a living natural person either directly or by reference to an identifier such as a name, identification number, location data, an online identifier or one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity.

	Non-personal data includes both data that is non-personal when generated and data that is personal at first and then anonymised.
Open data	Data that anyone can access, use and share. For data to be considered 'open', it must be: accessible, which usually means published on the web; available in a machine-readable format; and have a licence that permits anyone to access, use and share it - commercially and non-commercially.
Personal data	Information that relates to an identified or identifiable natural person. Data which allows for the identification of a living natural person either directly or by reference to an identifier such as a name, identification number, location data, an online identifier or one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.
Pseudoynmised data	Data that has been processed in such a manner that the personal data can no longer be attributed to a specific subject without the use of additional information. For data to be treated as pseudonymous, that additional information must be kept separately and subject to technical and organisational measures to ensure that the personal data is not attributed to a natural person.
Real-time, archive or reference data	Data that describes events as they happen is real-time or near-real-time. The same data collected over time is archive data. Data that describes something at only a single point in time is reference data.
Re-use	Use of data other than for the purpose it was originally produced or collected for. This could be for commercial or non-commercial purposes.
Security-by-design	A system that has been designed bottom-up to be secure.
Service-critical and safety-critical data	Data that is necessary to achieve the correct operation of services can be described as service-critical, an extension of this is where data is necessary to ensure the wellbeing of individuals which can be described as safety-critical.
Shared data	Shared data is data that a specific person or organisation, who are not the data subject, owner or holder can access, use and re-use. Whenever data crosses organisational boundaries, that data is defined as being shared. There are three main types of data sharing identified in this report: • Horizontal data sharing: sharing that occurs between organisations involved in the same commercial or non-commercial point of the value chain (e.g. businesses selling the same product in the same market place) or for use in benchmarking (e.g. comparing service levels between municipalities). • Vertical data sharing: sharing that occurs between organisations who have a customer or supplier relationship, directly or indirectly, e.g. to build and improve services or quality based on user-level feedback.

- External data sharing: sharing that occurs with organisations outside the sector, engaged in different commercial or non-commercial activities, e.g. sharing with third parties to analyse and buy-back aggregated datasets that have been contributed to by multiple parties.

Foreword

The rise of the machines and the €1.4trn prize

The modern internet at the centre of daily life for billions of people first emerged around 20 years ago. Over the years since, the personal data of each individual internet user has become the world's most valuable commodity, exploited to great - but, at times, perhaps also terrifying - effect.

As a result, policymakers worldwide are – rightly – beginning to focus on ensuring that personal data is used responsibly. The outcomes are people-centric rules designed to govern interactions between people and involving personal information. In Europe, the GDPR has just entered into effect – a wide-ranging and comprehensive approach to protecting the citizen's personal data, and which is likely to become a new global benchmark in future.

But there is a wholly new kind of internet now emerging where the interactions are purely between machines. People control the processes, and they benefit from the outcomes – but they are not the intermediaries.

The so-called 'Internet of Things' (IoT) raises important policy questions. For example, should intelligent networked devices using licensed cellular networks to communicate with each other be regulated using the same rules applied when consumers post a photo on social media via a smartphone? What rules should apply when an IoT device crosses international borders? Some of the answers may seem obvious; others, perhaps less so.

One aspect that has not been sufficiently understood until now is the scale of the economic opportunity to be achieved by reaching the right answers in response to questions such as these. The study by Deloitte – commissioned by Vodafone, and set out on the following pages – provides a vivid illustration of the potential.

Policies designed to enable the seamless transfer of non-personal machine-generated data between devices and different actors would have a dramatically positive impact on the European economy. Deloitte's modelling, which is based on conservative assumptions, indicates a potential annual addition of around 1.4 trillion EUR to European Union GDP in 2027. This is equivalent to the current GDP of Spain – the EU's fifth largest economy – and would exceed, in one year, the Union's total budget for the period 2014-2020. The prize is immense.

Yet, as this report also shows, there are many reasons why the opportunities to be derived from a more open data-sharing regime may not be realised. Existing legislation at Member State level and within current EU frameworks may be difficult to amend or retrofit without fear of creating unintended consequences in other parts of the digital infrastructure and services markets. Companies may be reluctant to allow non-personal machine-generated data to be shared with third-parties, either for commercial or intellectual property reasons or out of concern that doing so would create cyber security vulnerabilities.

However, these factors are not insurmountable. The Deloitte study examines a number of examples of effective data-sharing in practice and sets out compelling arguments that the benefits outweigh the risks. Moreover, the analysis demonstrates that data-sharing on a voluntary and contractual, rather than mandated, basis is more likely to be effective; just as the 'Internet of People' grew through openness and collaboration, so too will the 'Internet of Things'.

One clearly bad outcome would be for policymakers and regulators to do nothing. By sitting on its hands, Europe will not be able to tap into the potential of IoT and machine-generated data. It is not much of a forward looking strategy to assume that existing rules designed for the age of the smartphone (at best: some are much older than that) will somehow morph into an effective regulatory regime for an estimated 18 billion IoT devices

worldwide – with more than 50% of business processes using IoT sensing or control systems–– by 2022. Neither would Europe be well-served by a new set of regulations whose starting point was that this second internet revolution should somehow be constrained within national borders, and on a detailed ex ante basis.

There is a lot of discussion ahead, and a lot at stake. We hope this study will inform the debate.

Joakim Reiter
Vodafone Group External Affairs Director

Executive Summary

Vodafone Group has commissioned Deloitte to independently consider the economic potential of machine-generated, non-personal data in the European Union (EU). This has covered estimating the potential value of this form of data; how this value is derived from sharing; identifying barriers to that value being realised; and making policy recommendations to ease those barriers. For the purposes of this report, machine-generated, non-personal data is defined as data recorded, collected or produced by a connected device, network or asset without any direct human intervention and which, at the time of it being created, does not identify a living natural person, either directly or in combination with other data.

While personal data has been the focus of EU and Member State policy thus far, as reflected in a number of initiatives being enacted such as GDPR, the policy considerations relating to non-personal data generated by Internet of Things (IoT) devices are quite different. As the European Commission (Commission) turns its attention to enacting new policy in the area of non-personal data, the intention of this report is to contribute to new evidence and new ideas to support that process.[1] Indeed, without the same implications for privacy as in the case of personal data, the policy focus for non-personal data is on realising the economic and societal benefits of using and re-using non-personal data, whilst addressing legitimate commercial and security concerns.

Machine-generated, non-personal data is generated through an enormous array of connected devices. These devices, which may be geographically dispersed, collect and record many different types of data via their sensors. Individual data feeds can then be aggregated into much larger and more powerful datasets. The analysis in this report, building on the existing literature and expert engagement, suggests that across five major sectors of the European economy the potential from using and re-using this form of data is significant. Over the next decade, the sharing of machine-generated, non-personal data in European economies can, among other impacts:

- Create €35 billion in value in agriculture by raising yields.
- Reduce costs from road vehicle damage, maintenance and repairs by €40 billion.
- Generate efficiencies in resource management and prevent drug counterfeiting in the healthcare sector, saving €14 billion. Other uses combining with personal data can add further value.
- Save €2 billion in Smart Cities by improving the energy efficiency of street lighting (plus a range of other uses in Smart Cities).
- Create €1,300 billion of value in manufacturing by improving productivity.

Specific use cases cited in this report include using sensors and remote sensing in crops and livestock to enable the optimisation of agricultural processes; analysing asset condition data to undertake predictive maintenance for equipment in the healthcare, automotive and manufacturing sectors; and calibrating smart energy systems based on demand peaks to improve urban energy efficiency.

[1] On 25 April 2018, the Commission published a set of measures to increase the availability of data in the EU building on its framework for the free flow of non-personal data set out in September 2017.

These benefits depend, in large part, on the successful sharing of data between organisations. Sharing is vital as there may be a wide range of potential uses for the data once it has been transmitted from the connected device and many of those uses may not take place within the organisation that holds the data or may require different skills. Data can be shared with customers and suppliers throughout the supply chain (vertical data sharing); with peers to create larger data sets including more variation (horizontal data sharing); and outside the sector with stakeholders who can make use of the data in new settings (external data sharing).

However, the analysis in this report has identified a number of obstacles to data sharing which could prevent the full value of machine-generated, non-personal data being realised across the EU. The most important obstacles are:

- Commercial barriers to sharing – the concern that firms will have a limited incentive to share data when they might not appropriate the gains themselves.
- Security barriers to sharing – the concern that sharing will create vulnerabilities to intentional or inadvertent security breaches.
- Legal barriers to sharing – concerns, particularly in the healthcare sector, that a range of existing rules may not allow data sharing. This is particularly salient to the extent it is challenging in practice to distinguish between personal and non-personal data.

Other concerns highlighted by experts and the literature include contractual restrictions on sharing, technical limitations to sharing and cultural obstacles that disincentivise sharing. However, these obstacles were found to be both less important and more likely to be addressed through market-led initiatives without policy intervention by regulators and other authorities.

The Commission has a policy programme that seeks to address many important barriers to non-personal data sharing. With respect to non-personal data, its recent measures seek to improve access to and the reusability of public sector data and to provide guidance for businesses operating in the EU on the legal and technical principles that should govern data sharing collaboration in the private sector. However, this report recommends additional measures for consideration, particularly:

- The development of clear regulatory principles for the circumstances in which data sharing should be encouraged.
- The development of targeted policy measures that can facilitate data sharing and reflect those principles (particularly accreditation to ease sharing between organisations without an established commercial relationship and regulatory guidance).
- The promotion of appropriate data sharing models across sectors, e.g. the extended vehicle and neutral server concepts that are being developed in the automotive sector could be used in other sectors with OEM manufacturers (e.g. agricultural equipment).

The impact of investments being made in collecting data and new analytical tools will be maximised when data is shared across national and organisational boundaries to find the most valuable uses. Targeted interventions from the Commission and other policymakers can support the role of data sharing in enabling European companies to compete effectively in global markets, which are reacting to rapid technological change.

1 Introduction

The proliferation of connected devices and the data they generate holds the promise of significant economic and social benefits for European citizens. The Commission seeks to maximise these benefits by ensuring data flows freely across borders and sectors and has proposed a framework for non-personal data. This report contributes to the emerging evidence base on machine-generated, non-personal data and makes recommendations on how to ensure the expected benefits are realised across five economic sectors: agriculture, automotive, healthcare, manufacturing and Smart Cities.

1.1 Context of this report

The number of connected devices has risen rapidly in recent years. According to research by Gartner, globally there were over 6.3 billion devices connected to the Internet at the start of 2017, with the number predicted to rise to 20.8 billion by 2020.[2] The additional devices include consumer devices, industrial devices and vertical-specific business applications. The increase primarily results from the growth of the Internet of Things (IoT).

Those personal, industrial and business devices will be able to collect and generate data. Some forecasts suggest that they will be responsible for a tenth of the world's information in 2020, which equates to 44 zettabytes (or 44 trillion gigabytes).[3] Industrial and business IoT devices alone are expected to account for around a third of all IoT devices by 2020. Investment associated with industrial IoT devices is expected to grow from $1.4 trillion in 2016 to nearly $3 trillion in 2020. A significant proportion of this will take place within the EU.

The sharing and re-use of IoT data can stimulate innovation, reduce barriers to entry and expansion and facilitate new business models that can address perennial social and economic concerns. In this way, the free flow of data is a pre-requisite for a competitive economy. Policymakers need to address the resulting imperative to remove obstacles to data sharing while addressing legitimate concerns around the implications for security and competition.

With the implementation of the General Data Protection Regulation (GDPR) across European economies, the focus has been on the conditions around how personal data generated by consumers and consumer IoT devices can be used and re-used in a way that respects and protects privacy. However, as more and more businesses and public agencies implement IoT solutions that generate non-personal data, it will become increasingly important to provide clarity and guidance on how this type of data can be shared safely and efficiently. This is important in order to ensure machine-generated, non-personal data does not stay in organisational siloes by default. However, questions around technical standards and commercial risks of sharing non-personal data could act as strong barriers to data sharing within a sector, across supply chains and with other third parties. Policymakers will want to overcome those barriers in a way that does not inadvertently contribute to growing cybercrime threats or permit collusion.

[2] Quoted in: https://www.sam-solutions.com/blog/how-much-data-will-iot-create-2017/ accessed 21 January 2018.
[3] Source: https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm, accessed 7 February 2018.

1.2 Purpose and scope of this report

It is important that any new policy framework reflects rapid changes in the marketplace, technologies and consumer and business attitudes. This report, commissioned by Vodafone Group and undertaken by Deloitte, seeks to develop an evidence base to inform the ongoing development of that policy framework by the Commission and other policymakers. It also sets out new ideas for the principles that can underpin data sharing and how those principles can be reflected in policy.

Vodafone Group is actively involved in building the European data economy and the Digital Single Market. Its IoT Barometer 2017/18, which captures how enterprises are using IoT technologies globally, suggests there are sizeable potential benefits to the economy in the next five years. Vodafone has established a platform for IoT devices that includes a consumer offer (with an IoT marketplace and simple billing) and an enterprise-level offering.[4] It therefore has a stake in the new coordinated and pan-European approach to the development of policy in relation to non-personal data.

This report adds to the existing literature by quantifying the benefits of sharing machine-generated, non-personal data. Much of the analysis to date has been focused on personal data, and there is limited literature on the value of non-personal data sharing. This report is one of the first to measure the potential benefits of sharing non-personal data specific to the EU, and provides a framework for considering the barriers to data sharing within industries, through the supply chain and into adjacent industries. These barriers, while similar to those hindering personal data sharing, vary in their scope and scale. The analysis conducted allows for the identification of the main obstacles in different sectors and an understanding of where the biggest gains can be made within sectors. This report adds to the EU debate by proposing policy recommendations for how these obstacles might be overcome.

The report focuses on five key sectors of the European economy, each of which will be involved in the Digital Single Market.

These sectors have been chosen due to their size and contribution to the European economy; the amount of connected devices present; and the expected volume of machine-generated, non-personal data being generated.

1.3 Acknowledgements

Deloitte would like to thank the expert respondents surveyed and interviewed for their time and insight. The over 50 experts consulted have over 900 years of collective experience in their respective sectors; were split evenly across the five sectors above; and covered 11 EU Member States. Their responses, which were used to test and extend the existing literature, underpin the quantitative analysis in this report and have informed the qualitative analysis.

1.4 Report limitations

It should be noted that there is an inherent uncertainty around how emerging IoT technology will develop, which means the potential costs and benefits are also uncertain. With IoT being deployed in new technical and institutional settings, and this being subject to any novel policy development, the estimates provided should be treated as indicative and this report does not seek to quantify the importance of specific elements or impacts within sectors.

[4] Source: http://www.vodafone.com/business/iot/iotbarometer

The quantitative estimates presented in this report are based on assumptions derived from expert opinion testing and extending the findings of a literature review. While this embeds the best available understanding now of the likely development of IoT devices it necessarily results in a degree of subjectivity. Revised assumptions based on new observed data may lead the estimates to be revised upwards or downwards.

1.5 Structure of this document

This report is structured around the following sections:

- Chapter 2 provides more detail on what is meant by machine-generated, non-personal data and how it arises.
- Chapter 3 considers the importance of data sharing in generating benefits, the types of data sharing and how it can occur.
- Chapter 4 explores the benefits of the data in terms of productivity, cost savings, innovation and social impacts. It does this via a number of sectoral examples and model estimates.
- Chapter 4 explores how these benefits arise and the importance of different types of sharing or portability.
- Chapter 5 outlines the obstacles to data sharing that may cause the European economy to lose out on significant benefits, based on the analysis undertaken in this report.
- Chapter 6 presents potential policy recommendations to improve sharing in order to maximise the benefits to European economies from machine-generated, non-personal data.

2 Machine-generated, non-personal data

Machine-generated, non-personal data is generated across a well-defined value chain. The extent to which it can be used or re-used is determined by the degree of availability to third parties (i.e. its openness).

2.1 Machine-generated, non-personal data

Machine-generated, non-personal data is characterised by two distinct elements.

- It is machine-generated data, i.e. the data that is recorded, collected or produced by a connected device, network or asset independent of any direct human intervention.
- It is non-personal data at the time of collection.

2.1.1 Examples

Figure 2-1 lists some examples of machine-generated, non-personal data. Later in this report, more specific examples of the types of data that are generated are reported (e.g. in Figure 2-5).

Figure 2-1: Examples of machine-generated, non-personal data

Example	1. Vehicle traffic data	2. Population flows in a city area data	3. Asset condition data	4. Inventory stock data
How data is generated	Smart sensors on roads	Smart street lighting	Sensors and monitoring devices in manufacturing	Sensors, tags or barcode scanners for stock control
Nature of usage	To monitor flows into and out of areas	To facilitate smart lighting that is dependent on the presence of people, weather conditions, etc.	To monitor the quality and condition of production assets	To monitor stock levels and quality
Why it is non-personal	Data refers to aggregate flows, not individual vehicles	Data refers to aggregate movements not individuals	Cannot be used to identify natural person	Cannot be used to identify natural person
Can it be shared beyond the data holder?	Yes	Yes	Yes, subject to Intellectual Property Right (IPR) limitations	Yes, subject to competition law
Potential insights for users and re-uses	Traffic management schemes	Public safety schemes, energy management	Can lead to predictive maintenance	Can be used to identify and prevent counterfeiting and improve stock control

Source: Deloitte analysis

As Figure 2-1 shows, non-personal data can be generated by a range of devices. Devices may be involved in manufacturing processes or parts of street infrastructure and others may specifically measure asset conditions or

passenger flows. Other examples include organisational, distributional, safety, location, emissions and network level data. The common feature in all of these is that data is generated and collected independent of direct human intervention and is aggregated, measured or stored in a way that means it cannot be used to identify individuals. Importantly, as will be discussed in subsequent chapters, there is nothing intrinsic that prevents this data from being shared outside of the organisation that collected it.

Machines, sometimes the same machines that collect non-personal data, can also generate personal data such as user location, health conditions or spending patterns. This is outside the scope of this report. Equally, non-personal data can be generated by human endeavour, e.g. surveys or other non-automated data collection activities. This is also outside the scope of this report.

2.1.2 Operational challenges in classifying and separating personal and non-personal data

This report's definition of machine-generated, non-personal data encompasses a broad and growing range of datasets generated in a wide range of business settings. While the conceptual standard for personal data is clear, there is uncertainty in practice where even expert audiences were unclear or inconsistent in their view of how data is likely to be treated under GDPR.

In practice, almost any personal data can be anonymised to constitute non-personal data (although this might preclude some use cases) and almost any non-personal data can be used to identify individuals with sufficient analytical effort and in combination with other data sets (although this only means it should be treated as personal if the combination is sufficiently likely). Data can also be pseudonymised, i.e. maintaining a key to allow individuals to be identified, which means the data is treated as personal but diminishes the risk and allows some use cases not possible with anonymisation.

Any complexity in establishing whether certain data types are personal or non-personal could lead to inconsistencies between businesses. Companies may have an internal understanding of how different data types are likely to be treated, but inconsistencies could lead to data not being treated as both sides of a data sharing transaction expect. In the context of data sharing, that might expose the sharing or the recipient organisation to legal risk and/or deter sharing.

The GDPR can also affect the use of non-personal data in practice because the two categories are not always used separately. Non-personal data might be more valuable to the extent that it can be combined with personal data for holistic analysis of systems including natural persons. This is particularly the case in healthcare where productivity depends on the alignment of staff and patients as well as machines, beds and other physical assets. This implies that the legal distinction between personal and non-personal data will not always be reflected in a clear operational boundary when the data is being used. In many valuable use cases non-personal data will be combined with personal data to understand how people interact with machines, supplies and other assets. The use of the personal data involved will have to be GDPR-compliant and the non-personal data may then need to be treated as personal, because of the processing involved.

2.2 Characteristics of machine-generated, non-personal data

It is useful to consider the characteristics of machine-generated, non-personal data as some forms are more usable than others or generate more value than others.

Figure 2-2: Machine-generated, non-personal data characteristics

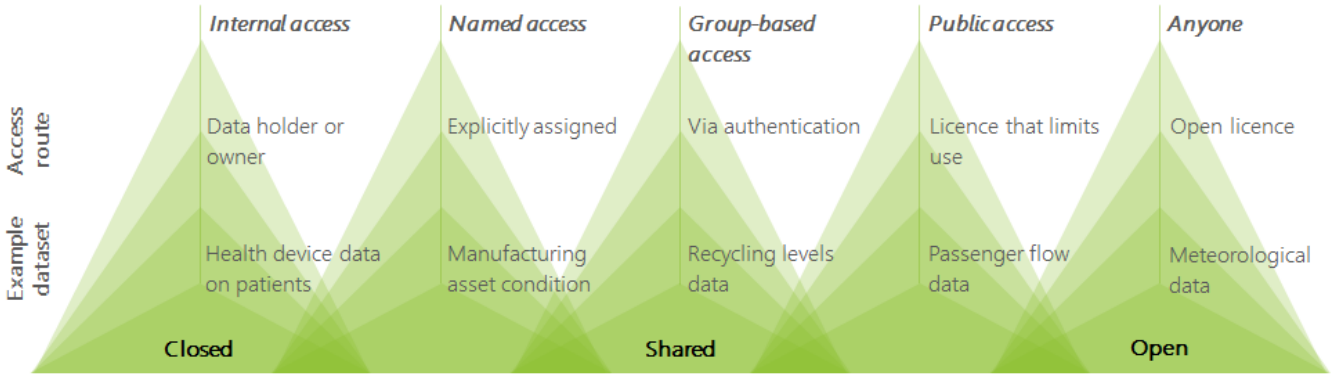
Data characteristic	Description	Variants
Content	The subject covered by the dataset in question. For example, does it relate to an individual asset, aggregated movement data, geospatial, movement and environmental data, performance data, usage data, etc.	NA

The Four V's[5]	The data's *volume* (size), *velocity* (speed of refresh); *variety* (the format it comes in) and its *veracity* (its quality and accuracy).	Volume – is it big data or not? Velocity – continuous, scheduled, intermittent or one-off updates? Variety – file format of data and whether it is machine readable? Veracity – what are its limitations, is metadata available, etc.?
Ownership	Who owns the data and has liability for it?	Commons licence or not?
Ability to share	Whether the data is open (i.e. available to be used and re-used without restriction and cost) or whether it is closed or only shared under certain conditions.	Closed, shared or open data?

Source: Deloitte analysis

It is also useful to consider the spectrum on which data can be classified in terms of its openness: the Open Data Institute maps different datasets against this spectrum and a customised version for machine-generated data is shown below.

Figure 2-3: Data spectrum for machine-generated data[6]



Source: Deloitte analysis based on ODI data spectrum

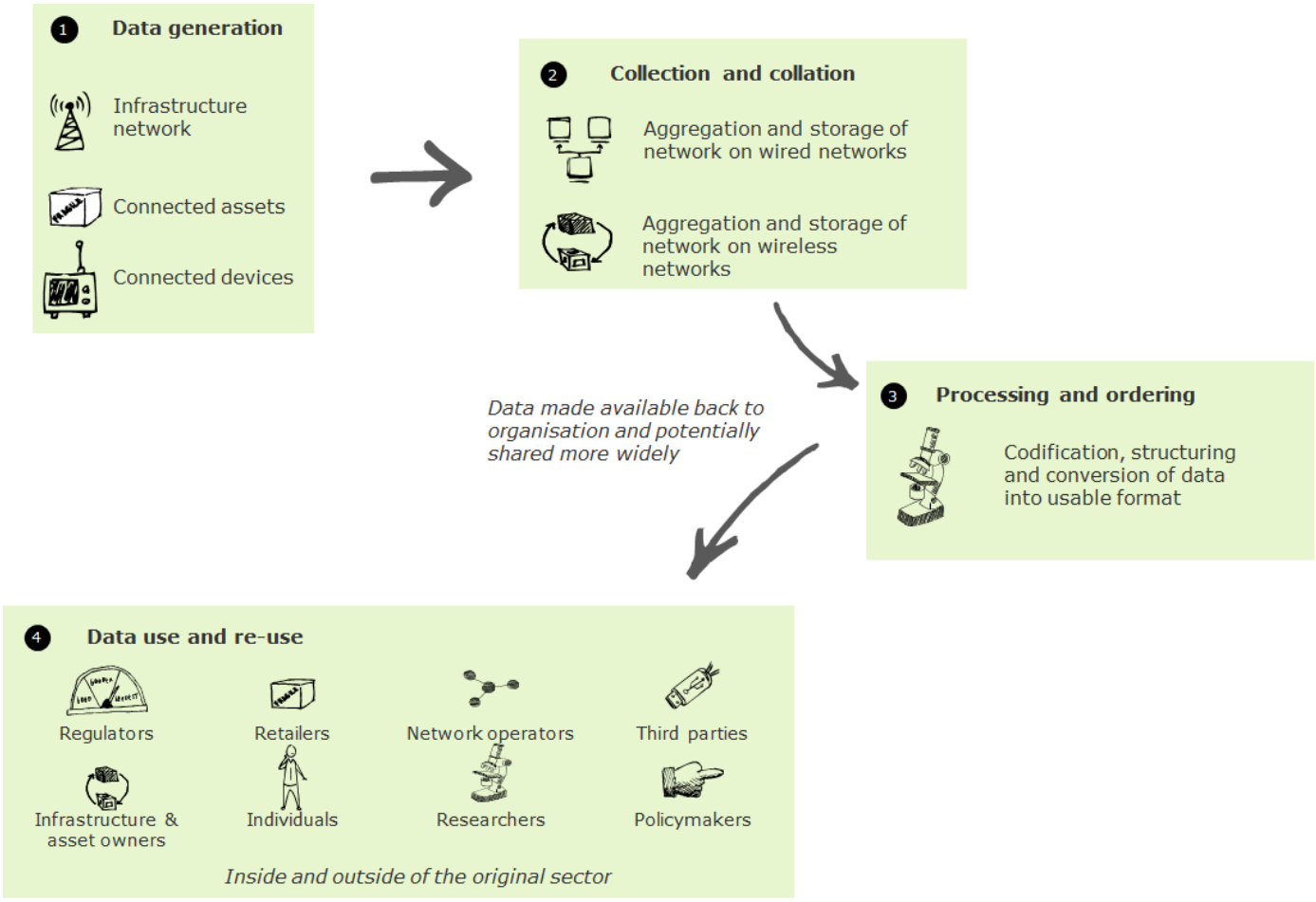
Data can move between being closed (held within the organisation in which it is generated, in the example case above normally the hospital), shared and open. Open data can be regarded a public good and means providing unrestricted access to everyone. On the other hand, shared data means providing restricted access to the data for certain entities only. This may be because it provides a revenue stream and is therefore only available to those willing to pay for it, or it may be because it is sensitive in some way.

2.3 Machine-generated, non-personal data value chain

There are some common stages in how data is generated from machines and then ultimately used or re-used by data holders or third parties. The value chain is set out in Figure 2-4, charting how non-personal data is generated by devices and ultimately used and re-used. It is important to note, however, that this entire process may take place in real time with data being generated, collected, processed and used or re-used immediately or in seconds (even where third parties are involved). Alternatively data may be generated and then stored for months or years before being processed and/or used or re-used.

[5] Based on: http://www.ibmbigdatahub.com/infographic/four-vs-big-data
[6] Based on: https://theodi.org/data-spectrum

Figure 2-4: Value chain

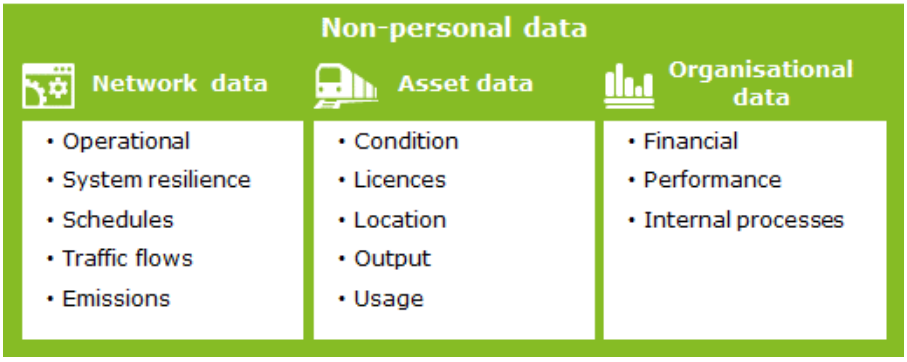


Source: Deloitte analysis

2.3.1 Data generation

Data is generated from connected devices, connected assets (that may not be moveable) and the infrastructure network. It can cover a range of topics.

Figure 2-5: Types of non-personal data



Source: Deloitte analysis

Data can be generated by machines through a diverse range of sensors intended to gather data regarding external conditions (e.g. temperature or movement) or sense the presence of artificial markers (e.g. radio-frequency identification tags). In other instances, data is not collected deliberately, but is rather a by-product of a machine's core function. This is known as exhaust data and could include a sensor noting road conditions when its main

function is geospatial location tracking. Aside from sensors, non-personal data from machines may also be generated through use alone, i.e. data on when the device is switched on, being used or left idle.

Projects using machine-generated, non-personal data can make use of existing sensors or involve the deployment of new sensors. Using existing sensors is less expensive, but the data may be less appropriate to the specific use case. For example, there may be existing data resulting from the production process in a factory currently employing machinery which is able to predict equipment failures. However, it may be more effective to add sensors that more directly measure the operating condition of the machinery. New sensors could lead or follow the development of new analytical tools.

2.3.2 Collection and collation

Individual machines can generate gigabytes of data each day in operation. Given that these are all connected devices, data from multiple machines is likely to be aggregated and combined centrally. The next stage in the value chain is thus the collection and compilation of data from multiple devices. Data collection might take place over wired or wireless networks, including over mobile networks. Collecting data for processing can involve a number of challenges:

- devices need to be connected;
- devices may be geographically dispersed, across a facility or a city or around the world, and conditions may mean it is difficult to connect to a network; and
- the data will need to be secured in transmission.

Indeed, the volume of data being transferred often means that network conditions are highly relevant to the practicality of collecting data.

2.3.3 Processing and ordering

In many cases, the collected data may be unstructured or unordered. For this data to be usable it needs to codified and converted into a format and structure that can be used and understood by others. It should be noted that this stage does not always occur, especially with exhaust data that may be of seemingly little value to the data generator. The processing of machine-generated, non-personal data is enabled by:

- cloud services that allow for the storage and processing of large volumes of data without unduly high fixed costs;
- machine learning algorithms that can identify patterns in large datasets despite gaps in the structural understanding of the problem; and
- falls in the cost of other physical components required, particularly sensors.

This stage transforms datasets so that they are suitable for use and re-use by a wide range of audiences.

2.3.4 Data use and re-use

The final stage of the value chain is where data is applied to generate outputs, outcomes and ultimately commercial and social value. Data can be used in a range of ways by regulators, consumers, researchers and industry to inform decision-making, power algorithms and new business models and support research. Subsequent analysis can attempt to address research questions with which it is possible to discern rules that improve performance, such as what pattern of planting will tend to maximise yield. Alternatively, it might take the form of operational optimisation, such as what pattern of planting will maximise yield given the characteristics of the field being sown. Over time, those two kinds of processing are likely to support one another.

Who is able to use and re-use the data will depend on how the data is made available. This relates to where on the spectrum of openness it sits, i.e. open, shared or closed. The importance of sharing data is explored in the next chapter.

3 The importance of data sharing in generating benefits

Data sharing allows data to be put to new and valuable uses. Many of the benefits of using and re-using machine-generated, non-personal data are made possible, or enhanced, by third party organisations accessing data generated by others.

3.1 The benefits of data sharing

Data sharing can bring enormous value to the economy, as it allows data to be put to new and valuable uses. Thus, limiting data to the boundaries of the generating organisation precludes a range of ways in which that data might be used to generate additional value.[7]

Firstly, specialists who are able to add value in processing data are often not the same as those tasked with being data holders. Putting non-personal data in the hands of organisations and individuals that are best equipped to use it effectively, rather than those who are simply custodians of it, can generate an increase in labour productivity and increased efficiency.

Secondly, organisations in different sectors may face similar business challenges or be affected by the same underlying constraints. It could be the case that they are able to learn lessons from using and re-using non-personal data from other sectors. Putting data in to the hands of other organisations operating in different sectors to the data holder could yield new insights, or create innovative uses outside the original intention of the data.

Thirdly, there may be circumstances in which combining personal with non-personal data (or different types of non-personal data) yields larger benefits than if they were analysed in isolation. Consequently, there will be opportunities in which sharing, aggregating and combining datasets may be appropriate in order to allow more complex and holistic analysis to reach firmer conclusions or more valuable insights.

Finally, providing wider access to non-personal data lowers barriers to entry for third parties and other players that could potentially enable innovative uses of the data. This could, in turn, stimulate market entry and competition.[8]

3.2 Types of data sharing and how it can occur

Broadly speaking, data sharing can be categorised on the basis of who the data is being shared with:

1. **Horizontal data sharing:** Sharing that occurs **between organisations** involved in the **same commercial or non-commercial point of the value chain,** e.g. businesses selling the same product in the same market place, or for use in benchmarking (to compare service levels between municipalities).

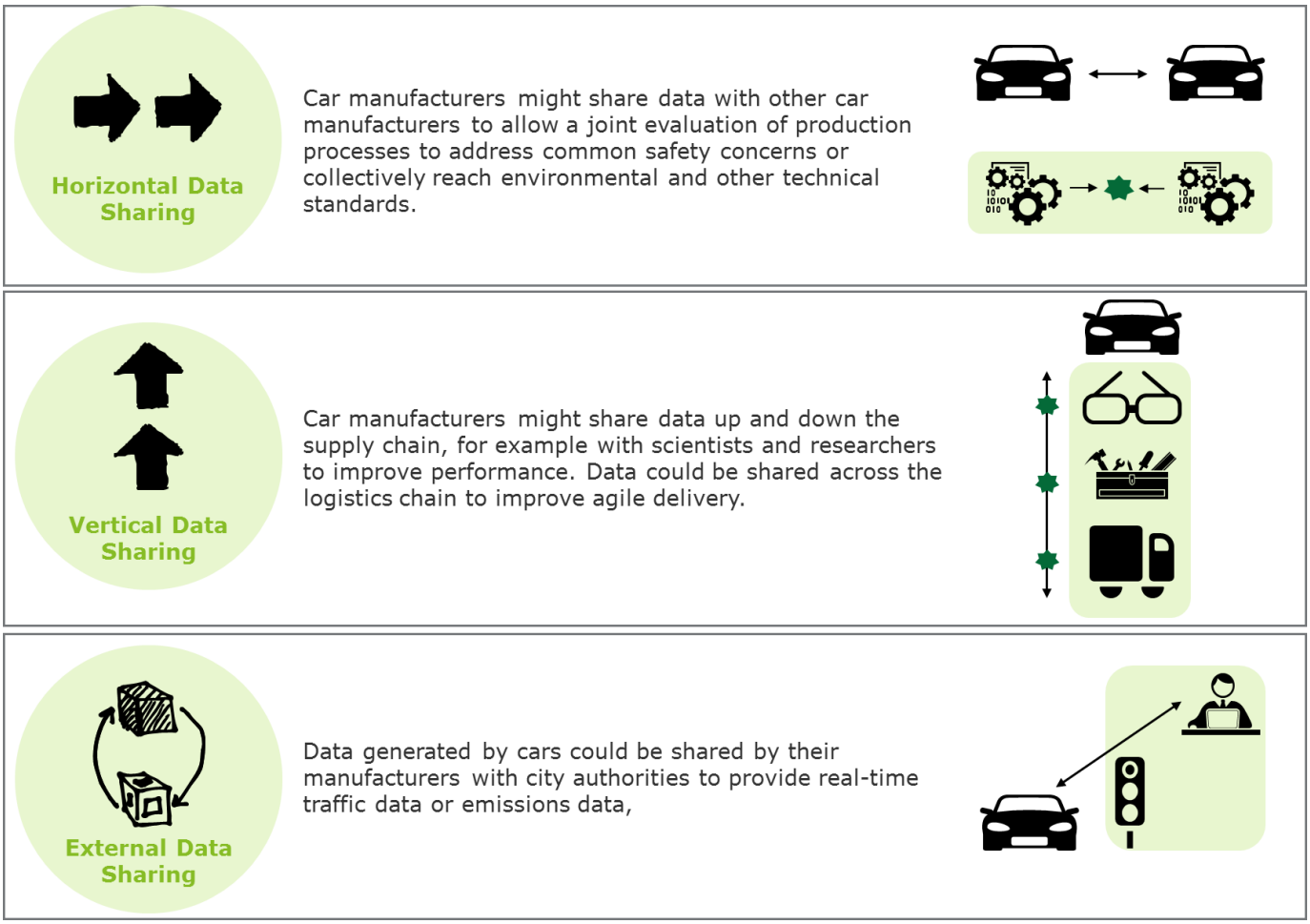
[7] Note, we are not considering intra-organisation barriers.
[8] There is also an established literature on the potential for the sharing of some data, e.g. prices, to be anti-competitive. This is discussed in the Policy recommendations section.

2. **Vertical data sharing:** Sharing that occurs **between organisations** who have a **customer or supplier relationship, directly or indirectly,** e.g. to build and improve services or quality based on user-level feedback.

3. **External data sharing:** Sharing that occurs with organisations **outside the sector** engaged **in different commercial or non-commercial activities**, e.g. sharing with third parties to analyse and buy-back aggregated datasets that have been contributed to by multiple parties.

Some examples are described below using the automotive sector for the purpose of illustration.

Figure 3-1: Data sharing typology and examples



Source: Deloitte analysis

Data can be shared through a number of mechanisms depending on the purpose, nature and intended openness, including:

- **Shared as open data**, so it can be used, re-used and redistributed without restrictions subject only, at most, to the requirement to attribute and share alike.[9]

- **Shared via commercial agreements** (involving a fee or not) between data holders and interested parties.

[9] Source: http://opendatahandbook.org/guide/en/what-is-open-data/ accessed 13 February 2018

- **Shared via restricted access facilities** such as data labs or sandbox environments where users can use the data under certain restrictions.

In terms of ways this can be implemented practically, there are a range of options available to data-sharing organisations. Data can shared technically via APIs, as static downloads, as part of data portals and even in hard copy. For instance, Google Maps APIs let developers embed Google Maps on webpages using JavaScript or Flash interface, and Amazon Product Advertising API gives developers access to Amazon's product selection and discovery functionality to advertise Amazon products and monetise a website.

Data can also be shared by the initial holder of the data or by a third party. The inclusion of third parties is envisaged for the European automotive sector with OEMs (original equipment manufacturers) sharing data through a neutral server. In the neutral server model, service providers looking to use data generated by vehicles can either do so directly through OEM systems, or through neutral servers operated by independent third parties. Additional data can be requested by the neutral server operator without giving away the identity or plans of third parties to use that data. This allows for novel business models to be developed without the concern that the best new concepts will simply be hijacked by OEMs. It will also allow for horizontal data sharing to the extent that one neutral server can offer data sourced from vehicles built and operated by multiple OEMs. This approach has been adopted by the OEM trade association.[10]

[10] Source: http://www.acea.be/uploads/publications/ACEA_Position_Paper_Access_to_vehicle_data_for_third-party_services.pdf, accessed 9 March 2018

4 The overall benefits of machine-generated, non-personal data in the chosen sectors

The use and re-use of machine-generated, non-personal data can result in a number of benefits for the economy and society. These include contributing to improvements in productivity and efficiency, stimulating innovation and contributing to societal improvements.

4.1 Identifying and quantifying the benefits of machine-generated, non-personal data

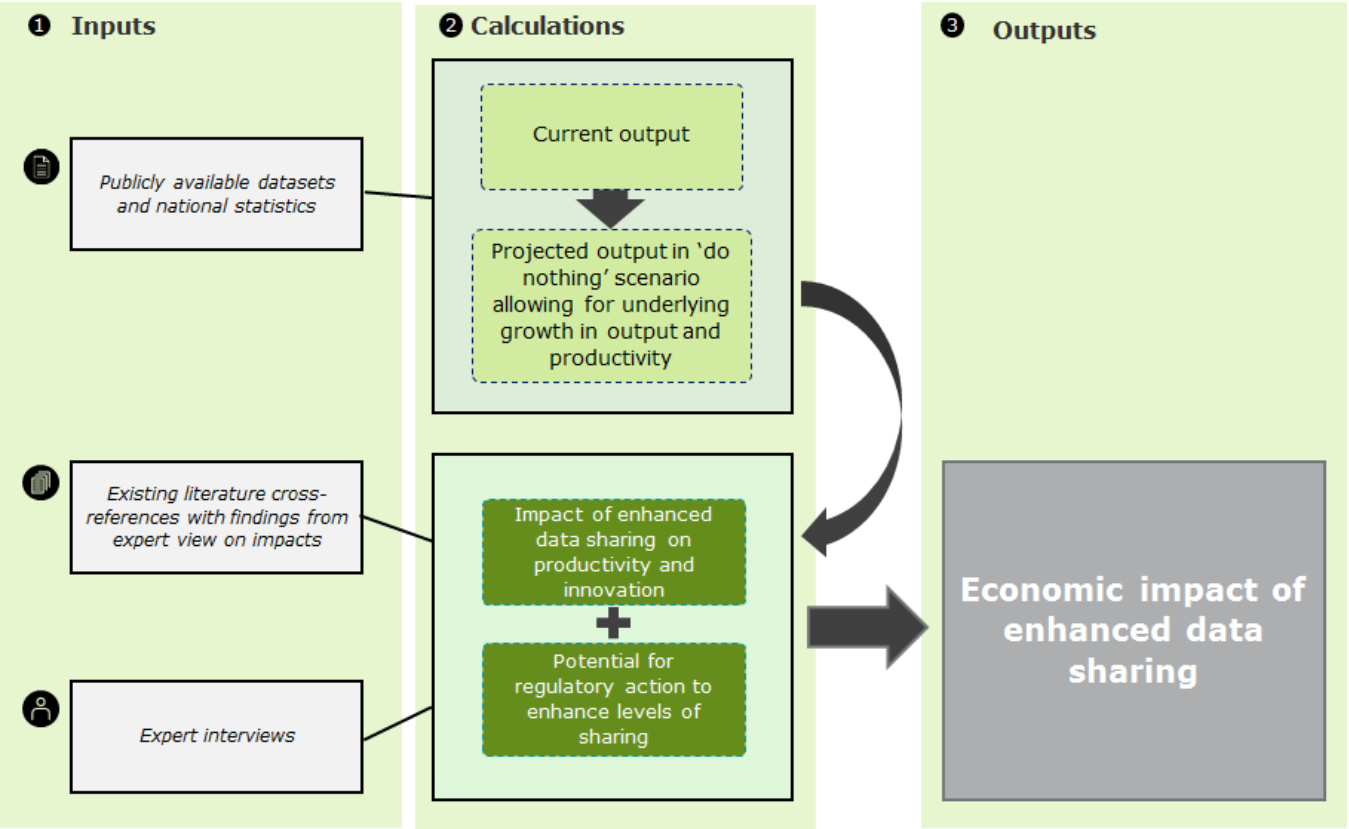
In order to estimate the size and value of benefits of machine-generated, non-personal data and the impact of the obstacles preventing greater data sharing, this report has developed a model to illustrate the scale of possible impacts that could be achieved through enabling a more open approach to sharing.

The high-level quantitative model for the five sectors studied is structured in the following way:

- We have identified the output indicators across five sectors that are most likely to be influenced by IoT and the use and re-use of machine-generated, non-personal data. Data has been collected on these indicators (current and future) from Eurostat and similar established sources.
- Based on the existing literature and discussions with experts, the contribution of machine-generated, non-personal data being shared between organisations is estimated and then tested in an expert survey.
- The survey also explores levels of data sharing of machine-generated, non-personal across five sectors, the impact of increasing sharing and the key barriers to sharing.

The experts consulted have over 900 years of collective experience across the sectors considered in this study. Their views, alongside the existing literature and publicly available data, form the basis of the modelling carried out for this study. A visual representation of the high-level approach is given in Figure 4-1 and further detail on the quantitative methods used to assign empirical values to benefits and the impact of barriers is provided in the Annex: Model approach.

Figure 4-1: High-level approach to estimating the economic value of enhanced data sharing



It should be noted that the results shown in this chapter and the next are based on a number of assumptions, subjective expert views and the existing literature on a nascent market. As better quality data is available and more use cases emerge, the results will inevitably change and provide more insight.

This chapter begins with a qualitative discussion of categories of benefits arising from the use and re-use of machine-generated, non-personal data. It then moves to present a number of detailed examples before presenting the quantification of the economic benefits of non-personal data sharing.

4.2 Types of benefits resulting from machine-generated non-personal data

Our analysis for this report suggests three categories of benefits arising from the use and re-use of machine-generated, non-personal data:

- **Productivity and efficiency improvements.** For instance, data re-use could result in higher crop yields and enable the roll-out of new resource management programmes that reduce costs and prolong asset life through optimised maintenance regimes.
- **Greater levels of innovation.** For instance, providing the data to underpin new technologies that use artificial intelligence, and enable the development of disruptive new business models.
- **Societal improvements.** For instance, through supporting the optimisation of traffic and energy usage.

Below we set out some use cases of machine-generated, non-personal data produced by IoT devices leading to these types of benefits.

4.2.1 Productivity and efficiency

Productivity and efficiency benefits arise from the use and re-use of machine-generated, non-personal data to inform better decision-making and resource allocation, which in turn can increase outputs, reduce waste and better align processes. Data can also be used to improve the resilience of assets and systems and thereby reduce the frequency, duration and impact of disruptive events. The results can be seen in, for example, higher crop yields or shorter and quicker logistics chains. Some specific examples from the chosen sectors are outlined below. Improved productivity efficiency can in turn create benefits for European citizens through higher wages and lower prices.

Figure 4-2: Productivity and efficiency benefits from machine-generated, non-personal data

Agricultural sector

IoT analytics can help farmers analyse real time data like weather, temperature, moisture, prices or GPS signals and provide insights on how to optimise and increase yield, improve farm planning and make smarter decisions about the level of resources needed.[11] Most data generated in the agricultural sector from smart devices is non-personal in nature. The data generated by sensors or agricultural drones covers meteorology, soil, livestock, crops, the use of water and fertiliser, feeding rates and so on.

The application of this non-personal data allows farmers to manage their crop production more precisely through so-called 'Precision Agriculture' and is typically used for field crops such as corn, soybeans, wheat and rice. Using GPS-based field maps and soil sampling and yield monitor data, farm machinery equipped with variable-rate technologies (VRT) can use insights from data to adjust seed planting density and application rates for herbicides, pesticides and nutrients based on variations in soil quality, topography, moisture, weeds and other factors. This can lead to increased yields and lower input costs.[12]

As another example of application, a major agribusiness is providing soil sensors to farmers under a fertiliser supply contract to provide data on what each field and area of fields needs in terms of nutrient enrichment to maximise yields. This data is shared with farmers, which in turn helps them understand their land better and be more effective in land utilisation.

CASE STUDY: MOOCALL

MooCall collects data from sensors attached to pregnant cows and alerts farmers before the onset of calving, reducing the amount of time farmers need to spend monitoring livestock and helping where possible in avoiding the costs of lost cows and/or calves.[13]

Manufacturing sector

The deployment of IoT devices and connected assets and networks in manufacturing is increasingly important in Europe as it is perceived as a way of creating competitive advantage through advanced analytics. The growing trend towards automation and data exchange is often called Industry 4.0. While manufacturers have used connected devices for some time, new advanced software and architecture (such as edge analytics and AI) is allowing them to make more automated decisions.

- **Predictive maintenance**: Manufacturers are increasingly equipping their products with sensors with which product owners can monitor the condition of those products. These generate new streams of data that can be used for predictive maintenance, predictive quality and other supply chain applications. The ideal outcome is to predict outcomes at a high level of confidence, optimise real-time operations and asset management and minimise failures.

[11] IBM (https://www.ibm.com/blogs/watson/2016/12/five-ways-agriculture-benefit-artificial-intelligence/, accessed January 2018).
[12] IFAMA (https://www.ifama.org/resources/Documents/v19ia/320150137.pdf, accessed January 2018).
[13] See https://www.vodafone.com/business/news-and-insights/case-study/moocall for more details.

- **Shop floor operational improvements**: Low cost sensors attached to machines, wireless connectivity and big data processing tools has made it progressively easier to collect performance data and monitor equipment condition. For instance, critical machine tools are designed to operate within certain temperature and vibration ranges. Sensors can work to prevent malfunctions by sending an alert when tools deviate from these parameters.
- **Suppliers and supply chain**: By connecting plants to suppliers, parties involved in a supply chain can trace interdependencies to help to identify issues before they happen and reduce inventories. Systems can also conduct remote monitoring of inventory and track parts/products as they move through the supply chain.

There are also examples of car manufacturers using non-personal data to undertake Predictive Quality Management (PQM), where streams of real-time data are used to underpin predictive models that make processes smarter by reducing failure rates. The creation of a predictive model to measure vibration and temperature of a spot welding machine can predict the likelihood of a weld failure. The productivity and efficiency benefit is substantial: each car has thousands of spot welds that hold the car together and failures are around 5-7%, which can add to the cost of production. Eliminating these failures can therefore reduce costs.

Healthcare sector

Connected devices, assets and networks allow different elements of procedures across the healthcare sector to be better linked and synchronised. IoT applications will require connecting hospital assets (instruments, biological processes, drugs, medical devices, scanners, other physical assets such as beds) and aggregating that data with information about patients and staff to create statistically significant datasets that can be used for multiple purposes.

Machine-generated data could be used in a wide range of applications that will improve productivity and efficiency in healthcare solutions, from automated intervention and monitoring using smart devices all the way through to augmented reality-supported or even AI-implemented surgery. Predictive analytics could also be used to predict the clinical "flight path" of a patient in the healthcare sector. By leveraging historical data from other patients with similar conditions, predictive algorithms can be created to predict the trajectory of a patient over time. Using a combination of machine learning approaches, clinical flight path models account for historical and cohort trends, and have the ability to forecast likely patient outcomes in terms of cost and complications. This can then inform the optimum protocol for treating that patient as quickly and cost effectively as possible.[14]

Source: Deloitte analysis

4.2.2 Innovation benefits

It is harder to identify specific metrics around innovation to quantify its impact, but the impact of machine-generated, non-personal data sharing can be seen in the form of new products, services and business models. Making data available widely can lower barriers to entry as new market players can access and use this data to power disruptive services, products and business models. This can in turn increase competitive pressures leading to lower prices for consumers and more choice. Innovative IoT applications can therefore create benefits for European citizens through new employment and investment opportunities.

Figure 4-3: Innovation benefits from machine-generated, non-personal data

Automotive sector

A number of automotive manufacturers are developing connected car initiatives via telematics platforms. The data is shared with insurance companies, with the intention that an insurance company will be able to specifically tailor a policy for each driver based on their driving behaviour. Around 3 billion driven miles of data have already been collected, helping insurance companies develop personalised products.

[14] Health Catalyst (https://www.healthcatalyst.com/big-data-in-healthcare-made-simple, accessed January 2018).

There are then two central concepts intended to support the wider sharing of data from OEMs in the automotive sector.

First, the extended vehicle concept (ISO standard 20077-1). This allows external organisations to access data through a range of interfaces (e.g. an on-board diagnostics – OBD – interface for emissions control, diagnosis and repair and maintenance and a web interface for more general use).

Second, the neutral server concept where third parties are able to access data (which may relate to vehicles from multiple OEMs) through an independent server. The intention is to allow innovative new business models to develop without the involvement of OEMs. Some stakeholders have criticised this approach as insufficient, arguing for a more general sharing of car data and more direct, immediate and bi-directional (i.e. data can be sent and received) connection between third party services and the data generated by the vehicle.[15]

Much of this data is personal to the extent it is connected to a specific vehicle identification number and thereby the owner. However, it can be anonymised for the purposes of sharing. If an individual owner were operating such a vehicle, the data generated would be personal (at least initially). However, personal data is less relevant in the fleet setting, as in the Panda Bus Dynamic Shuttle mobility experiment run by Ford at Mobile World Congress in Shanghai, which tested the potential for more flexible scheduling in Dalian, China.

Finally, possibilities for external data sharing are being explored in a range of ways. This includes plans to connect the emergency services to the hazard warning lights in cars, providing an early signal that there may be problems on a particular stretch of road.

Manufacturing

CASE STUDY: SEMIOTIC LABS

Semiotic Labs works with businesses operating rotating equipment in predicting failures in order to improve performance. The applications of this include: defence, naval and manufacturing industries (anything powered by AC induction motors). Projects will often involve both installing new sensors (existing data collected in the manufacturing process) and analysis.

There are two source of value inherent in their operations: (i) improved safety; (ii) improved productivity due to diminished downtime. Sharing is crucial to this business model in two respects: firstly, vertical sharing, which is integral and takes place between Semiotic Labs and its client; and, secondly, horizontal sharing, since there is a general provision (from which firms could but do not opt out) where insights from different clients are shared to optimise overall performance. It is a quid pro quo, where in return for sharing insights generated from their own data, they gain access to insights gained from other firms' data.

Semiotic Labs noted that sufficient connectivity infrastructure (including mobile connectivity) is required to facilitate data sharing. Individual machines are capable of generating gigabytes of data each day and some factories could have dozens operating.

Source: Deloitte analysis

4.2.3 Social benefits

The wider benefits to society are manifested in cost reduction, quality improvement and greater choice for consumers. Benefits such as reduced healthcare costs, improved levels of care and reduced environmental degradation that are derived from more intelligent and efficient systems accrue to society as a whole, not just particular sectors or groups of consumers. However, higher business productivity and increased energy and transport efficiency have to be balanced against risks to security and resilience, both known and unanticipated. For instance, issues surrounding collection and use of data should be sensitive to context; the data produced by

[15] Manifesto for fair digitalisation opportunities (https://www.grupoaseguranza.com/adjuntos/fichero_25280_20180423.pdf, accessed April 2018)

a sensing device with a specific individual is very different to the environmental data produced by a buoy floating in the ocean and must therefore be handled differently.

Figure 4-4: Social benefits from machine-generated, non-personal data

Smart Cities

Cities are already generating a lot of data that is only now just beginning to be shared with citizens and third parties and find its way into products, services and decision-making. In many ways, Smart Cities are ideally placed to use machine-generated, non-personal data to affect social change as this data can be applied at a system-wide level.

For example, machine-generated data on energy consumption and usage patterns in a city can be input into smart grids to gain useful insights, detect anomalies and better align demand and supply. Such predictions play a role in planning future aggregated electricity demand, future system supply and estimating flexibility in electricity distribution networks.[16]

CASE STUDY: CASCAIS

Cascais is a coastal city in Portugal that has deployed innovative solutions in a number of areas in recent years, including participatory budgets and integrated transport services. Other areas were felt to be progressing more slowly, however, so they introduced a command centre model based around sharing the data in a Digital Command Centre in order to facilitate real-time collaboration and more sophisticated analytics. In practice, this connection would take place through a series of APIs connecting different functions in the city administration.

The first domains included will be waste management, civic protection and emergency management and mobility. The waste management system alone is expected to save the city around €900,000 a year with sensors to track optimum fill level of more than 400 underground recycling bins, allowing the city to optimise routes for collection trucks.

Integration through this command centre approach is expected to deliver distinctive savings which might otherwise not be realised. For example, connecting the waste management operation to the mobility services and thereby data on road construction and repairs. This would allow for the timing and routing of collection trucks to reflect the operation of other city services. Other areas are also set to be included: security and surveillance; energy (street lights and buildings); health; education; green spaces and environmental control; and water and sanitation, all of which will bring additional benefits.

Progress so far has varied by area and much of the impact is expected to result from sharing across departments and enabling deeper analytics. Exchanging data requires the implementation of new APIs and a forum. The digital command centre is intended to provide that unified vision and the capacity for cross-cutting analysis.

Dutch cities

Dutch Smart Cities are also pioneering new approaches to data sharing, each taking charge of innovating in a specific space:

- Amsterdam is focused on the circular economy – making more efficient use of resources with more efficient production processes and sharing and reuse of physical goods.
- Utrecht is focused on healthy urban living, with extensive support for increasing cycling and a reduction in parking spaces to almost zero. This includes a focus on mobility-as-a-service.
- The Hague is focused on safety and security – making individuals and businesses safer both physically and in cyber space.
- Rotterdam is focused on resilience against both criminal acts and acts of nature.

While some of these will naturally involve personal data, there is also extensive non-personal data generation and sharing. This includes the numbers travelling down a specific street at a specific time and therefore the number of bicycles expected,

[16] IEEE Smart Grid Big Data Analytics, 'Big Data Analytics in the Smart Grid' (https://smartgrid.ieee.org/images/files/pdf/big_data_analytics_white_paper.pdf, accessed January 2018).

improving forecasts for congestion. It also includes the data collected by smart lighting and used (among other things) to deter crime without a visible police response which might exacerbate the situation, e.g. by changing the lighting in reaction to a disturbance. That data can be non-personal noise tracking with identification only once a problem has been identified and the camera activated in response.

Amsterdam is considering a requirement for companies that wish to collect data in a public setting to obtain a licence that would then require data sharing (in addition to creating a source of revenue). This provides a potential wider model for the public sector in its role as a regulator to encourage data sharing (outside the setting of economic regulation where it is already the norm).

Healthcare

Healthcare organisations are considering new initiatives to track staff, patients and physical assets. Tracking can take place through staff name badges, patient tags and tags on equipment. Vodafone and Deloitte are developing an end-to-end IoT medical device solution to improve the delivery of patient care. This has included the development of a Connected Medical Device application for a nutrition infusion pump, a prototype of which was demonstrated at the Mobile World Congress 2018 in Barcelona.

The use of beds, patient progress through emergency departmetns and other areas crucial to hospital productivity can be tracked. Data can then be shared by the primary care organisation with suppliers, analysts, peers (e.g. other hospitals) and external stakeholders stakeholders (e.g. health ministries).

The data collected will often be a mix of personal (patient and staff) and non-personal data. Patient data is naturally often particularly sensitive. It could be that, if using personal data becomes more challenging over time, optimisation within hospitals could take place using non-personal data alone. This might focus on the optimisation of the use and maintenance of equipment.

Source: Deloitte consultation

4.3 Quantified impacts of machine-generated, non-personal data across five sectors

4.3.1 Current and future output

The value of current output and future projected output (2027) is derived across five economic sectors of interest. Rather than estimating the value of the entire sector across the EU economy, the analysis instead focuses on a selection of specific example use cases within the sector that are most relevant to the use and re-use of machine-generated, non-personal data. This means that in some instances, such as in the healthcare sector where there are many potential use cases for IoT data, the vast majority of those use cases involve personal data rather than non-personal data and are therefore outside the scope of this report.

These sector segments have been further identified based on the applicability of benefits to data sharing (i.e. where in the overall sector the most non-personal data is generated) and the availability of reliable EU-level data that can be reasonably quantified. Analysis of relevant segments yields more meaningful insights on the potential benefits of data sharing, which can later be extended as markets mature. This baseline data is used to estimate the gross value of IoT in a number of use cases in 2027.

The current or baseline value of this output across the five sectors' relevant segments is shown below.

Figure 4-5: Value of use-cases measured in each sector (2027)

Sector	Baseline data (EU) and relevant segments	Value in 2027
Agriculture	Output of the agricultural industry, particularly meat and crop production.	€434bn

Automotive	Focused on end use, covering: i. Fuel consumption. Measured by the cost of emissions from cars. ii. Maintenance and repair services. Measured using costs and passenger vehicles in the EU plus a reduction in vehicle damage.	€87bn
Healthcare	Focused on patient care, covering: i. Expenditure on resource management. Measured by current curative and rehabilitative expenditure plus expenditure on laboratory services, imaging services, patient transportation and therapeutic appliances, and other medical durable goods. ii. Counterfeit drugs. Measured by pharmaceutical expenditure as a percentage of health spending and an approximate proportion of counterfeit drugs.	€123bn
Smart City	Expenditure on energy for streetlights.	€3bn
Manufacturing	Total value of the production of manufactured goods.	€5,000bn

Source: Deloitte analysis based on Eurostat data

4.3.2 Current levels of data sharing

As discussed, one of the key drivers of benefits is data sharing. However, very little publicly available data exists on the extent to which valuable data is currently being shared. As part of our survey, we sought views on levels of non-personal data sharing.

Of the three types of data sharing, vertical data sharing is the most common type of sharing taking place currently across all sectors, and it is highest in the healthcare and manufacturing sectors.

Figure 4-6 illustrates, for non-personal data being generated by IoT devices, the extent to which data is currently being shared along the continuum of no sharing to sharing of all relevant data. The table shows that only 25% of valuable data sharing is currently taking place at the horizontal level in agriculture, for example.

Figure 4-6: Extent of data sharing currently across the EU28

Type of data sharing	Across all sectors considered	Healthcare	Manufacturing	Automotive	Smart Cities	Agriculture
Horizontal	32%	33%	35%	30%	35%	25%
Vertical	47%	48%	58%	43%	43%	45%
External	31%	20%	33%	38%	25%	40%

Source: Deloitte analysis. The percentages given in this table do not sum. Instead they reflect on a scale of 0% (where no data is being shared) to 100% (where all relevant data is being shared where it would be valuable to do so) the proportion of data being shared at the horizontal, vertical and external level for each sector.

Figure 4-6 demonstrates that, across all sectors, experts felt that horizontal and external data sharing were only a third of their potential, but that vertical data sharing was closer to half its potential levels. Levels differ between sectors, but horizontal and external sharing are consistently the lowest. With the exception of vertical data sharing in manufacturing, the proportion of data sharing is consistently below 50% meaning there is considerable scope for increasing the level of valuable sharing.

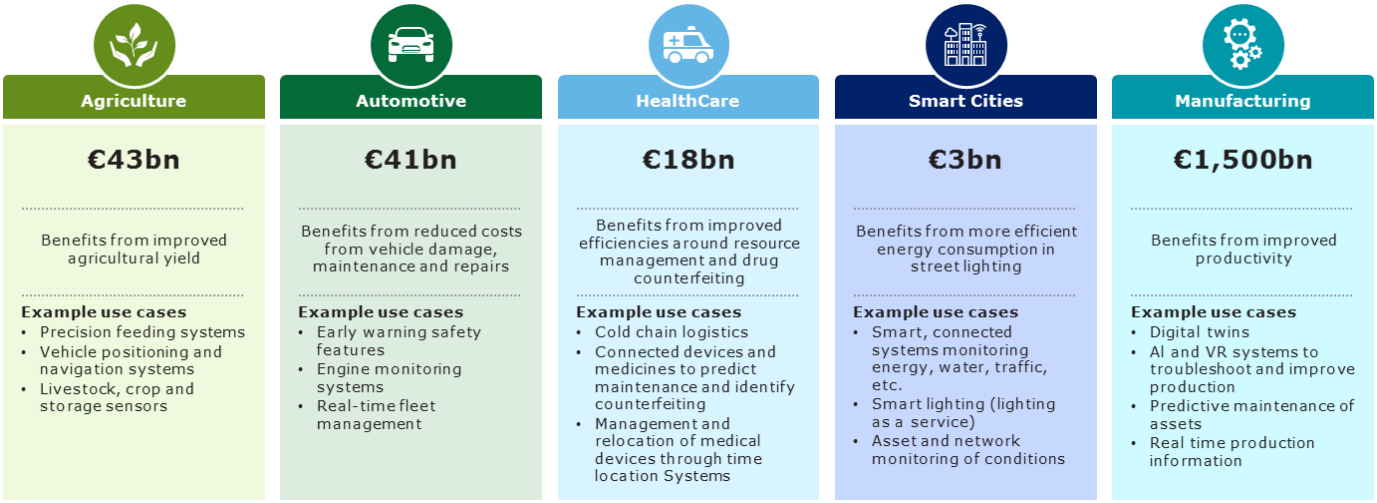
Given that the potential benefits identified are contingent on data sharing, these figures suggest that there is still a significant amount of economic value that could be realised if sharing increased.

4.3.3 The value of increased output caused by increased IoT penetration and the use and re-use of machine-generated, non-personal data – overall IoT benefits

Estimates for predicted IoT technology penetration for each sector in 2027 have been generated by conducting a literature review and testing expert opinion on the extent to which IoT penetration will grow over time. It is unrealistic to assume 100% IoT penetration.

Using the expected growth in IoT penetration, and leveraging assumptions on the expected use and re-use of machine-generated, non-personal data (based on the qualitative use cases), we have estimated the value of the benefits of the data.

Figure 4-7: Potential value of machine-generated, non-personal data in the EU in 2027, given predicted IoT penetration.



Source: Deloitte analysis. Note the above refers to economic benefits rather than additional revenue.

The largest quantified benefits from machine-generated, non-personal data are in the manufacturing sector, which reflects the large number of potential use cases and proliferation of IoT devices and assets. It also reflects that personal data is much less important than in most of the other sectors. The impacts are smaller in the other sectors for a range of reasons:

1. The agriculture sector is generally smaller as a share of EU28 GDP than the manufacturing sector. Any impact is necessarily less important in terms of aggregate economic output. However, the impact remains salient for policymakers particularly in considering measures that might drive rural economic development.

2. In the automotive and healthcare sectors many of the most important use cases relate to the use of personal data. Even non-personal data is largely expected to be used in combination with personal data. It is worth noting for the automotive sector in particular that even expert opinion is often unclear about the grey areas between personal and non-personal data. This may affect the reliability of the results.

3. In the Smart Cities sector the sheer diversity of the potential impacts means that the result estimated likely only captures part of the potential in the sector. Nonetheless, relative to the baseline (expenditure on energy for streetlights) the impact is large.

4.3.4 The value of increased output caused by increased IoT penetration and the use and re-use of machine-generated, non-personal data – attributable to sharing

Given that some of the impacts of machine-generated, non-personal data will be unrelated to data sharing, expert opinion was used to inform what share of the impact is likely to relate to horizontal, vertical and external data sharing.

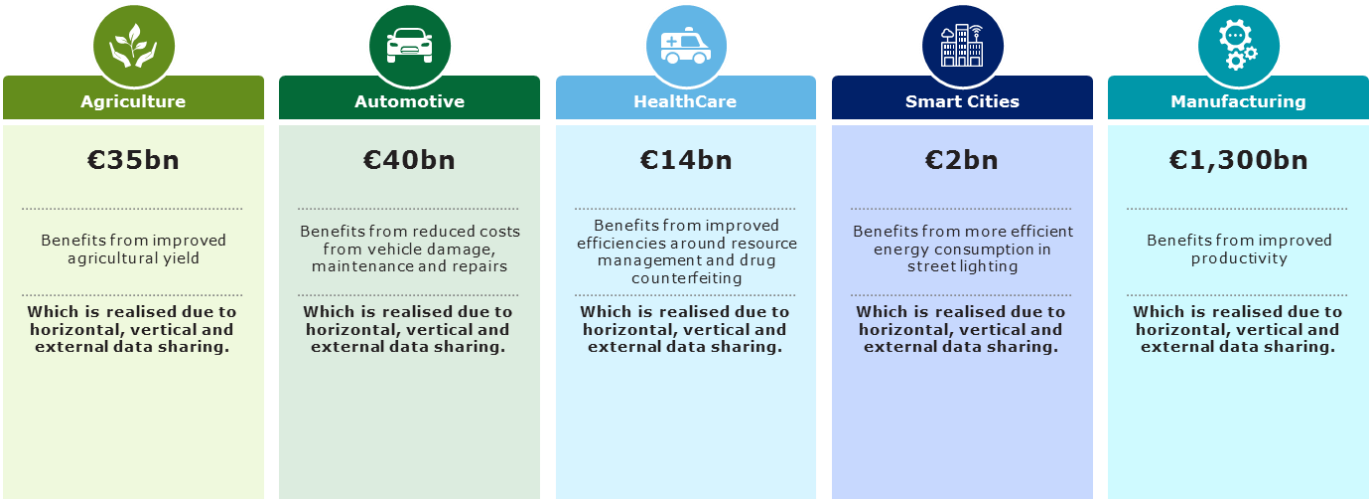
Figure 4-8: Share of the value attributable to types of data sharing

Type of data sharing	Healthcare	Manufacturing	Automotive	Smart Cities	Agriculture
Benefits unrelated to data sharing	19%	14%	8%	16%	18%
Benefits that depend on horizontal data sharing	20%	23%	24%	22%	20%
Benefits that depend on vertical data sharing	37%	42%	33%	29%	37%
Benefits that depend on external data sharing	24%	22%	36%	34%	26%

Source: Deloitte analysis.

Data sharing is found to be necessary to realise most of the benefits of machine-generated, non-personal data. Figure 4-9 illustrates the economic value of benefits related to data sharing (i.e. the overall benefits minus the share of benefits unrelated to data sharing).

Figure 4-9: Potential value of machine-generated, non-personal data sharing in the EU in 2027, attributable to data sharing.

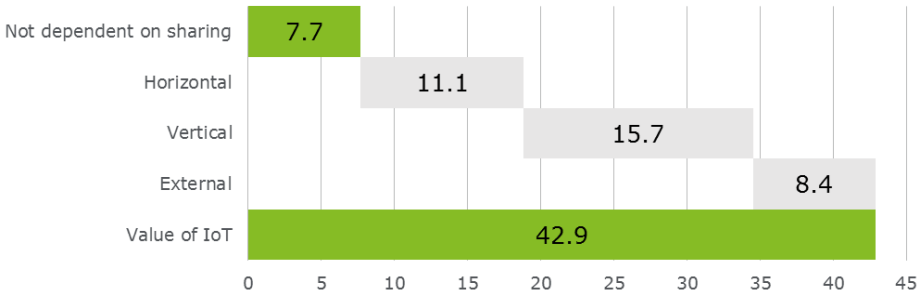


Source: Deloitte analysis. Note the above refers to economic benefits rather than additional revenue.

4.3.5 The value of increased output caused by increased IoT penetration and the use and re-use of machine-generated, non-personal data – by types of sharing

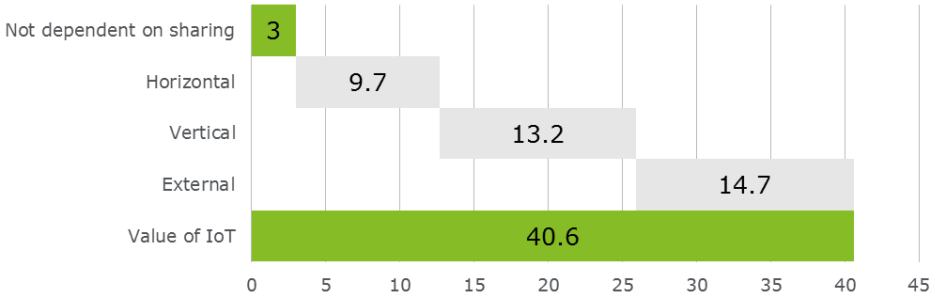
The analysis above can be further disaggregated by type of data sharing (i.e. horizontal, vertical and external) for each sector, as illustrated in the figures below.

Figure 4-10: IoT impact on EU agricultural yields in 2027, and shares depending on data sharing (€ billion)



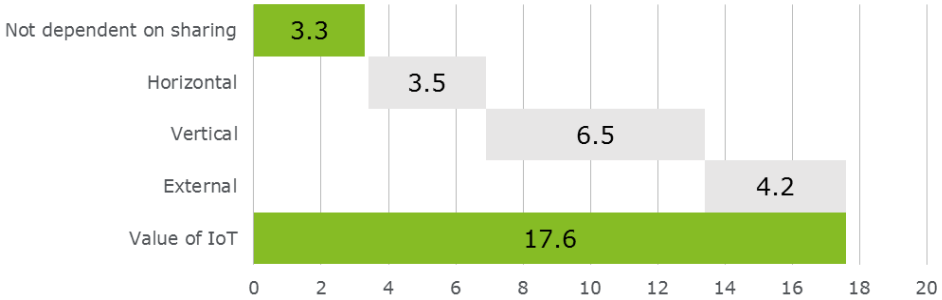
Source: Deloitte analysis. Note the above refers to economic benefits rather than additional revenue.

Figure 4-11: IoT impact on EU vehicle damage and repairs in 2027, and shares depending on data sharing (€ billion)



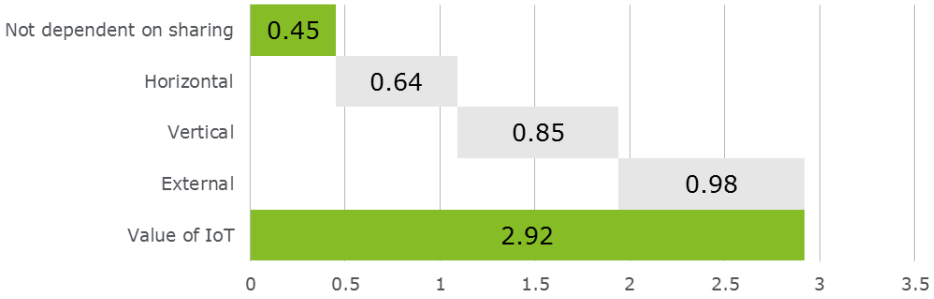
Source: Deloitte analysis. Note the above refers to economic benefits rather than additional revenue.

Figure 4-12: IoT impact on EU resource management and drug counterfeiting in 2027, and shares depending on data sharing (€ billion)



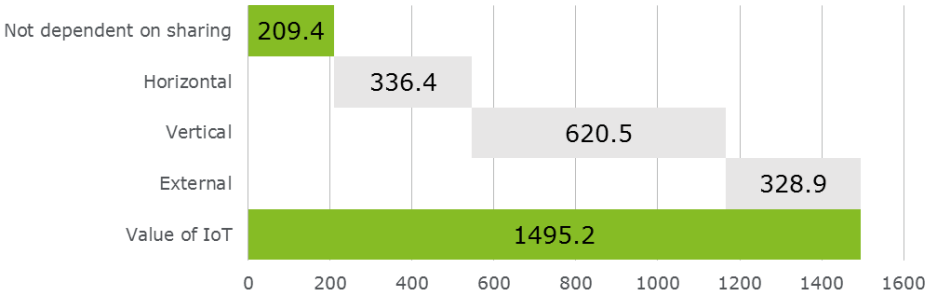
Source: Deloitte analysis. Note the above refers to economic benefits rather than additional revenue.

Figure 4-13: IoT impact on energy consumption in 2027, and shares depending on data sharing (€ billion)



Source: Deloitte analysis. Note the above refers to economic benefits rather than additional revenue.

Figure 4-14: IoT impact on manufacturing productivity in 2027, and shares depending on data sharing (€ billion)



Source: Deloitte analysis. Note the above refers to economic benefits rather than additional revenue.

4.3.6 Comparison with baseline scenario in which data sharing levels remain unchanged

In order to contextualise the added value of future developments in IoT penetration resulting from improved data, it is useful to consider a counterfactual with no improvement in data sharing. This is summarised in Figure 4-15, which isolates the impact of current horizontal, vertical and external barriers to data sharing in each sector.

Figure 4-15: Comparison of data sharing levels remaining unchanged or growing to potential levels

Sector	Potential value with increased data sharing in 2027 (€bn)	Value under current level of data sharing in 2027 (€bn)	Value added (€bn)
Agriculture	35	14	22
Automotive	40	15	25
Healthcare	14	5	9
Smart Cities	2	1	2
Manufacturing	1,300	581	704

Source: Deloitte analysis. Note: figures may not sum due to rounding

In the manufacturing sector alone, greater data sharing could increase output by €704 billion. This reflects the significant impact expected from the industrial IoT.

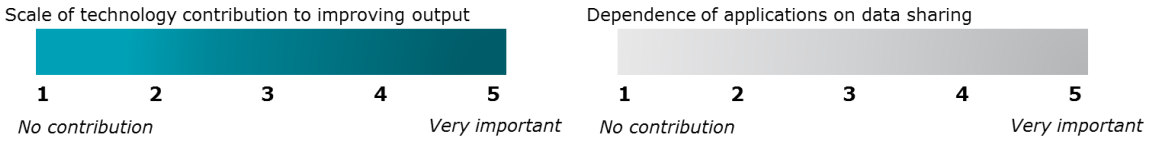
5 The sectoral benefits of machine-generated non-personal data in the chosen sectors

This chapter disaggregates the benefits across each sector and conducts a 'deep dive' into levels of current data sharing.

5.1 Sectoral findings

In this chapter, we present the analysis at a sector level. For each sector, we provide analysis of the level and dependence on data sharing for specific use cases within that sector using the continuum below.

Figure 5-1: Continuum of contribution of technology and its dependence on data sharing



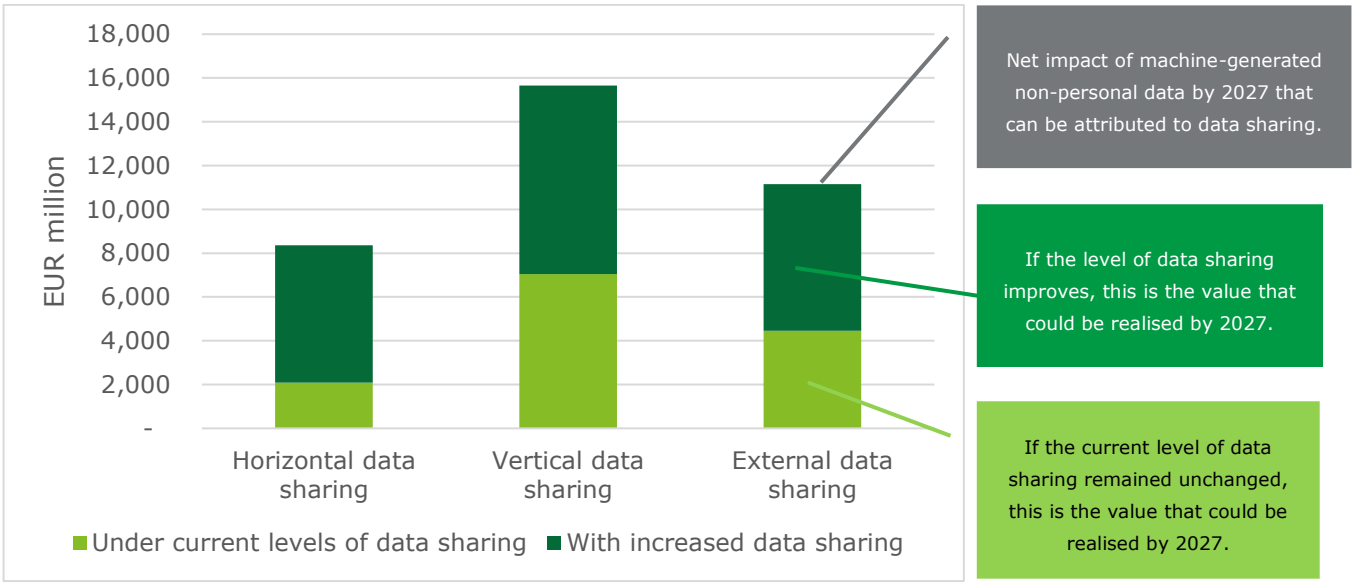
5.2 Agriculture

The most promising uses for machine-generated, non-personal data relate to sensors and remote sensing for crops and livestock and a better analytical understanding of how the farm is operating. However, a diverse set of uses are expected to add value.

The largest component in the expected impact in the agriculture sector relates to vertical data sharing. This reflects the organisational barrier between farms, which own (or lease) and operate agricultural equipment, and the manufacturers of agricultural machines. It may also reflect specialisation in the supply chain, e.g. specialist agricultural research operations in academia or business. External data sharing will also be important to the extent that farmers have an important role in environmental stewardship (increasingly important as the pretext upon which they receive financial support) and otherwise interact with a wide range of other sectors through their land use.

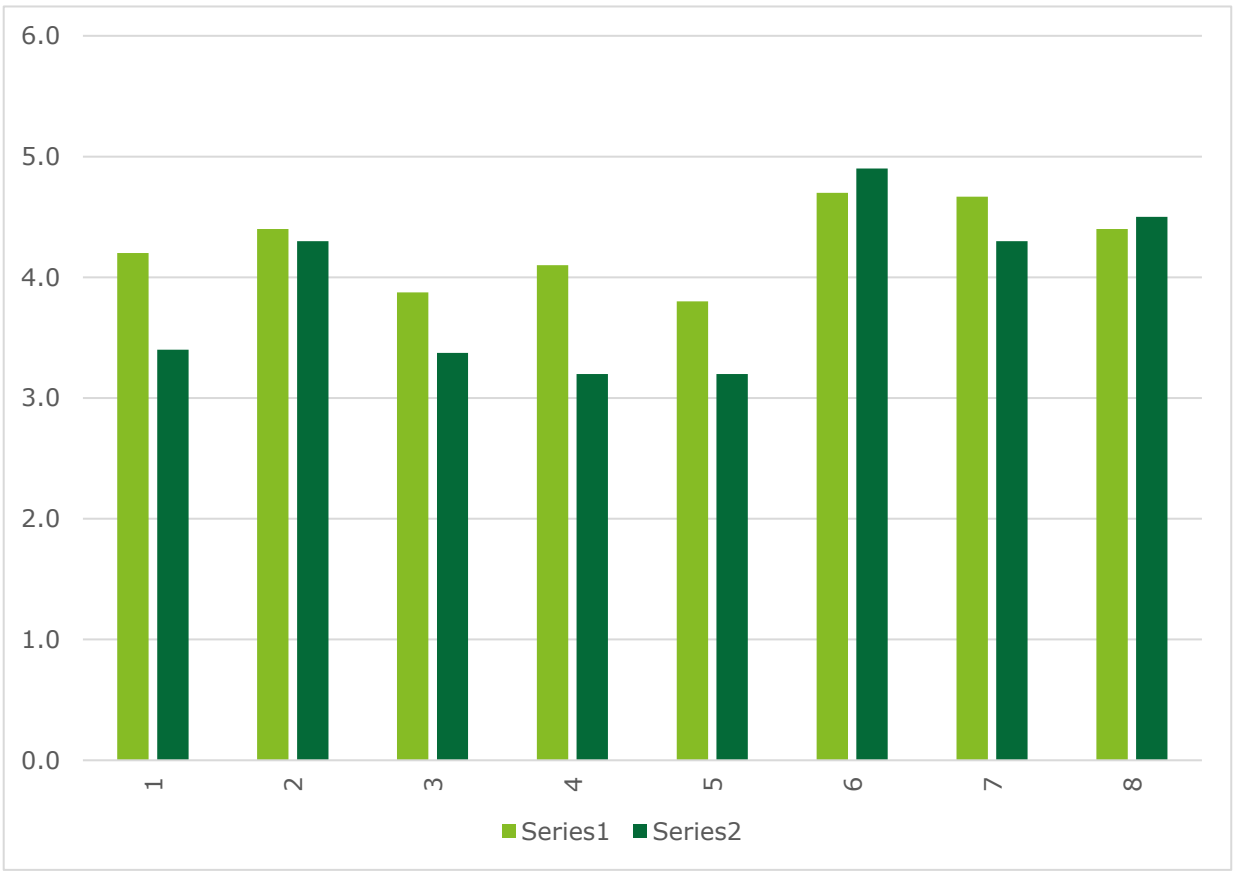
Horizontal data sharing is understood to be of modest importance now but is expected to grow in importance considerably over time. This would represent farms and other agricultural businesses sharing data to gain scale and thereby deepen potential insights.

Figure 5-2: Value of data sharing to agriculture in 2027



Source: Deloitte analysis.

Figure 5-3: Uses for non-personal, machine-generated data in agriculture



Source: Deloitte analysis.

5.3 Automotive

The uses expected to be most important in the automotive sector relate to improving reliability through predictive maintenance and car engine and systems monitoring. Much of the data in the automotive sector can be classed

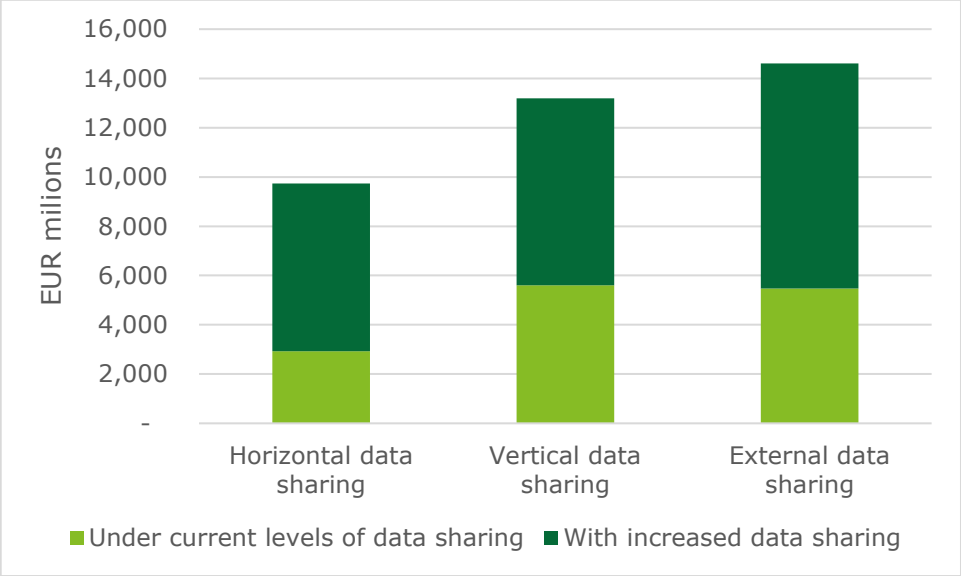
as personal data, and sharing mainly occurs in relation to the manufacturing process (e.g. asset and component performance data with supply chain partners which is generated by engine monitoring sensors). This allows the identification of commonly recurring faults and enables suppliers to improve the design of different components.

External data sharing is not extensive at present, but it is expected to rise once connected cars become the norm, with growing acceptance that this data could be shared externally to improve traffic management systems.

The role of non-personal data in the automotive sector might grow over time with vehicle and ride sharing and as self-driving cars lead to an increasing separation between vehicle use and ownership. The movements of a vehicle could still be linked to a vehicle but perhaps only through data sets held by the fleet owner rather than the manufacturer (although this could be same organisation in some planned business models). If the vehicle is owned by a corporate entity then its use and condition is much less likely to be personal. Real-time fleet management may therefore become steadily more important.

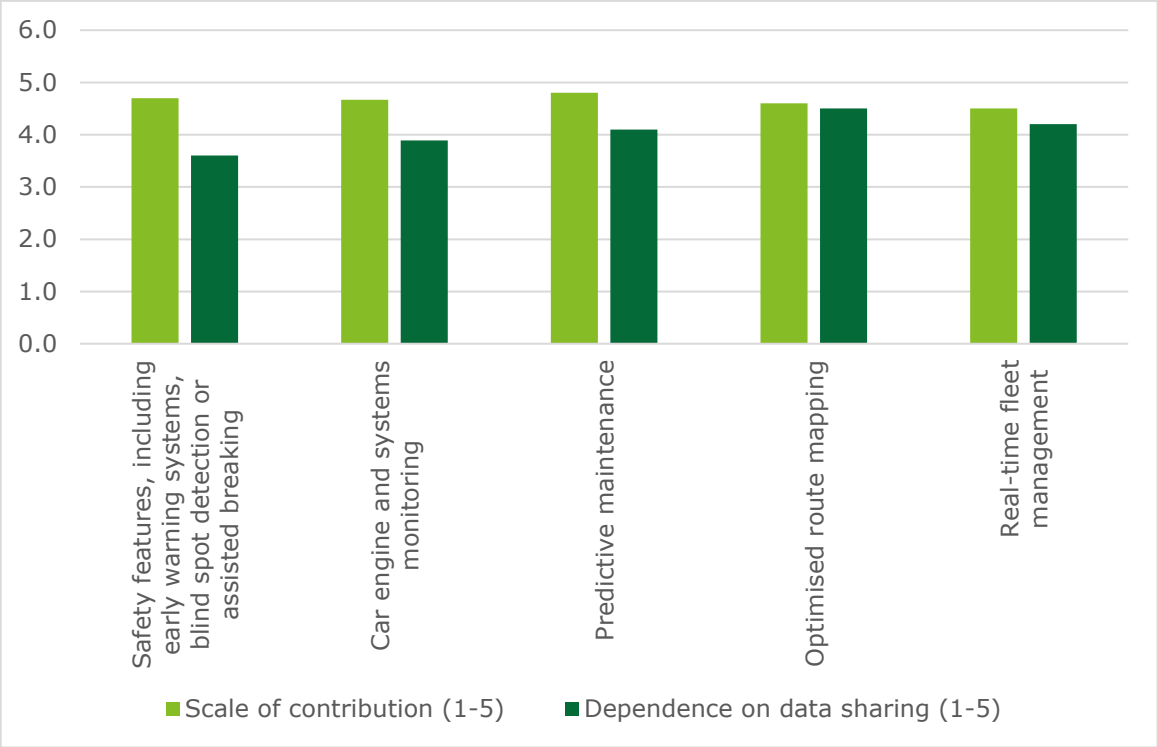
The largest components in the automotive sector are external data sharing. External data sharing has been highlighted as important in both the existing literature and our interviews. Data from cars can improve transport planning. At the same time, there are expected to be a range of service applications working with automotive data. Vertical data sharing is also expected to be extensive and grow considerably, which, in part, will reflect well-developed relationships between manufacturers and dealers and component manufacturers. It may also reflect an expectation that planned initiatives (e.g. the extended vehicle concept) will promote vertical sharing.

Figure 5-4: Value of data sharing to automotive in 2027



Source: Deloitte analysis.

Figure 5-5: Uses for non-personal, machine-generated data in the automotive sector



Source: Deloitte analysis.

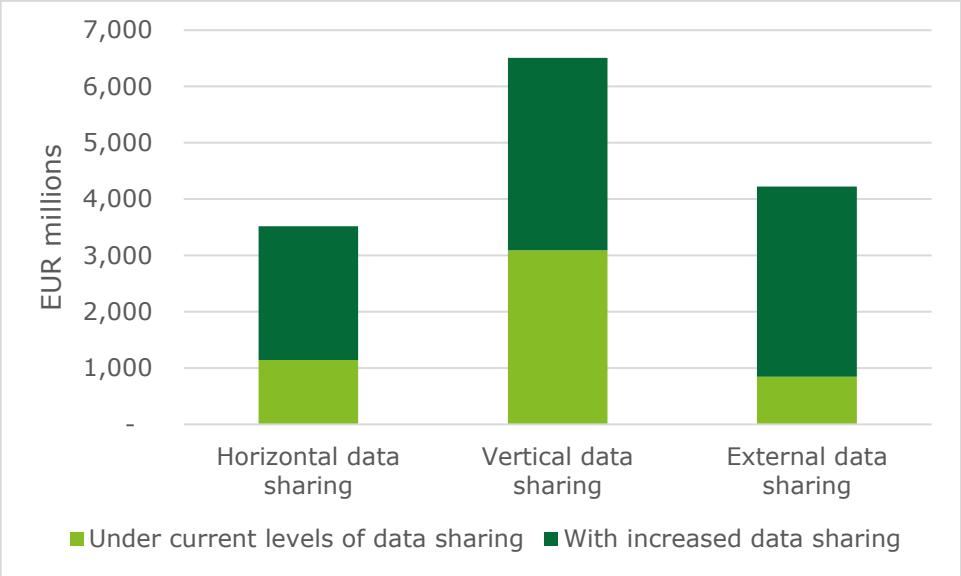
5.4 Healthcare

The main IoT use case in healthcare focuses on the benefits that could be achieved by harnessing patient level data to improve healthcare outcomes, but there are still significant benefits to applying IoT to asset-intensive environments such as hospitals, e.g. optimising hospital parameters such as temperature, humidity and other environmental controls. Maintenance and monitoring of expensive medical equipment (including X-ray, CT, MRI and ultrasound equipment) could significantly reduce downtime. Asset tracking solutions can also improve the efficiency of clinical operations through optimisations for hospital staff.

Vertical data sharing is expected to be particularly important in the healthcare sector. Many applications for horizontal and external data sharing will relate to personal data regarding patients and staff, however non-personal data can be valuably shared within the supply chain. This could allow for optimisation in the management (e.g. maintenance) of devices and other physical assets. Vertical data sharing is therefore particularly important in considering the use of non-personal data.

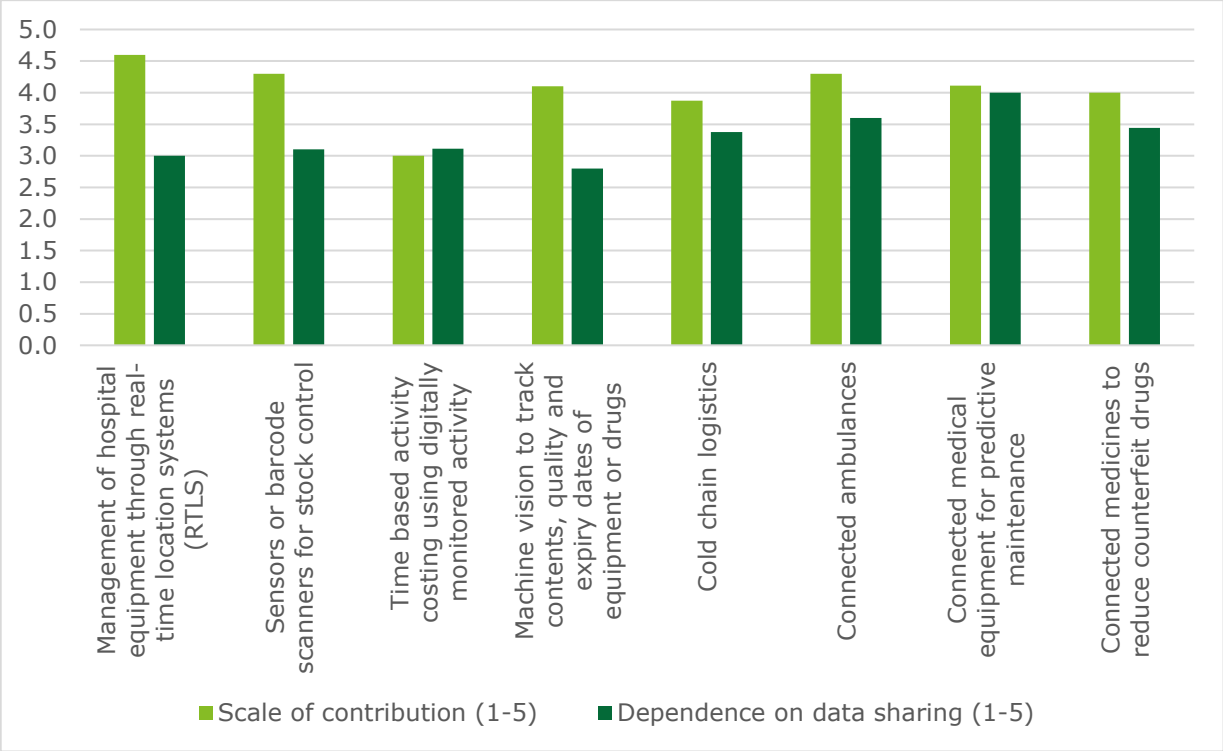
Finally, external data sharing is also expected to grow considerably. This is a heavily regulated sector and accounts for a significant share of public sector spending in EU economies. As the frontline deployment of IoT devices grows, the social interest in engaging with operational healthcare data will be considerable.

Figure 5-6: Value of data sharing to healthcare in 2027



Source: Deloitte analysis.

Figure 5-7: Uses for non-personal, machine-generated data in healthcare



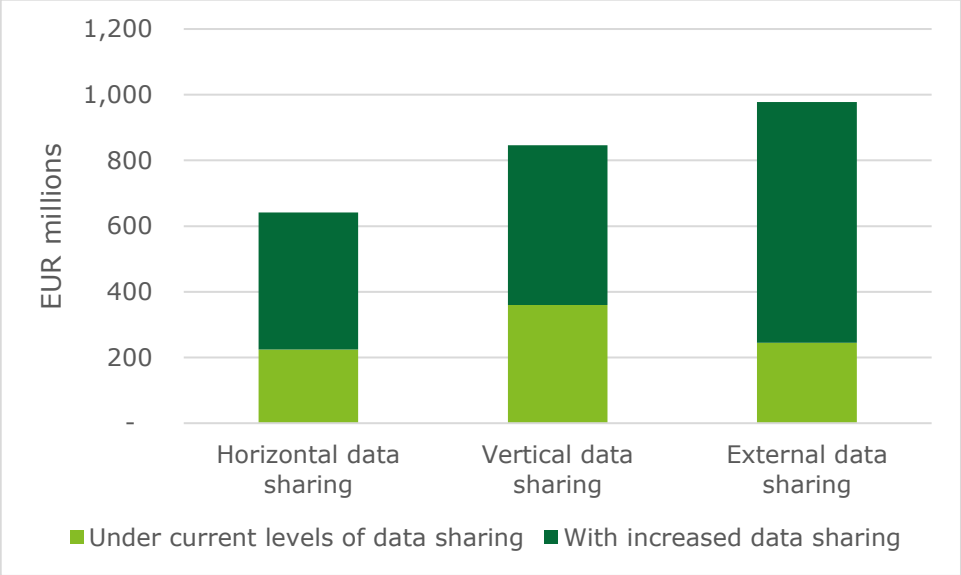
Source: Deloitte analysis.

5.5 Smart Cities

Improvements relating to a connected energy system, which is heavily dependent on data sharing, are expected to be the single most important element in Smart Cities. However, the link to the automotive sector and connected road traffic management is also expected to be significant and depend to a large degree on sharing.

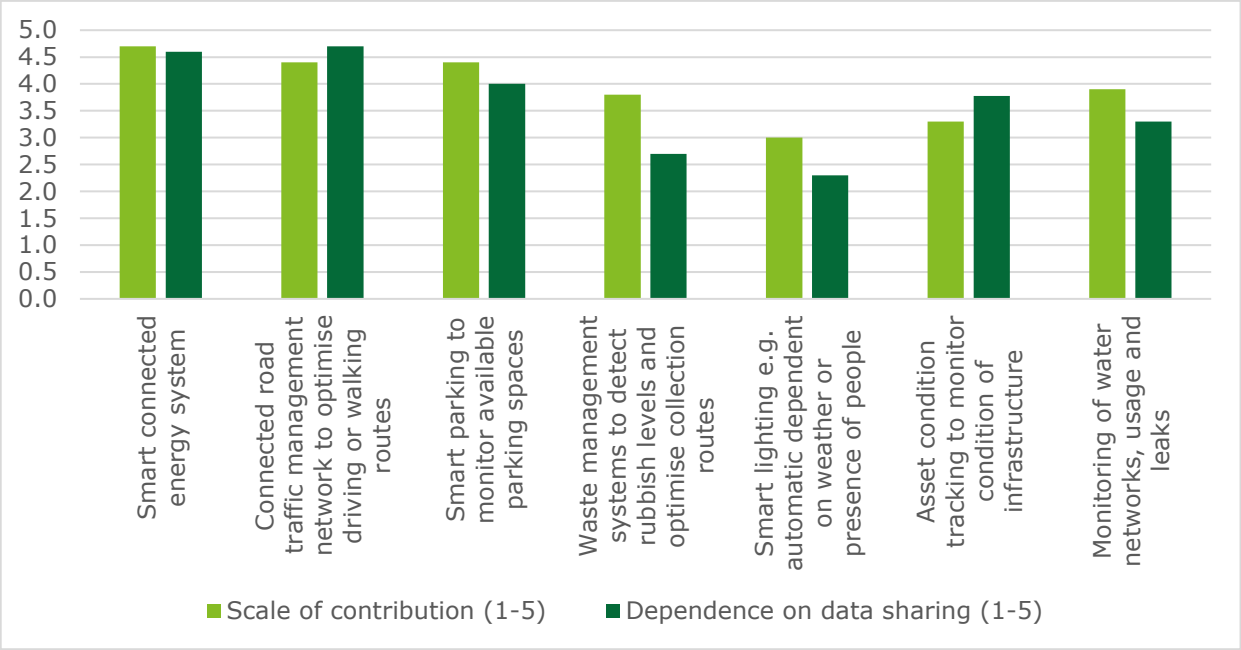
Smart City initiatives promote the sharing, analysis and operational use of data generated by municipal services. That progress can be extended with sharing data outside the sector (including with other sectors studied in this report, e.g. healthcare and automotive). This will open up new opportunities to optimise those municipal services and improve productivity.

Figure 5-8: Value of data sharing to Smart Cities in 2027



Source: Deloitte analysis.

Figure 5-9: Uses for non-personal, machine-generated data in Smart Cities



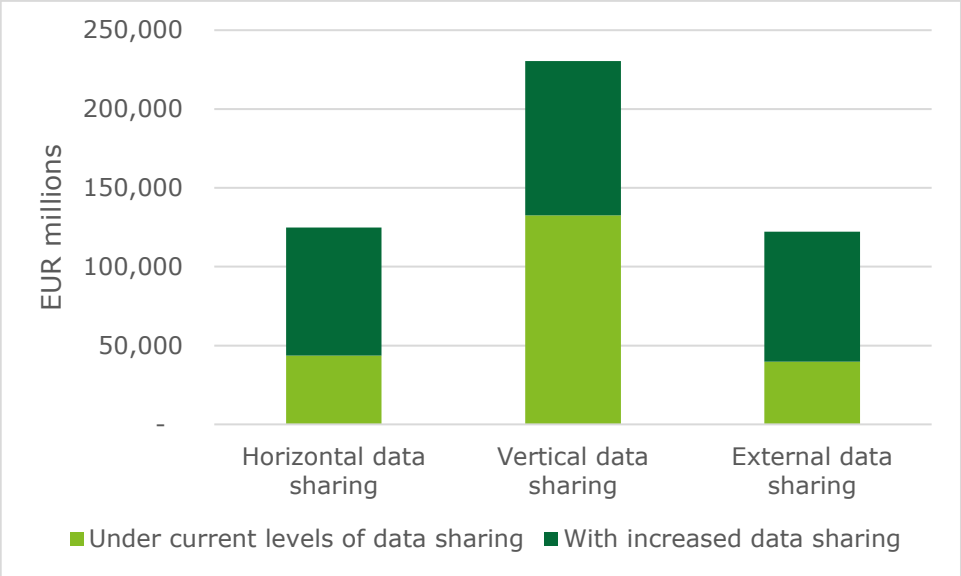
Source: Deloitte analysis.

5.6 Manufacturing

Similar to the healthcare and automotive sectors, the single most important use case for non-personal data is expected to be for predictive maintenance. However, there are a diverse range of other use cases and automated logistics suggest that the logistics sector itself is an important sector that should be considered for further investigation in how it might share IoT data (outside the scope of this study).

In the manufacturing sector, vertical sharing is expected to be particularly important. This reflects the complex supply chains prevalent in the sector, which would be costly to unwind through vertical integration. While commercial concerns might inhibit external or horizontal data sharing, this is less of a concern in vertical relationships where there are more likely to be cooperative norms. The results here therefore suggest that vertical data sharing will be the most important component in the overall growth of data exchange associated with Industry 4.0.

Figure 5-10: Value of data sharing to manufacturing in 2027



Source: Deloitte analysis.

Figure 5-11: Uses for non-personal, machine-generated data in manufacturing

Source: Deloitte analysis.

6 Obstacles to growing the benefits of non-personal data sharing

The analysis in this study has revealed a number of substantive barriers to sharing machine-generated, non-personal data. These barriers differ in scale and impact, but collectively they imply sharing non-personal data is less prevalent than it could be, and that the European economy could lose out on significant benefits.

6.1 Barriers to sharing machine-generated, non-personal data

Chapters 4 and 5 set out the approach to estimating the value of the expected benefits from non-personal data in 2027 across five economic sectors of interest. In particular, chapter 4 explored what share of the economic value derived from the rising adoption of machine-generated, non-personal data for each sector is likely to come from horizontal, vertical and external data sharing.

Figure 6-1: Types of data sharing in the five sectors

Type of data sharing	Healthcare	Manufacturing	Automotive	Smart Cities	Agriculture
Horizontal	33%	35%	30%	35%	25%
Vertical	48%	58%	43%	43%	45%
External	20%	33%	38%	25%	40%

Source: Deloitte analysis

This chapter discusses and quantifies the extent to which a number of barriers prevent the economic value of horizontal, vertical and external data sharing from being realised today and may continue to do so in the future. Industry experts were asked to score the importance of five barriers on the extent to which they currently prohibit horizontal, vertical and external data sharing. This allows for a clearer understanding of the most salient obstacles to data sharing in each sector and enables the estimation of the economic value that could be realised as a result of overcoming these barriers through targeted policy measures that can facilitate data sharing.

This chapter specifically considers how each barrier affects data sharing in the five sectors, i.e. the economic value that could be realised as a result of overcoming each one. The key metric used in this chapter is lost benefit

(in € million) which reflects the monetary value of the benefit that could be foregone if barriers to data sharing persist in 2027.[17]

6.2 Identified overall barriers to sharing machine-generated, non-personal data

The following key barriers to sharing machine-generated, non-personal data have been identified.

Figure 6-2: Barriers to sharing machine-generated, non-personal data

Source: Deloitte analysis.

6.3 Identified barriers within each sector

6.3.1 Agriculture

Commercial barriers are seen as the most important obstacle for agriculture in terms of their effect on reducing non-personal data sharing. Agricultural experts in the survey for this report suggest that commercial barriers are high, as "supply chain participants do not wish to concede competitive advantage by sharing their data". Industry experts also noted that data holders may not be sharing data as they have no direct commercial stake in the profits of organisations using and re-using the data. Another cited example of barriers in agriculture is the "lack of interoperability of data sharing platforms, where data from one source is not easily compatible onto another platform." Although in essence this is a technical barrier, it serves to reinforce commercial barriers by raising costs to organisations of providing data (whether through the costs of needing to invest in software or data sharing platforms, or the time costs of transforming data to a shareable state).

Figure 6-3: Key barriers to sharing non-personal data cited in the agricultural sector

Barrier type	Description
Technical	Lack of communication network infrastructure – agriculture generally takes place in remote areas, which are less well-connected than urban centres.
Technical	Taking full advantage of IoT investment could require analytical capabilities that smaller and more remote farms might struggle to access.
Commercial	Farmers need to invest in IoT sensors and local infrastructure, but lack the financial resources to do so.

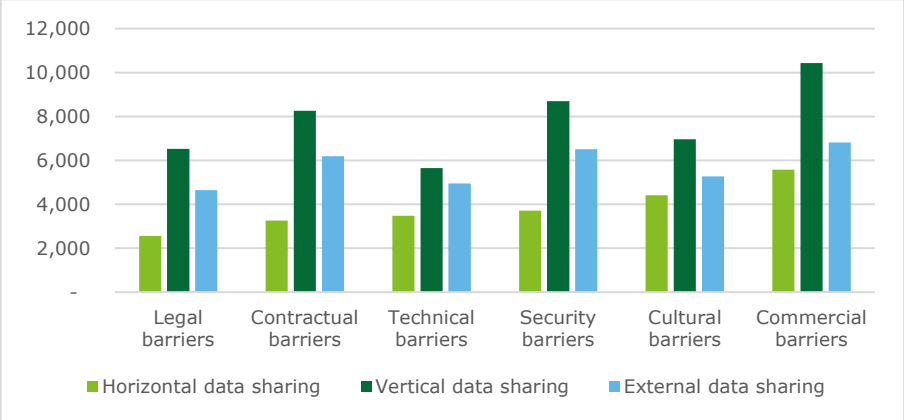
[17] It is important to note that these results are not additive across the vertical, horizontal and external levels, nor are they additive across the type of barriers examined. This is due to the fact that multiple barriers might contribute to preventing the same potential data sharing, and therefore it is difficult to isolate the extent to which an individual barrier prevents data sharing. The estimates presented in this chapter are derived by multiplying the percentages given in Figure 6-1 by the economic values of IoT impact on output in 2027 due to data sharing (disaggregated at the horizontal, vertical and external levels) as established in chapter 4.
Summing the economic values at the horizontal, vertical and external level will therefore not sum to the total. Rather, if the effect of commercial barriers is considered to limit 67% of valuable vertical data sharing in agriculture, this reflects up to €10.4 billion in lost benefits in 2027. However, this value will likely include the effect of any interaction with other barrier types.

Commercial/cultural	The return on investment on the implementation of IoT is not well understood by banks, meaning they will not lend for new systems to share data.
Commercial	Agriculture at the smallholder level, lack scale and are labour-intensive. In these cases, where farm machinery is limited, it is difficult to justify IoT investment.

Source: Deloitte analysis.

The quantified impacts of these barriers, in terms of lost benefits, are estimated below.

Figure 6-4: Most important obstacles to data sharing in agriculture (value in €million)



Source: Deloitte analysis.

Obstacles to vertical data sharing are the most important, accounting for between 31% and 67% of valuable data sharing prevented across the different barrier types. Commercial barriers to vertical data sharing are the largest single component and are therefore a particular focus for policy in our later recommendations, where we recommend adopting concepts from the automotive sector.

6.3.2 Healthcare

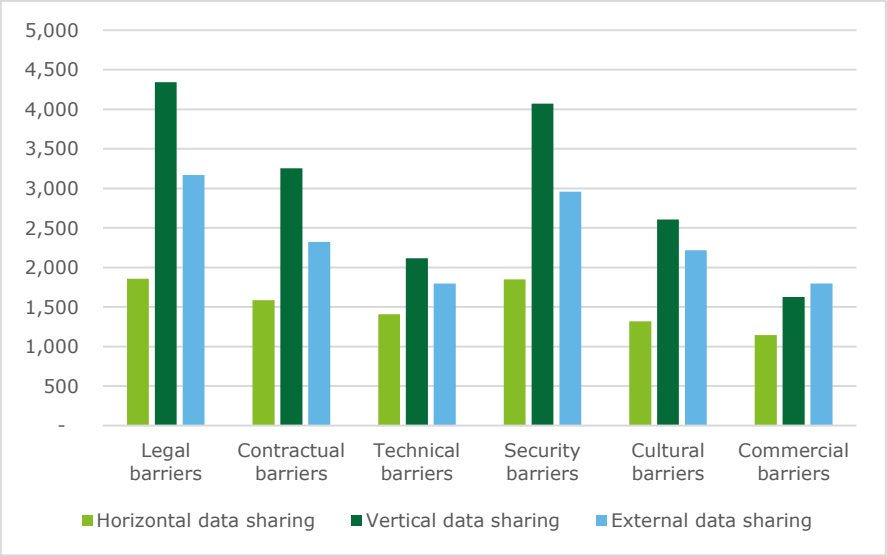
Figure 6-5: Key barriers to sharing non-personal data cited in the healthcare sector

Barrier type	Description
Cultural	Healthcare organisations might have a culture of keeping data close, reflecting the natural caution in organisations used to handling patient data.
Technical/cultural	Data format consistency between manufacturers, hospitals and third party technology companies will require collaboration in an industry that is not used to collaboration.
Legal/contractual	Data ownership will be a major barrier to sharing.
Commercial	Hospitals generate significant amounts of data, but it is costly to collect and store in a way that is usable.
Commercial	Sensors are relatively cheap and can be integrated into new devices at limited cost, but very expensive to retrofit into existing devices.

Source: Deloitte analysis.

The quantified impacts of these barriers, in terms of lost benefits, are estimated below.

Figure 6-6: Most important obstacles to data sharing in healthcare (value in €million)



Source: Deloitte analysis.

The most important barriers are found to relate to vertical data sharing, accounting for around 67% of potentially valuable sharing prevented. Legal barriers (e.g. fines under GDPR and earlier legislation) to vertical data sharing are the largest single component, reflecting concerns that hospitals, in particular, are loathe to share data due to a generally cautious attitude around the legal risks.

Patient-level data is clearly personal, but data about medical assets is not (provided any link between the medical asset and the patient is severed). However, concern about protecting personal data may indirectly lead to non-personal data being tagged as personal or not being used. For example, operational uses of medical devices are often focused on matching patients, staff and assets for healthcare delivery, thereby producing a mixed dataset of personal and non-personal data. Healthcare organisations may hold back on sharing any data from the mixed datasets, citing the concern that this would entail sharing personal data rather than sharing the non-personal data categories within that dataset independently. If these obstacles cannot be overcome, then the non-personal data, which may have independent value, may either be used independently or not used at all.

6.3.3 Manufacturing

Manufacturing in Europe is one of the most mature and advanced in terms of IoT adoption, as manufacturers have been connected to their equipment for a long time. However, recent advances in software and IT infrastructure will allow manufacturers to do more with data and make automated decisions (whilst at present it focuses primarily on identifying discrete operational efficiencies). Security barriers are considered to be the greatest obstacle in the manufacturing sector, though legal and commercial barriers also feature highly. Manufacturing industry experts cite that "concerns around security are definitely very important across the data sharing channels." These issues are particularly pertinent given increasing evidence around the threat posed by cybercrime. Market participants might be concerned that sharing in European economies could be exploited by malicious actors elsewhere if not subject to proper controls.

Legal barriers, including competition rules, also contribute to uncertainty within the sector and dampen data sharing levels. One particular point raised was that often a bespoke data sharing agreement was required per organisation, raising the cost of sharing.

Another issue raised is that of ownership. In the case where a machine vendor collects data, it is unclear whether the data belongs to the machine vendor or the machine operator. Horizontal data sharing is subject to commercial barriers, since there is a common perception that relinquishing production insights to others could undermine a firm's market and competitive advantage. Companies are focused on the downside risk rather than the upside potential to sharing, particularly since the cost of normalising data to be shared is very high.

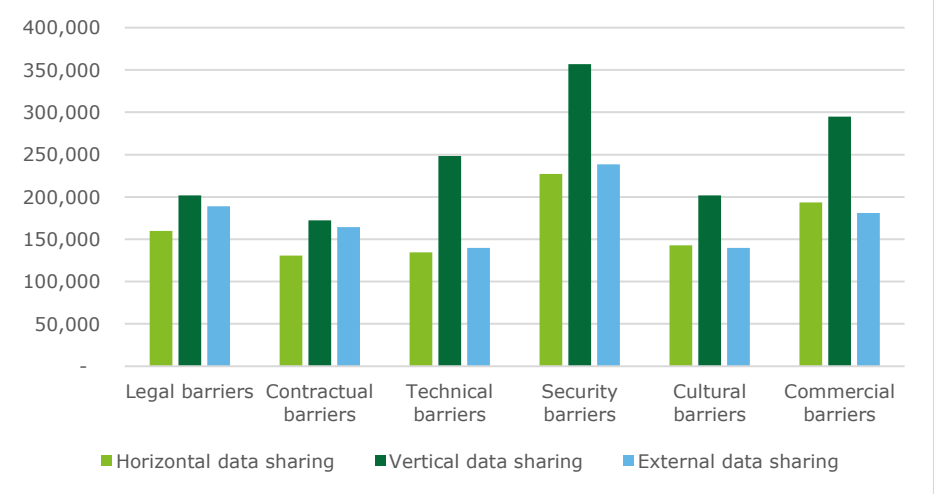
Figure 6-7: Key barriers citied in the manufacturing sector to share non-personal data

Barrier type	Description
Commercial/technical	The cost of normalising data to be shared is high
Legal	Legal procedures will need to replicated for every data-sharing partner, which is time consuming
Technical	Lack of common sharing protocols and standards
Security	Exposing machines to attack and/or inadvertently disclosing commercial secrets.

Source: Deloitte consultation

The quantified impacts of these barriers, in terms of lost benefits, are estimated below.

Figure 6-8: Most important obstacles to data sharing in manufacturing (value in €million)



Source: Deloitte analysis. In reaching these estimates, an assumption was made that manufacturing output will grow at the same rate as real GDP, which sees a steady increase in the value of the production of manufactured goods to 2027. Thus, a change in GDP projected annual growth rate would revise these estimates up or down, and there is likely to be variation between member states.

Barriers to vertical data sharing are again the most important. Though technical barriers are considered to prevent around 40% of valuable data sharing at the horizontal and vertical levels, the monetary value of lost benefit is higher in the latter, due to the relative importance of vertical obstacles in the manufacturing sector. Security barriers are worth approximately €227 million, €357 million and €238 million of lost benefits across horizontal, vertical and external dimensions respectively. Commercial barriers to vertical data sharing are also significant, thought to prevent up to 48% of valuable data sharing.

6.3.4 Automotive

Security is a particular issue in the automotive sector. Companies are concerned specifically about the reputational and financial implications in the event that their vehicles are hacked. Firms are also concerned with ensuring that third parties can protect their data, creating contractual barriers.

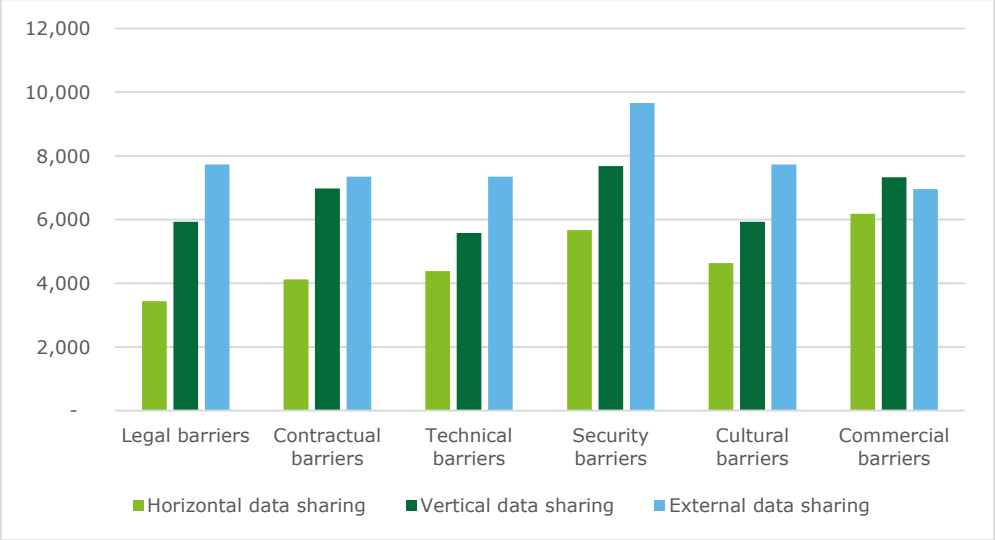
Figure 6-9: Key barriers to sharing non-personal data cited in the automotive sector

Barrier type	Description
Legal	There are extensive legal agreements required to share data vertically and externally and this often needs to be navigated in the context of commercial competition which raises the perceived risks.
Contractual	Similar to legal barriers, in that organisations need to take care in what data to share with supply chain partners that work with competitors.
Technical	There is no consistency in the way different organisations' IT and infrastructure systems are set up and the level of security.
Commercial	Applies to most horizontal data sharing. It is much easier to do with partners and non-competing entities.
Security	Applies to all data sharing and relates to the risk that third parties cannot adequately protect the data.

Source: Deloitte analysis.

The quantified impacts of these barriers, in terms of lost benefits, are estimated below.

Figure 6-10: Most important obstacles to data sharing in automotives (value in €million)



Source: Deloitte analysis.

In the automotive sector, barriers to external data sharing are the most important, accounting for between 45% and 63% of valuable data sharing prevented across the barrier types. Security barriers to external data sharing are the largest single component.

Legal and cultural barriers to external data sharing are also significant. This is partly a result of security barriers, with concerns over liability should anything go wrong, but also reflects the grey area between personal and non-personal data in this sector. Vehicle data is not in itself personal, relating to the car rather than a living natural person, but it becomes personal because a unique vehicle identifier allows it to be linked to a person (the owner). In the four data categories described by the VDA (German automotive manufactures' association), "personal data" is distinguished by not being anonymised. It is therefore only expected to be shared subject to controls for privacy. The same dataset could be personal if it includes fields which link the vehicle to the owner (i.e. the Vehicle Identification Number) and, without that field, non-personal. This is very different to the inherently non-personal datasets considered elsewhere in this report.

The value of benefits lost to commercial barriers are relatively high across all data sharing dimensions, reflecting an OEM preference to hold on to data noted by many stakeholders surveyed. Initiatives discussed later are in place to overcome this obstacle.

6.3.5 Smart Cities

The most developed thinking around external data sharing is occurring in Smart Cities. This is where external sharing is most important and therefore the most thought has gone into the regulatory model (e.g. in considering permits for data collection in public spaces, which are connected to an outright requirement for sharing). This sector will therefore be a potential source of best practice when it comes to policy initiatives to share data.

This notwithstanding, there remain some barriers to increased data sharing. Commercial, legal and security barriers are viewed as the most important obstacles in Smart Cities. Commercial arrangements that suit all parties involved can be challenging and there are generally more legal constraints placed upon contractual interactions between public bodies and third parties (e.g. more stringent procurement rules and judicial and political oversight).

As with sharing between businesses, there are particular security challenges which arise when data is shared outside the city. Other cities may have different norms, e.g. approaches may differ between those using cloud or on-premises IT infrastructure. There may also be large cultural differences between cities and companies, creating barriers to external data sharing. Cultural barriers are expected to diminish as markets mature, however, and municipal data sharing becomes more mainstream.

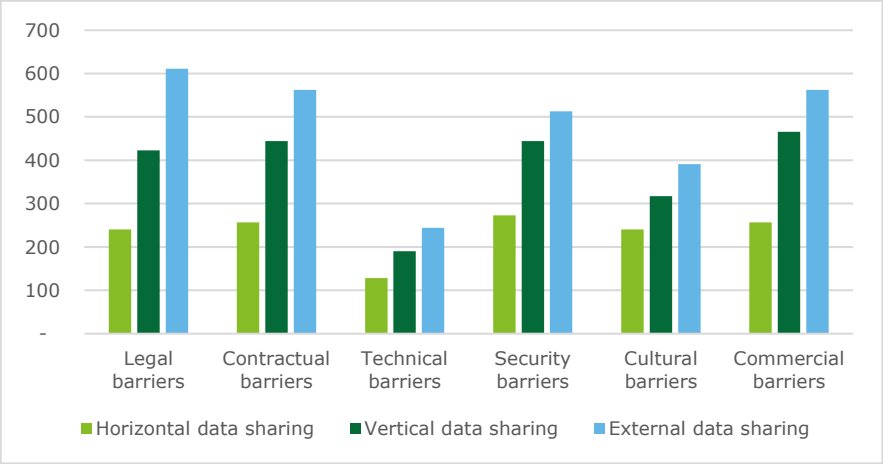
Figure 6-11: Key barriers to sharing non-personal data cited in Smart Cities

Barrier type	Description
Legal	This is a barrier in vertical and external data sharing: it is easier to share data within a city than outside because of the complicated legal procedures required to share data with third parties
Technical/cultural	Smaller cities are more focused on collaboration, and can act much faster to solve problems in comparison to bigger cities that operate on a very different scale
Cultural	There are significant cultural differences between cities of differing size, and also between cities and companies, in terms of organisation and motives.
Security	Disparities in security norms between different cities or between cities and corporates may create security risks.
Commercial	Different entities may find it challenging to create commercial arrangements that suit all parties.

Source: Deloitte analysis.

The quantified impacts of these barriers, in terms of lost benefits, are estimated below.

Figure 6-12: Most important obstacles to data sharing in Smart Cities (value in €million)



Source: Deloitte analysis.

In Smart Cities, the obstacles to external data sharing are the most important. This is particularly pronounced on the legal, contractual and commercial barriers. Vertical data sharing also faces considerable lost benefits across a number of barrier types, with several barriers preventing between 50% and 55% of potentially valuable sharing. Security barriers are the most important obstacles to horizontal data sharing, estimated to prevent up to 43% of potentially valuable data sharing.

It should be noted that these estimated values are based on a narrow view of some of the applications of Smart Cities. The range of smart city initiatives is extensive and growing over time. This analysis only considers street lighting. To the extent more use cases are considered, the estimated value of data sharing in Smart Cities will grow.

6.4 Observations on barriers

The overall conclusions below summarise the size and extent of barriers to sharing machine-generated, non-personal data.

Figure 6-13: Conclusions on barriers to data sharing

- Sharing across supply chains (vertical data sharing) yields some of the largest benefits from machine-generated, non-personal data, but this type of sharing also faces some of the most significant barriers including legal and security obstacles.
- Sharing between supply chains and third parties (external data sharing) is also constrained by commercial impediments such as the fear of losing a competitive advantage.
- Technical and cultural barriers are not thought to be preventing potential data sharing on a large scale, but they do exacerbate other barriers, e.g. amplifying legal, security and commercial barriers.
- Barriers vary between sectors. In the manufacturing sector, security is the most important barrier to sharing, whereas in healthcare the legal barriers are more significant.
- Legal barriers to data sharing result, in part, from operational challenges in implementing the concepts in GDPR consistently.
- Infrastructure, in the widest sense covering technologies, processes, assets and organisational structures, matters. The volumes of data being shared are considerable and infrastructure can be a more

fundamental technical and process challenge even than standardisation (where APIs offer an increasingly well-understood means to overcome problems at organisational boundaries). For instance, in the agricultural case where connected farms require affordable and reliable connectivity and infrastructure in rural areas to be able to benefit from these technologies. Therefore considering only the IoT devices will neglect the need to think about new processes, financing and ways of working.

Source: Deloitte analysis.

6.5 Will market initiatives resolve these barriers?

Before pursuing any policy intervention, it is important to consider whether there might be partial or complete market solutions that would overcome barriers to data sharing. Based on examples from other industries and analogous technologies, such solutions could take several forms:

- **Existing organisations might act as intermediaries to facilitate and smooth data sharing.** Such organisations could provide assurance on security and commercial impartiality with these intermediaries as common suppliers or customers. As noted in Chapter 3, Semiotic Labs is an example of this and acts as an intermediary between different operators of electric motors. Trade associations also often have an important role in defining technical standards or commercial norms.

- **New dedicated intermediaries might emerge.** The framework being put in place for the automotive sector anticipates this possibility, with neutral servers planned to aggregate data across multiple automotive brands. There is a market opportunity for new businesses to emerge that fulfil a similar role, enable new value and are compensated for doing so.

- **Organisations might merge to internalise the externality of low data sharing**. This is particularly relevant given that the greatest opportunities in several sectors relate to vertical data sharing. The development of IoT may have shifted the incentive towards greater vertical integration in manufacturing supply chains. This would fit with wider economic experience where "the technology intensity of downstream (producer) industries is positively correlated with the likelihood of integration whereas the intensity of upstream (supplier) industries is negatively correlated with it."[18]

Even businesses that currently envisage competing based on the data they generate might become more inclined to share more data over time. Smaller players who generate smaller amounts of data will face an incentive to cooperate and share with other smaller players in order to compete with larger players (who start with the most potential to generate data). This will naturally address the cases in which data sharing can make the greatest difference: where there is the greatest difference between the power of the data an organisation can generate itself and the potential with full sharing. The propensity to share and value of data sharing is therefore greatest among smaller players in a market.

Over time, unless the organisation holding data has pre-existing market power, attempting to establish market power by failing to share data will only lessen the ability of their product to compete with alternatives made by rivals more willing to share data. An analogy would be smart phones competing on the availability of third party apps.

Policymakers should not underestimate the potential for corporate and entrepreneurial action to address barriers to data sharing. However, market-based solutions are generally the most likely to prove ineffective in settings

[18] Acemoglu, D., Griffith, R., Aghion, P. & Zilibotti, F. (2010) Vertical Integration and Technology: Theory and Evidence, *Journal of the European Economic Association*, 8, 5, pp. 989-1033.

where transaction costs are high (i.e. where it is hard to come to an agreement which reflects market incentives). There is clearly the potential for this to be the case for data sharing, particularly because:

1. The consequences of a security breach can be high and it can be costly to establish that organisations sharing or receiving data are following proper security procedures.

2. The value of data can be hard to quantify and therefore it might be difficult to come to a price that both sides regard as acceptable.

Crucially, these transaction costs might diminish the volume of data sharing and affect the distribution of data sharing, biasing sharing towards larger and better-understood opportunities. This might lead to an under-sharing of data particularly for SMEs and innovative use cases. Existing and potential policy might achieve disproportionate results by mitigating these transaction costs and thereby catalysing market solutions.

7 Policy options

The Commission has intervened to address barriers to sharing, but further action could address the most salient barriers.

As noted earlier, work already in progress by the Commission represents a helpful market-focused approach to enablement of data sharing. However further action would be beneficial. As this chapter will discuss, these extensions to the Commission's approach fall into three broad categories:

- Development of clear principles for the circumstances in which data sharing should be encouraged.
- Targeted policy measures that can facilitate data sharing and reflect those principles.
- Developing and publicising practical models for data sharing.

This chapter briefly discusses the rational for intervention, considers the existing approach by the Commission and makes recommendations on how this could be extended and developed further.

Given our analysis that the most important barriers relate to security and commercial considerations, this is where the interventions proposed in this report focus.

7.1 Rationale for intervention

As this report has shown, the sharing of machine-generated, non-personal data has the potential to generate substantial benefits across European economies and societies. Indeed, the benefits of increasing access to data are increasingly being understood as a mechanism to promote economic growth and productivity:

- **European companies will find it easier to compete for labour and capital.** If obstacles to data sharing are addressed, European manufacturers, for example, might find themselves better able to compete with international rivals in intellectual property creation, and deployment/maximisation of new technologies.
- **More new use cases will emerge.** The technological potential of data sharing is still being understood, but as our analysis demonstrates, many new use cases are emerging. Data sharing has the potential to address perennial policy and business concerns.
- **Wider public policy goals can be achieved at lower cost.** For instance, increased healthcare productivity lowers the cost of delivering a given standard of healthcare and increased agricultural productivity reduces the cost of support for rural economies.

The previous chapter has considered the levels of data sharing currently without dedicated interventions by policymakers. The analysis suggests there exists the risk of a market failure emerging whereby data holders lack incentives to share or only do so in narrow situations, leading to a sub-optimal level of data sharing that fails to achieve the full benefits of data in the EU28. Nascent technologies may not mature and Europe could fall behind the curve in the adoption of new technologies.

For these reasons, there is a rationale for policymakers to intervene in a balanced and appropriate manner to address the market failure of too low levels of sharing machine-generated, non-personal data. It is important this

intervention is balanced – too heavy an intervention may stifle innovation; too light may not adequately protect justified commercial interests. Any intervention needs to be complementary to existing regulations, and where possible avoid creating bespoke regulations for specific industries or sectors.

Below we briefly review existing policy interventions aimed at promoting data sharing.

7.2 Existing EU policy

Policymakers recognise the market failure related to data sharing and have begun to address this. In January 2017, the Commission adopted a Communication[19] on "Building a European Data Economy" in which it looked at potential blockages to the free movement of data. It set out its plans to engage with Member States on addressing the issue of access to machine-generated, non-personal data. After consultation, the Commission then adopted a second Communication in April 2018: "Towards a common European data space".[20] This included:

- a proposal to review the Directive on the re-use of public sector information;
- an update of an earlier Recommendation on access to and preservation of scientific information; and
- guidance on sharing private sector data.

Public sector data sharing can be material in the healthcare and Smart Cities sector. It can also act as an example to overcome cultural aversions to sharing. The proposed measures would build on existing policy encouraging data sharing (Directive 2003/98/EC) and focus on improving access, e.g. with lower charges and increased use of machine-readable API interfaces for data access. This might make it easier for the data to be used in conjunction with machine-generated data (e.g. by combining machine-generated data and public sector data released in an API format to support systems with minimal human involvement).

Access to and preservation of scientific information is particularly relevant in the healthcare sector. The data is generally not likely to be machine-generated, and therefore falls outside the scope of this report, but scientific data could support advances that affect healthcare productivity alongside the operational use of IoT applications covered in this report.

In terms of the guidance on sharing private sector data, the Commission notes that stakeholders regarded the evolution of the sector as being at too early a stage for broad ex ante regulation to be appropriate. The Commission instead recommends principles for contracting in data sharing, leaving the question of whether to contract to business. Those principles are:

- **Transparency** over who will have access to what data and the purposes for using such data.
- **Shared value creation** with a recognition that where data is generated as a by-product of using a product or service, several parties have contributed to creating the data.
- **Respect for each other's commercial interests**.
- **Ensure undistorted competition** when exchanging commercially sensitive data.

[19] Source: https://ec.europa.eu/digital-single-market/en/news/communication-building-european-data-economy, accessed 3 May 2018.
[20] Source: https://ec.europa.eu/digital-single-market/en/news/communication-towards-common-european-data-space, accessed 3 May 2018.

- **Minimised data lock-in** with companies offering a product or service that generates data enabling data portability as far as possible. This element matches the recent initiatives on portability for data between cloud IT infrastructure providers.

Besides this, the Commission announced a number of supportive measures including assistance from the Support Centre for data sharing under the Connecting Europe Facility programme; fostering the use of APIs; and facilitating specific tests and demonstrations (e.g. the deployment of connected and autonomous mobility on digital cross-border corridors).

The Communication did not announce discrete actions on business data sharing with government. However, it did outline a set of principles for future policy, including:

- Proportionality in the use of private sector data.
- Purpose limitation (e.g. a limited horizon for the use of data shared).
- A 'do no harm' principle including the protection of commercial secrets.
- Conditions for data re-use with compensation reflecting any link to public interest goals (giving public sector bodies preferential treatment).
- Actions to mitigate limitations in private sector data (rather than simply require businesses do so).
- Transparency and societal participation.

7.3 Existing national and local policy

Given that the data sharing market failures also occur at national and local level, policy is not only being considered or implemented at the European level. National, regional and city governments are also seeking to require or facilitate data sharing. Interventions include:

- **Economic regulators requiring the sharing of data with regulators and other bodies.** For example, information about energy generation is often shared by regulators beyond the immediate requirements of sector regulation; healthcare regulators have been steadily increasing the amount of data that is being shared (generally focused on anonymised patient data to support clinical research); and extensive data is shared by property registries about the ownership of residential and commercial property. This interacts with European requirements, e.g. the requirement for banks to share data under the Revised Payment Service Directive and the initiatives on the sharing of scientific data anticipated in the Communication on data sharing in April 2018.
- **Local and regional authorities enacting public policy to require data to be shared.** This can include sharing their own data. There is a large volume of data from regional and local authorities available through the European Data Portal.[21] It can also include requiring sharing by commercial third parties, e.g. there are proposals in Amsterdam to require the sharing of data collected in a public space as part of the licence conditions to do so.

7.4 Report recommendations

The above policy interventions are a positive step. However, the nascent form of the IoT means that any assessment of the adequacy of current plans at the EU institutions and among national and local government is

[21] Available here: https://www.europeandataportal.eu/

necessarily provisional. Our discussions with experts and analysis of levels of data sharing and the barriers to sharing suggest these interventions could go further, particularly to address commercial and security barriers. This suggests two potential gaps in the Commission's current approach:

- It does not directly address security concerns, although some measures might help indirectly and some existing initiatives (e.g. neutral servers) are designed in a fashion intended to maintain security while opening up data beyond those organisations that have a direct commercial relationship with the OEM.

- Its consideration addresses the ideal form for contracting, but does not directly address the barriers to firms establishing a contract (i.e. the transactional costs in doing so). This raises the possibility of both a suboptimal volume of data sharing and an effective bias against SMEs and innovative use cases.

On this basis, this report concludes that to address these particularly important barriers it will be necessary to:

- Better articulate principles for the circumstances in which data sharing should be encouraged.

- Develop targeted policy measures that can facilitate data sharing and reflect those principles.

- Promote appropriate data sharing models across sectors.

This applies at the European, national and local levels.

7.5 Principles for data sharing

This report has identified substantial economic benefits that might result from enabling the sharing of machine-generated, non-personal data. However, there are potential risks including inadvertent disclosure of commercial secrets; inhibiting competition (even if data sharing often promotes competition); and infringing privacy (to the extent that the distinction between personal and non-personal data is not clear and consistent at an operational level). This suggests having an effective and comprehensive set of principles on data sharing that could help organisations to better understand the nature and severity of risks and ways to mitigate them, allowing those organisations to gain from the benefits of data sharing.

There is already a body of work that has considered the right principles to govern data sharing in some dimensions. The GDPR reflects an extensive engagement around the privacy implications of companies using personal data. The competition implications have also been considered by regulators internationally. While the principles of what data can and should be shared will likely evolve, policymakers could consider the following broad framework.[22]

7.5.1 The sharing of machine-generated, non-personal data creates economic benefits

To the extent that economic benefits are generated, this is generally likely to create a common interest in going ahead. Either the organisation sharing can be compensated, or they will also benefit from a strengthening of the ecosystem in which they are operating (this will be particularly likely in the context of vertical sharing).

7.5.2 The sharing of machine-generated, non-personal data does not reveal commercial secrets

To the extent that the data contains intellectual property, sharing might, in some cases, legitimately impair the commercial interests of the company sharing. While machine-generated data is unlikely to contain intellectual property directly, it might reveal commercial secrets inadvertently, (e.g. by revealing novel ways that a machine reacts to certain environmental conditions). Ensuring the protection of intellectual property and other commercial secrets will be important in maintaining incentives for organisations to bear the costs and risks associated with

[22] Note that these principles by design cover both how data is shared and the consequences of that sharing.

IoT innovation. This notwithstanding, policymakers should enact this principle carefully to avoid an excessive caution in sharing data.

7.5.3 The sharing of machine-generated, non-personal data is not related to pricing data

Machine-generated, non-personal data is typically not about pricing. However, when it is, the sharing of pricing data needs to be subject to more stringent controls to prevent collusion that undermines market competition. This is reflected in established practice in the EU[23] and internationally.[24]

7.5.4 The sharing of machine-generated, non-personal data does not necessarily involve access to IT systems operated by the sharer

Access to the IT systems of the data sharer can raise additional security risks. These can be avoided by (a) working with archive data; and (b) working through intermediaries (e.g. the neutral servers anticipated in the automotive setting) or APIs. Further, in many cases, having access to the raw data from IT systems may not be appropriate or necessary. For instance, real-time or raw data sharing might generally be unnecessary outside of specific use cases where time is of the essence (e.g. alerting emergency services to crashes on the roads).

7.5.5 The sharing of machine-generated, non-personal data means including an agreement not to use the data to identify individuals

The focus here is on data which is inherently non-personal. Nonetheless, privacy concerns and risks of GDPR breaches might occur either where non-personal data is used in combination with personal data or where non-personal data is the product of the anonymisation of personal data. Any resulting risks can be diminished to the extent data (a) is anonymised through recognised tools intended for that purpose; and/or (b) the likelihood of anyone being able to combine it with a dataset that would allow them to identify an individual is very low.

7.5.6 The sharing of machine-generated, non-personal data is subject to common norms around how data is shared and processed

Many risks of data sharing can be greatly diminished if the sharing organisation and the recipient have agreed protocols for a) transmission; b) secure storage; c) appropriate usage of data; d) destruction of data; and e) rules around transmission to third parties.

7.5.7 The sharing of machine-generated, non-personal data is part of medium-to-long-term commitment

In many cases, the benefits of the data will emerge over time as the data is better understood and extended series start to emerge.

7.6 Targeted policy measures that can facilitate data sharing

There are a number of steps that could be considered by policymakers and regulators to facilitate enhanced levels of data sharing, in particular:

1. A voluntary accreditation scheme or kite mark for data sharers and recipients.

[23] Source: https://www.lexology.com/library/detail.aspx?g=a4ab9a67-ddfa-4483-bf20-629269790791, accessed 3 May 2018.
[24] Source: https://www.ftc.gov/sites/default/files/attachments/us-submissions-oecd-and-other-international-competition-fora/1010informationexchanges.pdf, accessed 3 May 2018.

2. Further regulatory guidance on appropriate data sharing addressing some remaining sources of uncertainty for organisations considering data sharing.

7.6.1 Accreditation scheme

Trust is at the heart of data sharing and an accreditation system is one approach that could support this. This would be an accreditation for businesses planning to provide or receive shared data analogous to the identity verification schemes for individuals being created to facilitate e-government, e.g. the online ID service launched by the Government of Estonia, or private sector equivalents such as the Mobile Connect platform offered by the GSMA. It provides a central approval for those able to interact with a range of other market participants (versus each one having to establish separate verification systems, multiplying the administrative burden for participants).

The accreditation should:

- Be voluntary. It is likely that different forms of accreditation would likely be needed to reflect industry and sector differences in data sensitivity, requirements and so forth.
- Provide clarity on the conditions on which data is shared and attached to its use.
- Provide a mechanism for monitoring on the processes used by the recipient (and to some extent the sharer) consistent with the above principles, e.g. that the accredited organisation meets existing ISO standards for IT security (e.g. the relatively new 27017 for cloud IT security).
- Develop common technical standards on:

  (a) state of the art anonymization – at which point data is considered "non-personal data";

  (b) start of the art pseudonymisation – at which point personal data is considered pseudonymised in accordance with GDPR; and

  (c) common understandings of qualitative tests established in legislation or in case law (for example the "means reasonably likely to be used" legal test).
- Be endorsed and/or authorised by relevant regulators to promote confidence and uptake.

7.6.2 Regulatory guidance

A clear message from this research has been the concerns of stakeholders over potential regulatory barriers to the sharing of data. The proposed accreditation scheme is designed to assist with this but it is also recommended that:

- Further guidance from regulators on broader regulatory/legal concerns could help clarify and enable sharing. Examples include (a) setting out when anti-trust concerns may arise due to data sharing; and (b) the rights of ownership of data; and (c) other relevant factors under the principles above. In order to avoid complicating existing regulatory regimes, this could focus on creating 'safety zones' or 'safe harbours' in which data sharing can generally be expected to be approved (a concept adopted by the US Federal Trade Commission for competition purposes).[25]
- The accreditation schemes could provide regulators with an arbitration role in the event of disputes, e.g. over breaches of long-term commitments to share data.

[25] Source: https://www.ftc.gov/sites/default/files/attachments/us-submissions-oecd-and-other-international-competition-fora/1010informationexchanges.pdf, accessed 3 May 2018.

7.7 Promoting sharing model use cases

In addition to a supportive regulatory framework, there is also a need for a broader understanding of the potential for and benefits of data sharing. In our engagement to support the delivery of this report, we found that even sector experts were often unaware of potential models that might support sharing in their sectors. The Commission Staff Working Paper released alongside the Communication in April 2018 made a start on this process by detailing a number of formats for data sharing (e.g. an Open Data approach; a data marketplace approach; and data exchanges through a closed platform) and providing illustrative examples.[26]

There are also a range of sharing model use cases that companies, regulators and policymakers might explore depending on the institutional context in the sector concerned. Those institutional contexts might include, for example:

- Sectors that are generally competitive, but where data is concentrated in one part of the supply chain. In sectors where this is the case (e.g. certain markets for manufactured goods) the extended vehicle and neutral server models that the automotive sector is adopting could facilitate increased sharing. The expectation is that, over time, the market incentives for appropriate sharing will become apparent and therefore the policy requirement is for institutions that enable such sharing rather than compulsion.

- Sectors in which the public sector is a principal customer. The Staff Working Paper set out a number of models for business sharing data with government, including (a) donation (as a form of corporate social responsibility); (b) public sector prizes for solutions to specific social challenges; (c) mutually-beneficial partnerships; and (d) the creation of intermediaries to overcome a lack of trust, which might manifest itself as a security, commercial or cultural barrier to data sharing. Policymakers could extend this support for sharing with requirements for sharing as a condition in public sector commercial terms. This could be effective where there is a procurement relationship and the procuring body expects the provider to generate substantial operational data. Such terms could also require sharing with third parties.

- Sectors in which data gathering depends upon drawing on public goods, e.g. collecting data in public spaces. In this case, municipal authorities might legitimately require data sharing as a condition of allowing such collection to take place.

These and other use cases could be explored to enhance data sharing, in line with the principles outlined earlier, and overcome the barriers identified earlier in this report.

[26] Source: https://ec.europa.eu/digital-single-market/en/news/staff-working-document-guidance-sharing-private-sector-data-european-data-economy, accessed 3 May 2018.

Annex: Model approach

A.1 Data collection

The quantitative model used is informed by baseline data obtained through publicly available datasets and an expert survey. The sector-specific insights obtained through the survey are combined with quantitative analysis. The resulting findings quantify the economic value of barriers to machine-generated, non-personal data in Europe.

A.1.1 Baseline data

The full range of uses to which new sources of data can be put will only become clear with time as innovators in and outside each sector explore them. Nonetheless, this analysis uses publicly available datasets of established use cases for data sharing to obtain high-level estimates of the value of some applications of non-personal data sharing.

The figure below provides an overview of the data and sources used for each industry. The indicators listed were identified in the datasets as being relevant to machine-generated, non-personal data, following discussions with internal experts and a review of the existing literature.

Figure: B1: Data and sources used for baseline estimates

Data and sources for baseline data

Industry	Baseline data (EU)
Agriculture	Output of the agricultural industry • Meat and crop production projections (Eurostat)
Automotive	Fuel consumption • Cost of emissions from cars – Calculated using Emissions and Unit cost of emissions data (Eurostat) Maintenance and repair services • Calculated using Passenger vehicles in the EU (Eurostat), Repair and maintenance costs (BCG)* Vehicle damage • Vehicle damage (Insurance Europe)
Healthcare	Counterfeit drugs • Current healthcare expenditure; curative and rehabilitative care (Eurostat) • Pharmaceutical expenditure % of health spending (Eurostat) • Approximate proportion of counterfeit drugs (WHO) Estimated expenditure on resource management • Current healthcare expenditure; curative and rehabilitative care (Eurostat) • Sum of expenditures on laboratory services, imaging services, patient transportation and therapeutic appliances and other medical durable goods (Eurostat)
Smart City	Street lighting • Midpoint in expenditure per street light estimates – Annual running cost per street light (EU Commission) – UK local authority expenditure on electricity for streetlights (Green Investment Group)
Manufacturing	Value of production of manufactured goods • Total value of the production of manufactured goods (Eurostat)

**Excluding accident repairs to avoid double counting with insurance*

For those EU28 countries where data on the healthcare sector is not available, an approximate value is estimated using one of two methods: the EU28 average of the non-personal data related functions as a share of a country's total healthcare expenditure is applied; or, if no healthcare expenditure data is available, a value is estimated based on its share of EU28 population.

For the variables identified in each sector, this data was used to estimate the current 2017 value in millions of euros, and to estimate its value in 2027. Where projections were not available from the source data, future projections were estimated using Eurostat forecasts of GDP or population growth where appropriate.

Figure A2: Value of indicators in each industry using available baseline data (€m)

Value of each industry in 2017 and projected value in 2027 using baseline data

Industry	Value in 2017	Value in 2027
Agriculture	408,316	434,350
Automotive	123,651	111,790
Healthcare	119,195	122,770
Smart City	4,651	4,651
Manufacturing	4,189,648	4,910,374

Note: For the Smart City estimate, the number of traditional streetlights is assumed to be unchanged from 2017 so as not to overestimate the value in 2027 (since any additional street lights installed are likely to be intelligent)

A.1.2 Survey data

Surveys were conducted with subject matter experts across the EU in agriculture, healthcare, manufacturing, Smart Cities and automotive sectors. These interviews were used to ascertain: existing and potential data sharing; and the most important obstacles to data sharing.

The questions also helped to provide a more in-depth understanding of the opportunities of machine-generated, non-personal data, and the extent to which each obstacle is likely to be overcome without external intervention.

Each expert was asked five main questions, which are summarised in the figure below. These were a combination of open questions in which experts could provide views and elaborate on their answer, and closed questions where respondents were asked to provide a score or percentage based on their expert opinion.

Figure A3: Summary of industry expert interview questions

High level overview of interview questions

Question	Overview
1. IoT impact and penetration	• Respondents were asked to score the contribution of sector-specific IoT technologies to improving the industry (i.e. through efficiencies). • Respondents were asked the score the importance of data sharing outside organization on the value of these different applications of IoT. *Score: 1 (no contribution) to 5 (very important)*
2. IoT impact and penetration	• Respondents were asked for their views on estimates of existing and future IoT penetration. *Penetration percentage*
3. Types of non-personal machine generated data sharing	• Respondents were asked to consider horizontal, vertical and external data sharing. They were asked to consider what share of the overall value expected to come from deploying IoT in their sector would depend on each type of sharing. *Share of impact*
4. Level of data sharing	• Respondents were asked to consider horizontal, vertical and external data sharing. They were asked to score, of the data currently being generated by IoT, the extent to which it is currently being shared. *Score: 1 (no data being shared) to 5 (all relevant data being shared)*
5. Barriers to data sharing	• Respondents were asked to consider potential barriers to data sharing (legal, contractual, technical, security, cultural and commercial), noting that they may interact with one another. For each type of data sharing, respondents were asked to score the barriers in terms of importance. *Scale: 1 (no impact); 2 (preventing up to 25% of valuable data sharing); 3 (25-75%); 4 (>75%); and 5 (prohibitive)*

Using these survey results, sector-specific estimates could be established for:

- Current IoT sector penetration (2017).
- Future IoT sector penetration (2027).
- Share of the benefits that relate to the three types of data sharing (horizontal, vertical and external).
- Extent to which the data being generated by IoT is currently being shared, for each type of sharing.
- Importance of barriers to data sharing for each type of sharing.

Figure A4: IoT current (2017) and future (2027) penetration, and overall IoT impact

IoT sector penetration

Industry	Current penetration	Future penetration	IoT impact
Agriculture	10%	40%	10%
Automotive	10%	90%	19%
Healthcare	5%	30% - counterfeit drugs 72% - resource management	56% - counterfeit drugs 14% - resource management
Smart City	10%	82%	63%
Manufacturing	50%	87%	30%

The table above provides estimates of current (2017) and future (2027) penetration of IoT technology. The figures under "IoT impact" represent the overall impact of IoT accounting for estimated penetration, i.e. the analysis does not assume 100% penetration of IoT technology.

In order to understand the value of data sharing and where the most salient obstacles can be found, we consider data sharing relationships, which can be grouped into three categories (horizontal, vertical and external).

Figure A5: Share of the value that will come from different types of data sharing

For each sector, the share of the value that is expected to depend on the types of sharing

Benefits	Healthcare	Manufacturing	Automotive	Smart City	Agriculture
Benefits not related to data sharing	19%	14%	8%	16%	18%
Benefits that depend on horizontal data sharing	20%	23%	24%	22%	20%
Benefits that depend on vertical data sharing	37%	42%	33%	29%	37%
Benefits that depend on external data sharing	24%	22%	36%	34%	26%

The quantitative estimates presented in this study are based on assumptions derived from expert opinion and a literature review. Thus, while the assumptions are informed by expert opinion they carry a degree of subjectivity. Revised assumptions based on new observed data may lead the estimates to be revised upwards or downwards.

A.2 Estimating the value of data sharing

The rising volume of data is not, in itself, indicative of economic value. The potential of that data needs to be realised in economic opportunities resulting from existing corporates or new entrants to put it to use. The model used in this analysis tries to identify the opportunities presented by machine-generated, non-personal data in each sector by using one or more indicators for which non-personal data could be applied.

The baseline data provides current values for 2017 and potential values in 2027. The percentage impact of the IoT estimates and the share of value that is derived from types of data sharing, obtained through the survey, are then combined with the 2027 baseline data to estimate the gross value of the IoT in 2027 and the associated gross value of data sharing.

To understand the value of removing the barriers to data sharing, it is assumed that the current level of data sharing (estimate obtained through the survey) remains constant until 2027. This counterfactual scenario considers no improvement in the level of data sharing.

Deloitte.