



# Scoping Study into Deriving Transport Benefits from Big Data and the Internet of Things in Smart Cities

Final Report for Department for Transport

Contract No. CCZZ16A22

**Customer:**

**Department for Transport**

**Customer reference:**

Contract No. CCZZ16A22

**Confidentiality, copyright & reproduction:**

This report is the Copyright of the Department for Transport. It has been prepared by Ricardo Energy & Environment, a trading name of Ricardo-AEA Ltd, under contract to the Department for Transport dated 30/11/2016. The contents of this report may not be reproduced in whole or in part, nor passed to any organisation or person without the specific prior written permission of the Department for Transport or the Commercial Manager, Ricardo Energy & Environment. Ricardo Energy & Environment accepts no liability whatsoever to any third party for any loss or damage arising from any interpretation or use of the information contained in this report, or reliance on any views expressed therein.

**Contact:**

Nikolas Hill  
Ricardo Energy & Environment  
Gemini Building, Harwell, Didcot, OX11 0QR,  
United Kingdom

t: +44 (0) 1235 75 3522

e: nikolas.hill@ricardo.com

Ricardo-AEA Ltd is certificated to ISO9001 and ISO14001

**Author:**

Nikolas Hill, Gena Gibson, Elena Guidorzi, Sofia Amaral (Ricardo Energy & Environment); Dr Ajith Kumar Parlikad and Dr Ying Jin (University of Cambridge)

**Approved By:**

Sujith Kollamthodi

**Date:**

13 June 2017

**Ricardo Energy & Environment reference:**

Ref: ED10011- Issue Number 4

## Executive summary

Big Data refers to both large volumes of data with high levels of complexity and the analytical methods applied such data which require advanced techniques and technologies in order to derive meaningful information and insights in real time.

In the transport sector – and particularly in cities – use of Big Data has the potential to realise significant efficiencies along with innovative new products and services, greater competitiveness and economic growth. Despite the great potential, the level of Big Data exploitation in transport is at a much lower level of maturity compared to other sectors such as retail and healthcare.

The purpose of this report is to provide the Department for Transport with a better understanding of the potential opportunities, challenges, enablers, data needs and the possible business models to support Big Data applications in a Smart Cities context. The report was developed over a period of four months, through consultation with key stakeholder experts and a Rapid Evidence Assessment

The transport sector is a vast generator of Big Data - advancements in analytics, enhanced computing and improvements in connectivity through the Internet of Things have opened up the potential to combine and use these big datasets in new ways. The potential transport benefits could be considerable, and include:

- Greater predictive capacity, leading to better transport and land-use planning;
- New behavioural insights;
- Better customer services;
- Better network utilisation and performance in near-time; and
- Reductions in accidents, pollution and energy use.

Free flowing data also allows better integration of modes and services, thereby improving end-to-end journeys, and underpins commercial innovation, such as Mobility as a Service. Data and improved connectivity, via the Internet of Things, also allows transport to integrate with smart city infrastructure and energy systems leading to new services for citizens, efficiencies and environmental benefits.

However, evidence quantifying the potential value of these diverse benefits for the UK economy is currently lacking, as are transport specific use cases. Moreover, the benefits are only maximised when data is used and shared, and meshed with other (transport and non-transport) datasets. However, too often datasets reside in geographical or modal siloes.

The main challenges are not technical, but rather other considerations, including identifying suitable commercial business models that benefit all parties. These challenges are compounded by data skills shortages in the transport sector and fears that sharing data will breach privacy, security and competition laws.

There are no 'silver bullets' for unlocking the potential of big data, as the challenges are often varied and complex. Nevertheless, certain actions could help to unlock the potential in the short-term which are summarised below:

The key areas for action identified are summarised below, based upon the findings in the main report:

1.

**Raising awareness, buy-in and coordination:** DfT should promote the Government's vision, the value and importance of Big Data and Data Analytics, IoT and the Smart City agenda for transport. This will help provide a more fertile environment for the sharing of key datasets and the development of intelligent mobility services using Big Data.

2.

**Standards:** DfT should address uncertainties and potential gaps in standards for big data, IoT and open architectures. This will help remove perceived barriers and facilitate big data availability and the connectivity of public and other open/shared big datasets.

<b>3.</b>	<b>Connecting datasets:</b> DfT should ensure the development and maintenance of a publically available national transport data catalogue, and facilitate the standardisation of open architectures for public datasets. This will facilitate the improving of data connectivity, which is considered crucial to the development of intelligent mobility services and obtaining the most important potential benefits more broadly across the UK.
<b>4.</b>	<b>Closing data gaps:</b> DfT should conduct further detailed analysis of the degree, nature and importance of different data gaps and test innovative solutions for closing them. In the medium term this will help close important data gaps for enabling intelligent mobility services.
<b>5.</b>	<b>Privacy, security and commercial sensitivities:</b> DfT should to continue to work with public commercial organisations to identify ways to improve certainty and overcome concerns over data privacy, security and other commercial sensitivities for quick-win results in this area.
<b>6.</b>	<b>Skills and capability:</b> Action needs to be taken more broadly in educational establishments and in the workplace, supported by appropriate Government actions, to ensure sufficient availability of individuals with the skills necessary to support the development and deployment of Big Data applications in the UK.

To deliver a step-change in this area, a sustained multi-year work programme is required, including further research into valuing the transport benefits and the scope for greater standardisation. Real-world demonstrator projects are needed to help develop and deploy innovative and commercially sustainable solutions.

Action in transport has a wide range of wider potential policy co-benefits in the areas of planning, asset management, budgets, air quality, noise, energy, trade, business, skills and employment. Stronger interdepartmental collaboration, encouraging the development of cross-sectoral solutions, and ensuring effective ex-post evaluation of projects would also be beneficial to unlock these co-benefits.

The key role for Government and city authorities should be to set the direction, identify the main problems to be solved, and to provide guidance and support to allow the market to develop and deliver solutions.

Establishing an expert Strategy Panel to advise DfT on this work programme would be an important step in the process of developing strategy and prioritising actions in this area.

Ultimately, the benefits will only be fully realised through a collaborative and multi-modal approach involving industry, academia, central and local government.

## Table of contents

<b>Executive summary</b> .....	<b>ii</b>
<b>Table of contents</b> .....	<b>iv</b>
<b>List of figures</b> .....	<b>v</b>
<b>List of tables</b> .....	<b>v</b>
<b>List of boxes</b> .....	<b>vi</b>
<b>1 Introduction</b> .....	<b>1</b>
1.1 Background .....	1
1.2 Scope and objectives of this study .....	1
1.3 Methodological outline.....	1
1.3.1 Stakeholder consultation and Delphi survey .....	1
1.3.2 Identification and review of literature .....	2
1.4 Structure of this report.....	3
<b>2 Big Data and the benefits for transport in smart city applications</b> .....	<b>4</b>
2.1 Big Data and the IoT and the trends driving their potential .....	4
2.1.1 What is Big Data?.....	4
2.1.2 What is the role of the Internet of Things (IoT)? .....	6
2.1.3 What are the other trends driving Big Data and IoT in transport and mobility? .....	7
2.2 Transport benefits of Big Data applications in Smart Cities .....	9
2.3 Research gaps and recommendations .....	11
<b>3 Big Datasets and the role of open architectures and innovation platforms</b> .....	<b>13</b>
3.1 The availability and sharing of Big Datasets for transport.....	13
3.1.1 Identification of relevant Big Datasets for transport .....	13
3.1.2 The collection and sharing of transport-relevant datasets .....	15
3.1.3 The role of sensors.....	16
3.1.4 What are the key gaps identified in relevant datasets? .....	18
3.2 The role of open architectures and innovation platforms .....	23
3.2.1 What are open architectures and innovation platforms?.....	23
3.2.2 Technical feasibility .....	24
3.2.3 Economic feasibility .....	25
3.2.4 Case studies.....	27
3.2.5 Public acceptability.....	29
3.3 Research gaps and potential solutions for Big datasets and open architecture and innovation platforms.....	30
<b>4 Possible business models to accelerate the Big Data applications in the transport sector</b> .....	<b>33</b>
4.1 Taxonomy of business models.....	33
4.1.1 Data suppliers .....	35
4.1.2 Aggregators and enablers .....	37
4.1.3 Developers .....	38
4.2 The role of government .....	41
4.3 Research gaps and potential solutions for the business models for transport Big Data applications .....	42
<b>5 The skills and capability needed for Big Data services</b> .....	<b>43</b>
5.1 Identification of key skills and capabilities .....	43
5.2 Assessment of the current situation and future potential .....	44
5.3 Potential solutions to address skills and capability gaps .....	45
<b>6 Challenges and enablers to applying Big Data in the transport sector</b> .....	<b>47</b>
6.1 Challenges.....	47

6.2	Enablers .....	49
6.2.1	Actions for industry .....	51
6.2.2	Actions for Government.....	52
6.3	Research gaps and potential solutions for the challenges and enablers of transport-related Big Data applications .....	55
<b>7</b>	<b>Potential links and co-benefits of transport Big Data to other policy areas .....</b>	<b>57</b>
<b>8</b>	<b>Summary of findings and potential solutions.....</b>	<b>60</b>
<b>9</b>	<b>References .....</b>	<b>64</b>
<b>10</b>	<b>Acknowledgements .....</b>	<b>74</b>
	<b>Appendices .....</b>	<b>75</b>
<b>A1</b>	<b>Appendix 1 – Final Delphi Survey Summary.....</b>	<b>76</b>
<b>A2</b>	<b>Appendix 2 – Summary of Benefits, Challenges and Enablers .....</b>	<b>77</b>
<b>A3</b>	<b>Appendix 3 – Examples of key datasets and transport data portals.....</b>	<b>80</b>
A3.1	Examples of the meta-analysis of key datasets provided in Annex B of TSC (2015).....	80
A3.2	Examples of transport-related data portals .....	82
<b>A4</b>	<b>Appendix 4 – DfT Actions Proposed in the Smart Cities and Transport Workshop Report .....</b>	<b>83</b>
<b>A5</b>	<b>Appendix 5 – The potential of Big Data to inform new predictive analytics .....</b>	<b>85</b>
A5.1	Turning data into predictive capabilities .....	85
A5.2	The need for new thinking in Smart City data access and predictive analytics .....	85

## List of figures

Figure 2.1: The Data Spectrum.....	5
Figure 2.2: Illustration of the three basic layers for smart infrastructure connected by communications .....	7
Figure 2.3: Changes in data storage costs .....	8
Figure 3.1: Precision of geo-location technologies .....	15
Figure 3.2: Mechanisms for transport data collection .....	17
Figure 3.3: Stakeholder assessment on the nature of the identified data gaps.....	22
Figure 4.1: Examples of service types and overlaps between different roles.....	35
Figure 5.1: Roles needed at different stages to implement a Big Data project or strategy .....	44
Figure 6.1: Overview of possible solutions in open government data initiatives .....	53

## List of tables

Table 2.1: Transport-specific benefits of Big Data .....	9
Table 2.2: Estimates of the impacts of open data on GVA per capita .....	10
Table 2.3: Key economic benefits associated with the use of Big Data in the transport sector .....	11
Table 2.4: Key research gaps and recommended actions associated with understanding the benefits of Big Data applications in the transport sector .....	12

Table 3.1: Intelligent mobility data themes and example datasets .....	14
Table 3.2: Transport-related data gaps assessed in this project (red = most important; yellow = least important) .....	18
Table 3.3: Key understanding/research gaps and potential solutions to address Big Data gaps .....	30
Table 4.1: Key roles and business models .....	34
Table 4.2: Key understanding/research gaps and recommended actions on the business models for transport Big Data applications .....	42
Table 5.1: Key gaps and possible actions/solutions identified in the literature and consultation for this study with regards to ensure the skills needed are available in the UK in the future .....	45
Table 6.1: Challenges to applying Big Data in the transport sector .....	48
Table 6.2: Enablers for Big Data and IoT applications in the transport sector .....	50
Table 6.3: Key understanding/research gaps and recommended actions for the challenges and enablers for transport Big Data applications .....	55
Table 7.1: Summary of potential cross-departmental / policy agenda impacts resulting from transport-related Big Data and IoT development in Smart Cities .....	57
Table 7.2: Key understanding/research gaps and potential solutions for the potential links and co-benefits of transport Big Data to other policy areas .....	58
Table 8.1: Key recommended actions for the Department for Transport to facilitate the deriving of transport benefits from Big Data and IoT in a Smart Cities context .....	61
Table 10.1: Organisations contributing to the consultation and wider evidence base for the project ...	74

## List of boxes

Box 2.1: Summary of key findings on the transport benefits of Big Data for Smart City applications ....	4
Box 3.1: Summary of key points on open architectures and innovation platforms .....	13
Box 3.2: The role of metadata in open government data initiatives .....	16
Box 3.3: Key reasons for current data gaps identified by stakeholders .....	21
Box 3.4: International Examples of Smart Cities: Copenhagen .....	26
Box 3.5: FIWARE – lessons learned and key success factors .....	27
Box 3.6: Immense Simulations – lessons learned and key success factors .....	28
Box 4.1: Summary of key findings on business models .....	33
Box 4.2: International Examples of Smart Cities: Rio de Janeiro .....	40
Box 5.1: Summary of key findings on skills and capability .....	43
Box 6.1: Summary of key findings on challenges and enablers .....	47
Box 6.2: International Examples of Smart Cities: Singapore .....	54
Box 6.3: International Examples of Smart Cities: Barcelona .....	55
Box 7.1: Summary of key points on connections to other policy areas .....	57

# 1 Introduction

## 1.1 Background

Big Data refers to both large volumes of data with high levels of complexity and the advanced analytical techniques and technologies applied to such data to derive meaningful information and insights in real time. Big Data is already transforming the way we conduct science, deliver healthcare, run cities and operate businesses. In the transport sector – and particularly in cities, due to their complex intermodal systems – use of Big Data could help to maximise efficiency gains, improve asset management and enhance the experience of transport users. Such advances underpin the effective delivery of a number of emerging and transformative technological developments, i.e. Cooperative Intelligent Transport Systems (C-ITS), Smart Cities, Mobility as a Service (MaaS), and Connected and Autonomous Vehicles (CAV).

Despite this potential, the level of Big Data exploitation in transport is at a much lower level of maturity compared to other sectors. There are many reasons for this, including the fragmented availability of transport data, data privacy issues, institutional barriers and the availability of suitable business models for encouraging data sharing, amongst others.

The Department for Transport has a strong interest in data science, open data and Big Data and how this could deliver tangible transport benefits and contribute to the wider government agenda. In order to realise the benefits of Big Data in the transport sector, it is important to learn from best practice elsewhere and to understand the opportunities, challenges and enablers, business models and data needs.

## 1.2 Scope and objectives of this study

The Department for Transport (DfT) commissioned Ricardo and the University of Cambridge to undertake a scoping study into *“Deriving transport benefits from Big Data and IoT in a Smart Cities context”*.

This study includes the following elements:

- A review of the existing evidence base.
- A gap analysis to identify research gaps and recommended actions.
- Identification of the available ‘big’ datasets.
- A discussion on the benefits, the challenges and enablers associated with Big Data and IoT.
- An assessment of the feasibility of open architectures and innovation platforms.
- An assessment of whether there are any skills shortages.
- Identification of new business models.
- A discussion of the areas of the connectivity between policy agendas across Government (e.g. health; access to education, training and employment; service provision and the environment).

We developed this report over a period of four months, using consultation with key stakeholder experts (via a Delphi survey and one-to-one interviews) and a Rapid Evidence Assessment (REA).

## 1.3 Methodological outline

### 1.3.1 Stakeholder consultation and Delphi survey

Our main method for engaging with expert stakeholders was the Delphi method, which is widely recognised as a robust and systematic approach for engaging with multiple experts on complex topics. The Delphi method is based on structural surveys that are sent to experts in two or more ‘rounds’; where each respondent sees aggregated, anonymised results from earlier rounds of questions. This process allows for effective consensus building when considering long-term issues and/or high uncertainty; and improves the robustness and legitimacy of the analysis.

Due to the very short time-frame for the project, the survey process was limited to two rounds carried out between January and March 2017, using a combination of telephone interviews and an online survey. We developed the survey questions in discussion with the Department for Transport at the start of the process. The two rounds were as follows:

- **Survey Round 1:** To maximise efficiency, Round 1 was largely based on telephone interviews covering a range of quantitative and qualitative questions, with complex issues broken down into sub-topics that could be judged individually. Controlled feedback was provided on the results, including:
  - Numerical feedback, to show the breadth of the opinions on, for example the importance rating of particular elements.
  - Qualitative feedback, summarising the arguments and counterarguments of the various answers provided by those participating in the survey.
- **Survey Round 2:** In the second round we asked the participating experts to complete an online survey questionnaire, covering the same issues from Round 1, but including the opinions of the other experts that we gathered during the first round. We also added a limited number of qualifying questions in Round 2 to explore in more detail some of the findings from Round 1.

In addition to the Delphi survey, we also carried out other interviews to inform specific topics covered by the study in more detail. For example, we consulted with experts on open data architectures and innovation platforms for Smart Cities, and used this to inform our analysis (i.e. technical feasibility, economic/market feasibility and legal/political feasibility). We used the information gathered from these interviews to develop a set of case studies (see Section 3.2.4).

Overall, we consulted with experts from over 30 organisations from academia, public sector organisations and private sector organisations. Consultees included universities, Government departments, local authorities, civil society organisations, data platforms, representatives from the technology industry and representatives from the transport industry (see Section 10 for a full list of the organisations consulted and Appendix 1 for full details of the survey results).

Additional information and insights were obtained from stakeholders participating in DfT's Smart Cities and Transport workshops (Department for Transport, 2017), and discussion of the initial survey results with the IET (Institution of Engineering and Technology) transport policy panel.

### 1.3.2 Identification and review of literature

In parallel to the stakeholder consultation activities, we conducted a review of the available literature. This process included the elements:

1. **Identification of evidence sources:** we expanded the initial list of published research provided by DfT at the start of the project through further searches at the start of the study. The research strategy and search criteria were agreed in advance with DfT; this resulted in us identifying of over 120 pieces of evidence in total, which we catalogued and provided to DfT.
2. **An initial, high level review of the literature:** to help assess and filter the evidence, we conducted a Rapid Evidence Assessment (REA) of the literature identified, using standard approaches for REAs that we have previously used when carrying out research for Government Departments.
3. **Identification and prioritisation of the key pieces of literature:** We assessed the literature according to the evidence from the REA (e.g. on topic coverage, authorship and level of detail on the study topics) to identify the key literature sources that we could review in detail within the project timescales.
4. **In-depth analysis of the key literature sources:** We reviewed more than thirty pieces of evidence in greater detail to extract the information and insights necessary to address the topics under investigation and also to cross-corroborate the findings of the consultation where feasible. We also identified and reviewed a range of additional evidence during the stakeholder consultation process.

## 1.4 Structure of this report

This report is structured as follows into eight core chapters; following these, the key references cited in this report are provided, as well as a list of consulted organisations acknowledged for their contributions to the project and its findings:

- *Section 1, Introduction:* This introductory section provides a brief overview of the project aim and objectives, and a summary of the methodologies applied to gather evidence to inform the development of the report.
- *Section 2, Big Data and the benefits for transport in smart city applications:* This section provides an overview of Big Data, the Internet of Things (IoT) and the trends driving them and application to transport. The anticipated benefits in Smart City applications is also provided based on the evidence review and stakeholder consultation.
- *Section 3, Big Datasets and the role of open architectures and innovation platforms:* This section provides a summary of the availability and sharing of Big datasets, and the role of open architectures and innovation platforms in helping to improve availability and sharing of such datasets and realise the potential of them.
- *Section 4, Possible business models to accelerate the Big Data applications in the transport sector:* This section summarises the evidence on the business models available to help build the business/commercial case for Big Data applications in the transport sector, addressing some of the issues/challenges identified to deliver tangible benefits in cities.
- *Section 5, The skills and capability needed for Big Data services:* This section provides an assessment of whether there are any skills shortages, in both the public and private sectors, that are impeding transport applications and the resulting benefits realisation; and what measures could be undertaken to address these.
- *Section 6, Challenges and enablers to applying Big Data in the transport sector:* This section provides a summary of the key barriers (i.e. challenges, complexities, etc.) which have been identified to be restricting the potential applications, and the potential enablers that have been proposed to encourage more rapid uptake of Big Data and IoT enabled intelligent transport services.
- *Section 7, Potential links and co-benefits of transport Big Data to other policy areas:* This section identifies the key areas of connectivity between policy agendas across Government, including how transport smart initiatives in our cities could support the policies of other government departments.
- *Section 8, Summary of findings and potential solutions:* In this final section the findings of this work have been used to identify research gaps and provide recommendations on possible actions that could be taken to accelerate the unlocking of the potential of Big Data and IoT for transport.

## 2 Big Data and the benefits for transport in smart city applications

### Box 2.1: Summary of key findings on the transport benefits of Big Data for Smart City applications

#### *Big Data, the IoT and the trends driving their potential in the transport sector*

- The potential for Big Data to enable intelligent mobility has increased in recent years due to a combination of developments: (i) extensive deployment of sensors producing real-time data streams, (ii) the development of sophisticated techniques for in-stream analysis, (iii) new analytic frameworks allowing the interrogation of Big Datasets, and (iv) advances in data storage and computational power.
- Further trends that are driving Big Data and IoT include: crowd-sourcing as a data source (e.g. from social media), entry of telecoms operators and advanced technology companies such as Apple and Google to the transport sector, and the development of intelligent mobility services and technologies (e.g. Co-operative Intelligent Transport Systems (C-ITS) and Connected and Autonomous Vehicles (CAV)).

#### *The benefits of Big Data applications in Smart Cities*

- The use of Big Data in Smart City applications is beneficial to both the public and private sectors – through fostering commercial innovation, increasing the efficiency of public services and improving engagement of citizens.
- Transport-specific benefits include: improved planning and operational efficiency; better knowledge on transport network operations; higher levels of customer service; improved safety; better air quality, reductions in vehicle emissions; and increased social value.
- However, there is a lack of information on the quantitative value of such benefits, and such information is needed to properly communicate the value of data (and what to do with it) to key stakeholders, support the growth of business models for deriving transport benefits from Big Data and prioritise areas for development.



### 2.1 Big Data and the IoT and the trends driving their potential

#### 2.1.1 What is Big Data?

Big Data is a term that is used to describe extremely large and/or complex datasets and the analysis of such datasets to identify useful or interesting patterns and trends. The key characteristics of Big Data are frequently expressed in terms of the '4Vs' (IBM, 2017), (InsideBIGDATA, 2013), (DeVan, 2016):

1. **Volume:** One of the most iconic characteristics of Big Data is its sheer volume, with massive amounts of data being generated by machines, networks and human interaction on systems like social media.
2. **Variety:** This refers to the many sources and types of data, both structured (e.g. databases or other similar formats) and unstructured (e.g. emails, photos, videos, monitoring devices, PDFs, audio, etc.) The latter types of data create significant challenges for data storage, mining and analysis.
3. **Velocity:** This relates to the enormous rates at which new data are created, including in real-time, as well as the rates at which analysis needs to take place in order to utilise it.

- 4. **Veracity:** This concerns the biases, noise and abnormality in data – i.e. the quality of the data itself. Data must be consolidated, cleansed, consistent, and current to facilitate the right decisions.

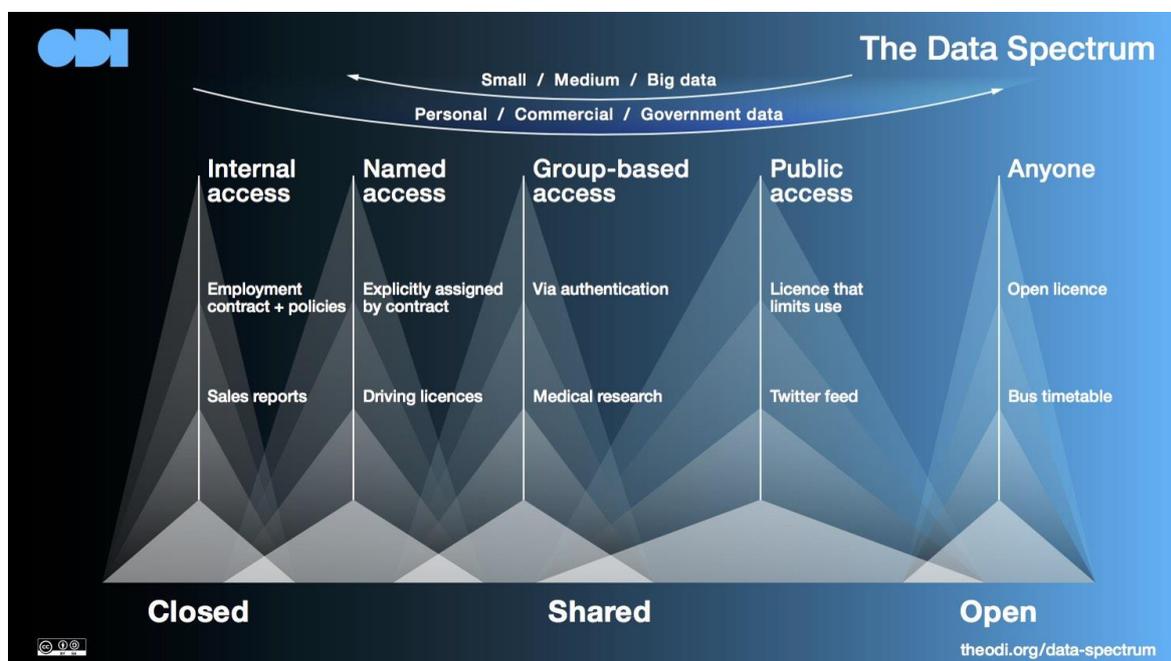
There are a number of additional 'Vs' that are also sometimes cited, such as:

- 5. **Variability:** This refers to data whose meaning is constantly changing, which can have a huge impact on data homogenisation. Companies have to develop sophisticated programmes which can 'understand' context and decode the precise meaning of words through it.
- 6. **Visualisation:** refers to the use of charts and graphs in order to present large amounts of complex data in ways that convey meaning in a more effective manner than reports.
- 7. **Value:** refers to the benefits that result from big datasets once they are ready for use. The value of data is not absolute – in particular, the value of raw data may be different to data that has been analysed and processed (such actions can increase the value of data by, for example, making it more accessible).

An important concept, that we will discuss in more depth later in this report, is "Open Data", i.e. data that is made freely available for anyone to use and republish as they wish, without restrictions due to copyright, patents or other mechanisms of control (Policy Exchange, 2016). Currently, open data are mainly (though not exclusively) provided by public sector organisations in the UK. Examples include: data.gov.uk and the various datastores set up by cities authorities including Glasgow, Nottingham and London (see also later Section 3.1).

In practice, there is a spectrum of data currently available, ranging from closed to open, as illustrated in Figure 2.1. Open data are free for anyone to access, whereas "Shared data" includes data to which access is granted to certain named people, groups or to the public in general under more restrictive terms compared to open data. Closed data can only be accessed internally by its subject, owner or holder (ODI, 2017).

Figure 2.1: The Data Spectrum



Source: (ODI, 2017).

### 2.1.2 What is the role of the Internet of Things (IoT)?

The Internet of Things (IoT) refers to the integration of the physical world with the digital world, where objects around us interact and cooperate with each other to provide improved information, enable better decision making and to reach common goals (Sarma et al., 1999).

IoT technologies are already shaping a number of different applications in the transport sector, such as intelligent transport systems (ITS), and the use of sensors, actuators and data capture technologies is rapidly growing<sup>1</sup>. ITS make use of information from cameras and microcontrollers to optimise public transport, monitor the environment and run security applications. Examples of where IoT technologies are already playing a role in ITS and the transport sector more widely include vehicle-to-vehicle and vehicle-to-infrastructure communication to transmit and use traffic data; machine-to-machine communication for connected vehicles to provide added features such as controlling heat settings or charge settings for hybrid/electric cars (Merswolke, 2003); electronic on-board recorders in trucks delivering data on load/unload times, travel times, driver hours, truck driver logs; pallet or trailer tags delivering data on transit and dwell times; and GPS updates from vehicle fleet sensors. (Rusitschka & Curry, 2016). Furthermore, modern vehicles include numerous connected systems that provide drivers with a range of different abilities ranging from streaming media, providing navigation, finding parking spaces and sensing the outside environment (Venturebeat, 2015). Crowd-sourcing from personal communication devices constitutes an increasingly significant source of data in the transport sector. Such data can be generated as a by-product from the use of smart device applications (“apps”) such as social networks, or be generated directly from users in order to develop the apps (as in the case of apps like Waze<sup>2</sup>) (Venturebeat, 2015a).

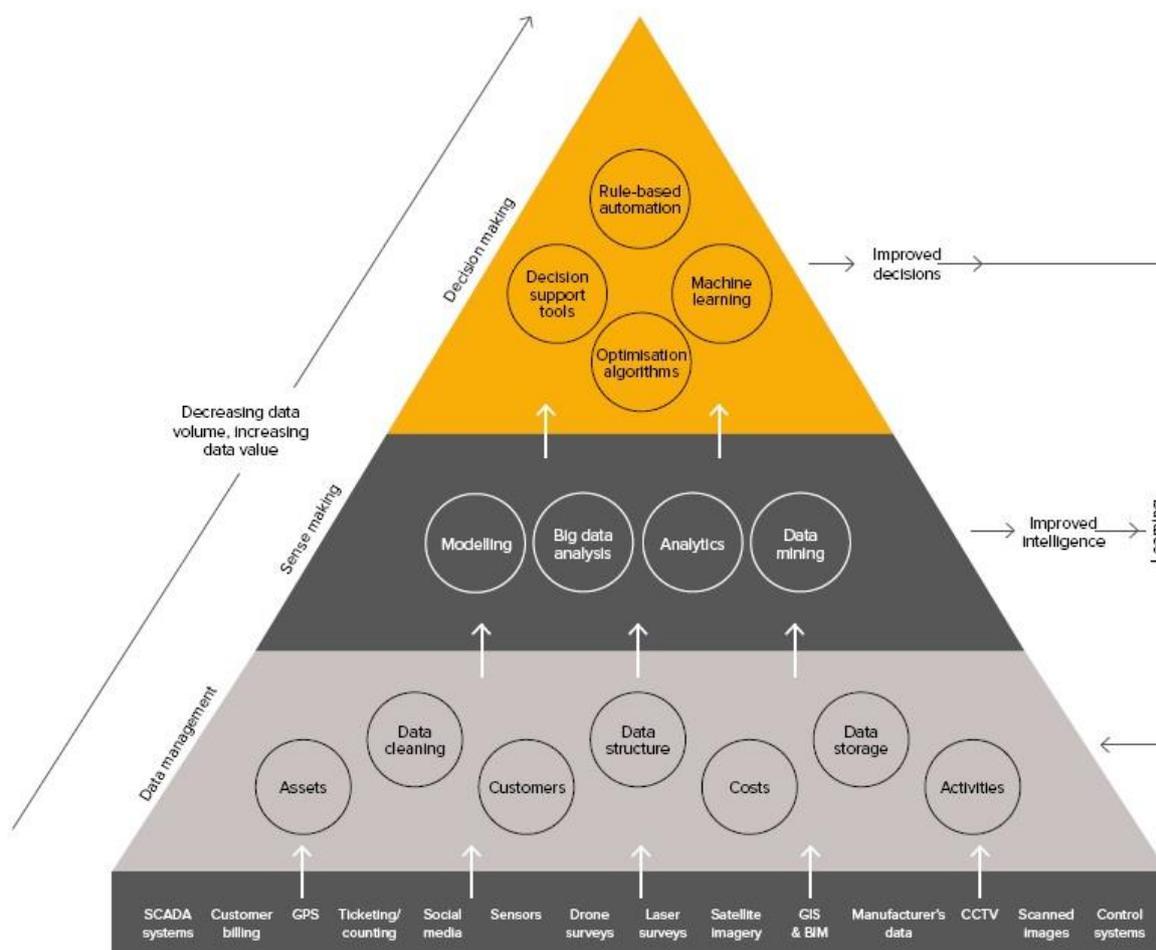
In addition to the concept of IoT, there are also critical advancements occurring in artificial intelligence, machine learning and autonomous vehicles. For instance, the emerging concept of ‘smart infrastructure’ – an extension to the IoT concept – is the result of combining physical infrastructure with digital infrastructure, providing improved information to enable better, faster and more cost-effective decision making (University of Cambridge, 2016), as illustrated in Figure 2.2. Such smart infrastructure involves monitoring the condition and performance of infrastructure assets and systems such as bridges, pavements and tunnels using sensing systems such as distributed fibre optic sensors and wireless sensor networks (Soga et al., 2016) in addition to robotic inspection and advanced analytics applied to satellite images. The growth of smart infrastructure solutions is being further accelerated by the emergence of low power miniature sensors and energy harvesting sensors that address critical issues of size and power management, as well as increasing interoperability requirements driven by standards.

---

<sup>1</sup> Sensors are used to monitor and actuators are used execute selected tasks in response to the analysis of sensor data.

<sup>2</sup> The Waze navigation app gathers map data and traffic information from its users to provide routing and real-time traffic updates.

Figure 2.2: Illustration of the three basic layers for smart infrastructure connected by communications



Source: (University of Cambridge, 2016)

### 2.1.3 What are the other trends driving Big Data and IoT in transport and mobility?

A number of key trends driving Big Data and the IoT are anticipated to radically change how data is captured and exploited in the transport sector. This, in turn, is expected to bring significant changes to the provision of transport services and have an impact on the market forces in the sector. New mobility forms and services are expected to be developed based on Big Data. These would entail integrated, multi-modal, seamless and user-focused approach to transport services and solutions (e.g. MaaS). The technology to collect big data from the transport sector has existed for some years but its full potential has not yet been realised. This appears to be principally due to the challenges identified above, regarding the fragmentation of transport-related data and the difficulty of combining transport data with wider sources due to privacy and security concerns.

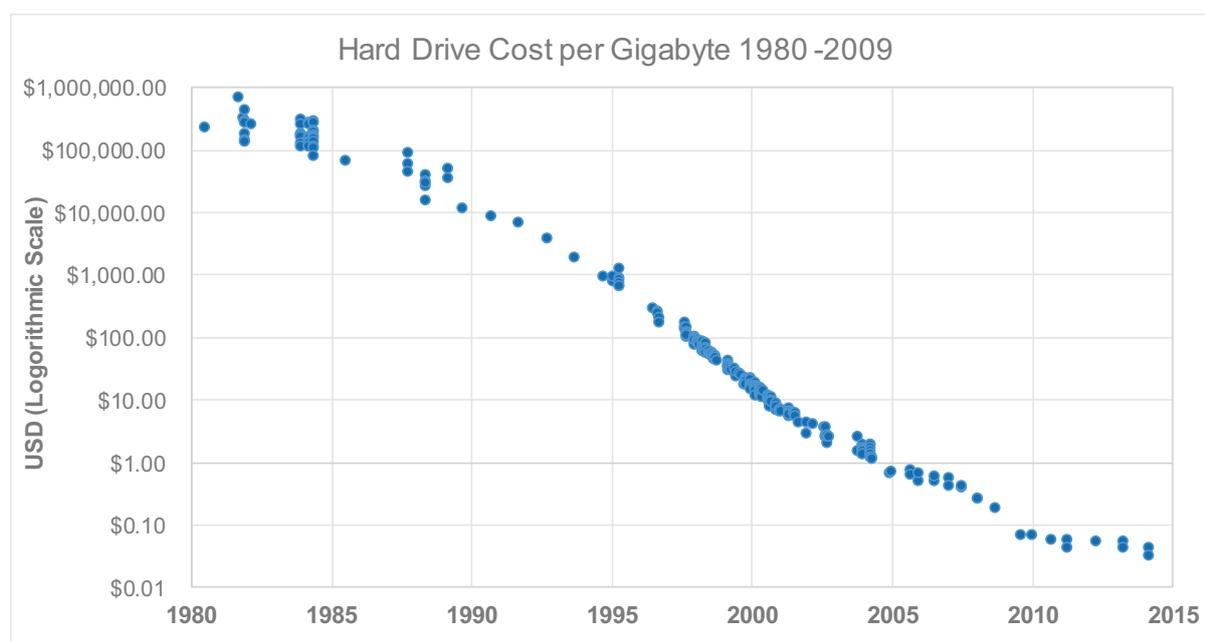
#### 2.1.3.1 Technological developments

Big Data is enabled by the increasing number of sensors, devices and ICT applications connected to the internet. The following four interlinked technological trends are driving the rapid growth of Big Data (ITF, 2015):

- i. **Extensive deployment of sensors producing real-time data streams:** real-time production of vast amounts of data enabled through the deployment of millions of sensing devices and extensive event logging (and storage).

- ii. **Development of sophisticated techniques for data analysis:** sophisticated algorithms and distributed computing capacity enable real-time analysis of data. New and powerful techniques have emerged, such as in-memory processing (i.e. analysis that occurs without data storage), cognitive computing and machine learning provided by public cloud data and streaming analytics services, that can be used to extract relevant data from unstructured analogue video or audio streams (Kobielus, 2016).
- iii. **New analytic frameworks allowing the interrogation of big datasets:** allowing efficient processing of very large datasets within the constraints of available runtime computing capacity, with many techniques released under open-source licences.
- iv. **Advances in data storage and computational power:** falling data storage costs (see Figure 2.3<sup>3</sup>) have facilitated huge increases in the amount of retained data; such data may previously have been discarded. Data is increasingly stored remotely - a key development is the use of cloud-based computing<sup>4</sup> to enable economies of scale and new possibilities in the analysis, storage and sharing of particularly large datasets.

Figure 2.3: Changes in data storage costs<sup>5</sup>



Source: adapted from (Komorowski, 2014)

### 2.1.3.2 The changing competitive landscape in the transport sector

Organisations that were previously external to the transport sector are now active participants. These companies are not only entering the market as data providers, but also expanding their role as developers of mobility services and directly competing with traditional players in the transport sector. Examples include Google's development of self-driving vehicle technology, and Amazon's trialling of drones to deliver their products.

On the other hand, data-driven business models are opening up new opportunities for incumbents, with organisations such as TomTom now expanding its services based on the data they collect from their Sat Nav devices and providing real-time traffic data to drivers. More recently they have partnered with

<sup>3</sup> Despite not being linear historic data of hard drive cost shows a clear decreasing trend over the 30-year period.

<sup>4</sup> Cloud computing allows the sharing of computer resources, rather than having local servers handle applications, allowing wider access to more robust applications and programs. The "cloud" refers to remote data storage centres as well as the suite of data transfer and networking protocols that allow access to and analysis of distributed data as if it were located on a single server

<sup>5</sup> The log scale was reproduced using data made available by the source. The majority of the data were retrieved from historical notes about the cost of hard drive storage space. Data from 2004 until 2009 was retrieved using the Wayback Machine archive.org.

the city of Moscow to help design a road transport programme for the city. TomTom built a city portal for Moscow to connect consumers with the city's public administration, to help improve access to road data and also to encourage the use of data analytics of such datasets to help reduce congestion and implement further the Smart City vision (TTT, 2016).

### 2.1.3.3 Mobility is becoming intelligent

The emergence of new intelligent mobility services is an important trend that is influencing developments in the transport sector. A range new mobility services are expected to be deployed over the next ten years:

- In the near-term, the focus will be on better integration of existing transport systems using data that are already available. Examples include the provision of real-time parking information and crowd-sourced transport on-demand services such as Uber and Lyft ride-hailing services, as well as the greater use of C-ITS (Co-operative Intelligent Transport Systems) technologies, which can improve knowledge on the locations and speeds of vehicles with consequential improvements in traffic management.
- In the medium term, the focus will be on optimisation of mobility across multiple transport modes and greater use of predictive analytics. Mobility may increasingly be provided as a service (Mobility as a Service - MaaS) in the sense that there could be a shift towards door-to-door mobility solutions which are provided as a consumer service, negating the need for individuals to own private vehicles. Furthermore, data is also expected to be used for improving transport planning and addressing demand constraints in the network.
- In the longer-term, new forms of on-demand mobility requiring real-time data sharing are expected to emerge. Intelligent mobility is likely to be in the form of connected and autonomous vehicles (CAVs) which further enhance the MaaS trend. For freight transport, drones and cargo airships could be deployed in the future.

## 2.2 Transport benefits of Big Data applications in Smart Cities

The new data sources discussed above will provide ample opportunities to combine different types of data for monitoring, measuring and understanding the supply of, and demand for transport. This can help overcome existing fragmentation in transport data and generate significant benefits in relation to understanding transport systems as a whole, particularly in cities that have complex multimodal provision of services. The fusion of purposely-sensed, opportunistically-sensed and crowd-sourced data generates new knowledge about transport activity and flows.

Table 2.1 provides an overview of key transport-specific benefits deriving from Big Data applications drawing from our literature review and discussions with experts. Although analysis of the results of our research did not convey a sense of prioritisation of the benefits identified, these benefits are, in general, associated with improved service delivery, market efficiency and customer experience.

A further discussion of the potential of Big Data to inform new predictive analytics is also provided in Appendix A5.

**Table 2.1: Transport-specific benefits of Big Data**

Transport benefits	
<b>Improved knowledge of transport planning and operations</b>	The implementation of new data analytic tools will produce new insights and knowledge on how to optimise the use of existing infrastructure. This should result in improved traffic and event management, better transport modelling and demand forecasting. Data can also be used to ensure improvements in asset maintenance in real time, or in a predictive manner, improving efficiency and effectiveness of transport systems.
<b>Improved effectiveness and efficiency of transport networks</b>	Thanks to the operational and management improvements, data analytic implementation tools are expected to improve effectiveness and efficiency of transport networks planning, management and operations.

Transport benefits	
<b>Improved customer service</b>	Gathering insights on customer needs and feedback (through for example, ticketing, social media and MaaS applications) allows for service improvements while satisfying the increasing demand of transport customers for timely and accurate information about door-to-door journeys.
<b>Improved safety and reduced risk</b>	Big Data enables causal factors and patterns to be identified and, in turn, this assists in developing and implementing measures to reduce safety risks. For example, transport and non-transport related datasets (e.g. on major events, weather data) can enable a faster response to emergencies and incidents and develop better-informed strategic plans for winter readiness.
<b>Improved air quality and reduced emissions</b>	Combined analysis of air quality monitoring and traffic speed information (e.g. C-ITS) can inform traffic management centres, which can take real-time informed action to divert traffic or introduce speed restrictions, which can have an impact on air quality in sensitive areas, such as schools or pedestrian areas.
<b>Increased social value</b>	The integration of transport-related data deriving from mobile GPS technology (e.g. smart phones, tablets) with social media feeds and other consumer-based data (e.g. feedback, sentiment data) allows for sentiment mapping and situation analysis associated with transport experiences. The results from such data analysis can improve the customer experience and influence travellers' behaviours. The social benefit associated with this relates to the time that travellers could save by having access to real-time and crowd-sourced information. Moreover, Big Data and IoT applications are expected to contribute to improve mobility accessibility for all.

Sources: (Rusitschka & Curry, 2016); (Royal Academy of Engineering, 2015); (Policy Exchange, 2016); (IET, 2017); (TSC, 2015) and the findings of the consultation with experts by Ricardo for this project.

With regard to the economic value associated with Big Data and IoT applications in transport, stakeholders consulted for this study and the literature indicated that there is often a lack of understanding on what the value of data is. At the moment the value of data is still an intangible asset in many companies, and the lack of recognition of this value hinders the opportunity to assess and measure the contribution of data to the UK economy. As a result, decision-making on how best to develop, trade, protect and exploit such assets is potentially compromised (Royal Academy of Engineering, 2015).

The evidence from the literature on quantifying the benefits of Big Data and IoT is currently weak, as there are currently few examples where ex-post analysis of project outcomes has been carried out. To address this gap, we have identified estimates from the literature of the impact of open data on Gross Value Added (GVA)<sup>6</sup> from across Europe. These estimates are summarised in Table 2.2 but we stress that these values are not specific to the transport sector.

**Table 2.2: Estimates of the impacts of open data on GVA per capita**

Source	GVA per capita
Gartner, Innovative use of public data, report to the National IT and Telecom Agency, and the Danish Agency for Science, Technology and Innovation (Danish), 2009	£54
Analysis by CS Transform Limited, based on modelling by Pollock, R. (2011), "Welfare gains from opening up Public Sector Information in the UK"	£122
Open data measures in the Autumn Statement 2011, Cabinet Office as quoted in <i>Open Data - driving growth, ingenuity and innovation</i> , Deloitte, 2012	£255

<sup>6</sup> Gross Value Added (GVA measures the contribution to the economy of each individual producer, industry or sector. To estimate GDP, the GVA is used to measure the production / contribution of each industry or sector. (Definition by the Office of National Statistics)

Source	GVA per capita
Data equity - Unlocking the value of Big Data, Centre for Economics and Business Research Ltd, April 2012	£650

Source: (TSC, 2015), pp. 66-67. Figures have been normalised to provide GVA uplift per capita.

In terms of local UK examples, a Transport Coordination Centre set up for the London 2012 Olympics to provide coordination between teams of TfL's Surface Transport & Traffic Operations Centre, Network Rail and the Highways Agency, among a number of other governmental departments, the Olympic Delivery Authority and the London Organising Committee of the Olympic Games. (Schneider Electric et al, 2014) estimated that if the TCC had been continued, it could have contributed around £4.4 million of economic benefits per year from time savings alone. In addition, a study for the UK Government has previously estimated the annual value of time saved resulting from the release of TfL's open data at up to £58m per year from an annual spend of less than £1m (Deloitte, 2013). Whilst these examples are specific to London, and therefore difficult to scale to the UK, they do give an indication of the potential size of the prize that could be available to the transport sector through open data sharing and smart city applications using of big datasets.

The key economic benefits for both public sector and private sector actors associated with the use of Big Data in the transport sector are presented in Table 2.3.

**Table 2.3: Key economic benefits associated with the use of Big Data in the transport sector**

- *Time and money savings for transport operators, service users, and the economy.*
- *Passenger/customer revenues arising from the use of new forms of mobility, and increased sales of related services.*
- *Payment for strategic planning and operational insights, and advice, derived using transport-related data.*
- *Payment for data itself, which will continue to flow between mobility service providers, organisations, and individuals as they interact.*
- *Revenues from partnerships/sales of tools and applications to major technology companies that are growing by acquisition.*

Notes: (TSC, 2015)

The challenge of defining the tangible value of data represents one of the main issues regarding the ability to develop viable business models. These aspects are explored in depth in Section 4.

## 2.3 Research gaps and recommendations

The transport benefits resulting from Big Data and IoT applications in smart cities could be significant and span a wide range of areas including economic and social improvements. However, while there is general agreement on the types of benefits and opportunities in this area, there is still a lack of research on the economic value associated with the use of Big Data in transport. In particular, it is difficult to put a value on data if there are no existing examples demonstrating its usefulness, which in turn makes it impossible to make optimal decisions about how to develop, trade and exploit such data. Gaps also exist regarding the potential environmental and social impacts associated with using Big Data to support the transport sector. These issues are further explored in Section 7, with the main research gaps summarised in Table 2.4.

**Table 2.4: Key research gaps and recommended actions associated with understanding the benefits of Big Data applications in the transport sector**

# Key research gaps	
1.	There is limited research that <b>quantifies the direct economic benefits</b> of Big Data applications <b>in the transport sector</b> and the <b>monetary value of data</b> , which were identified in the literature and by stakeholders as key challenges to unlocking the full benefits of Big Data. Research is needed to address these gaps, which in turn will help to inform a better understanding of potential business models.
2.	There is a lack of information regarding the <b>wider economic benefits of Big Data</b> , beyond the transport sector. It is anticipated that the use of data in transport will generate a multiplier effect in the economy but the literature fails to provide detailed information on this. There is a need to conduct research to develop a portfolio of case study evidence to demonstrate how to deliver such benefits in a Smart City context.
3.	More research is needed on the <b>environmental and societal benefits of Big Data</b> . The literature recognises the type of benefits to society and the environment, but an understanding of their potential scale and the means through which they can be achieved is limited.

## 3 Big Datasets and the role of open architectures and innovation platforms

### Box 3.1: Summary of key points on open architectures and innovation platforms

#### *Big Datasets for transport in Smart Cities*

- The five most important data gaps identified in discussion with experts were:
  1. Bus and rail vehicle occupancy data (e.g. from passenger ticket purchase, barriers, electronic ticket machines).
  2. Traffic speed and count data from urban traffic management control systems and non-strategic roads.
  3. Real time parking space availability for on and off-street car parks.
  4. Local weather data.
  5. Real time condition of infrastructure assets.
- The root causes of the gaps are the lack of sensor networks, the temporality with which transport-related datasets are collected, maintained and shared, and the private ownership of datasets.
- Many of the identified data gaps could potentially be overcome through the use of better business models that encourage data sharing in combination with the development and application of consistent (international) standards and open data platforms.



#### *The role of open architectures and innovation platforms*

- There is a strong argument for developing open architecture and innovation platforms in order to maximise the potential for connections between multiple creators and users of data in a city. Regional and national platforms would help to overcome issues facing smaller cities in terms of reaching critical mass and/or securing adequate funding.
- To ensure the feasibility of open architecture and innovation platforms, it is necessary to ensure:
  - Technical feasibility – especially improved standardisation, which influences the ease / extent to which data can be shared and linked – and consequently the resulting level of innovation.
  - Economic feasibility – especially clarifying the monetary value of data.
- Public acceptability– especially ensuring public acceptance and trust around protection and use of open data.

## 3.1 The availability and sharing of Big Datasets for transport

### 3.1.1 Identification of relevant Big Datasets for transport

A detailed meta-analysis has recently been conducted for the Transport Systems Catapult, culminating in a publically-available catalogue of transport-related datasets in the UK<sup>7</sup>. The different datasets can be categorised into nine different groups by theme and the overall intelligent mobility context, as summarised in Table 3.1. Rather than repeat this work we have summarised and sought to build upon some of the key findings in our scoping study. Moreover, the data portal run by the UK government (i.e. data.gov.uk – Opening Up Government<sup>8</sup>) provides an up-to date catalogue of transport related datasets relevant across the UK which are uploaded by government department, local councils and relevant agencies (e.g. DVLA).

<sup>7</sup> Available via the IM Data Index (<http://imdata.co.uk>)

<sup>8</sup> Available via: [https://data.gov.uk/data/search?sort=title\\_string+asc&theme-primary=Transport#search-sort-by](https://data.gov.uk/data/search?sort=title_string+asc&theme-primary=Transport#search-sort-by)

**Table 3.1: Intelligent mobility data themes and example datasets**

Theme	Definition	Example Datasets	
<b>Place &amp; Space</b>	Anything tangible that can be seen, touched or found.	TfL's digital speed limit dataset Rail station / bus stop locations	OSM/OS/Google/Apple/Nokia/Bing maps Locations of on and off-street car parking
<b>Environment</b>	Data relating to environmental trends and natural occurrences.	Real-time weather, sunrise/sunset times Rainfall, tides and marine conditions Floods	Earthquake monitoring Air quality Droughts
<b>People, Things &amp; Movement</b>	Data generated by individuals and things as they move around.	Location history from cell/smartphones Employment/health/education data	Stats19 road incident data Bike journey counters/origin-destination (O-D) data
<b>Disruption and event-related data</b>	Dynamic datasets related to physical events that impact transport networks.	Sporting fixtures Live and planned road closures	Real-time traffic incident reports Real-time data on road gritting in winter Roadwork locations
<b>Public Transport Services</b>	Scheduled and real-time data relating to the movement of public transport vehicles, and their characteristics.	Traveline National Dataset –public transport routes and schedules Traveline NextBus Real-time data	ATOC/National Rail Enquiries live departure boards TfL Journey planner
<b>Personal Automobility</b>	The spatial movement of powered personal vehicles (e.g. cars, motorcycles, taxis).	Highways Agency live traffic information DATEX II XML Waze/Google traffic speed data	TfL live traffic camera images Glasgow on-street parking data Vancouver EV charging stations
<b>Freight connections</b>	Data related to the movement of goods by road, rail, sea, air.	Shipfinder Road and port freight statistics	Track-Trace API Freight train movements
<b>International Connections</b>	International travel outside of the UK by Air, Rail, Sea.	Flight Radar 24Eurostar scheduled departures	Port schedules
<b>Consumption &amp; transaction data</b>	Individual preferences and retail choices –both directly and indirectly related to transport.	Oyster card derived travel data Credit card spend-data Domestic energy consumption	Petrol prices & Rail fares Loyalty card purchases Restaurant bookings

Notes: reproduced from (TSC, 2015)

### 3.1.2 The collection and sharing of transport-relevant datasets

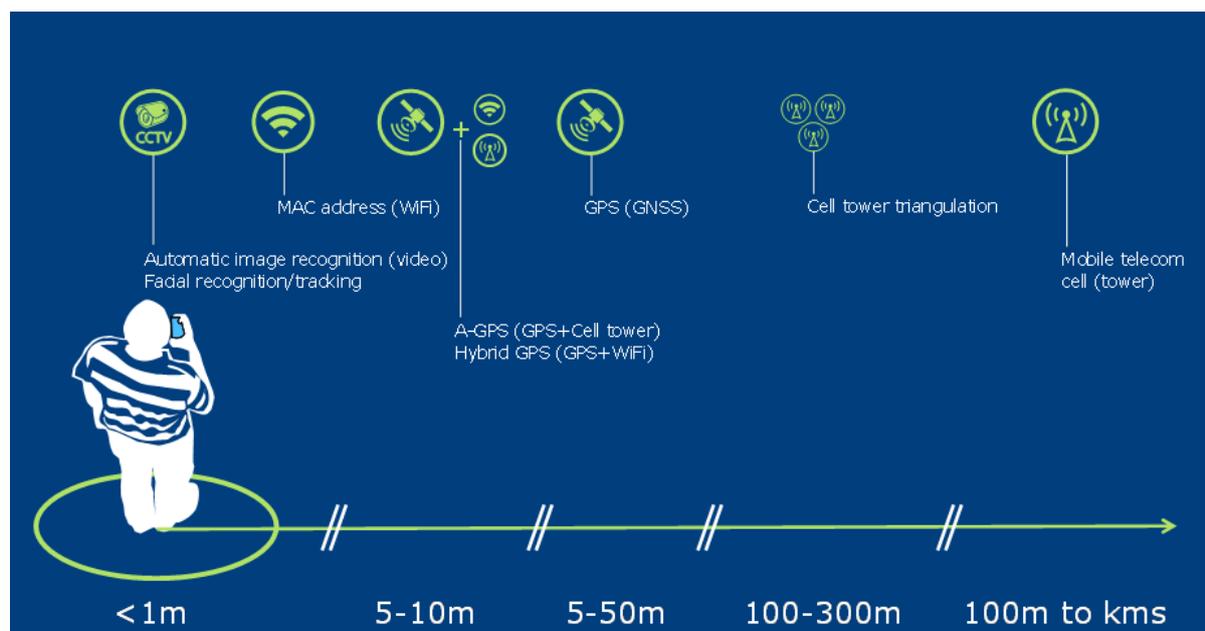
Broadly, data on transport activity is generated via a range of mechanisms: by transport service providers, manual collection, crowd-sourcing (overt and covert) and from sensors (fixed and mobile).

Sensors are responsible for the majority of the relevant real-time data that is currently available, which includes bus and train location data. Sensor-derived real-time data will also increasingly include information from cars (and also vans, lorries) with the roll-out of C-ITS technologies.

A vast amount of mobile location data is currently already generated, collected and utilised by organisations like Google to enhance their services (e.g. mapping and navigation, search, etc.). However, lower resolution mobile information can also be collected based on cell phone tower connections (GSP), and from Wi-Fi network connections. There are already examples of where this approach is utilised, such as in Newcastle through the city's Wi-Fi network.

Figure 3.1 provides an overview of the relative precision of various geo-location technologies. At the moment none of these datasets are open, although limited anonymised datasets may be sourced commercially in some cases. Such real-time information sources are particularly important to enable the full benefits of intelligent mobility (see section 2.2).

**Figure 3.1: Precision of geo-location technologies**



Source: (ITF, 2015)

Data is increasingly made available online, and may be grouped into four categories (TSC, 2015):

1. **Distributed data catalogues:** available datasets from multiple organisations are published via a searchable online catalogue. Datasets remain stored with the data owners.
2. **Individual data catalogues:** a single organisation's data is published via its own online catalogue. Datasets remain stored by each data owner, and may also be accessible via other data platforms.
3. **Consolidated 'Big' data platforms / ecosystems:** available data are consumed and stored ('warehoused') so they can be combined and served up on-demand from a single interface.
4. **Distributed data aggregation platforms / ecosystems:** these provide a gateway into available data from multiple sources using APIs in order to allow for real-time fusion and consumption of datasets.

Reviewing the key public and private sector data providers (e.g. as summarised in Appendix A3 Table) leads to the following conclusions and considerations:

- Data currently being shared openly is usually available at national or major city level.
- Data shared by the public sector tends to be open, but shared in different platforms which hinders its linkage. Conversely, private datasets are not usually easily accessible.
- There is no comprehensive online transport-related data catalogue in the UK.

Our review of the literature and the consultation with stakeholders has highlighted the problem of inconsistent (or sometimes absent) data standards. There is a general lack of consistent application of standards, data formats and structures for many types of data relevant to transport. For example, in many of the open data portals, raw data feeds are presented alongside CSV, XLS and XML files, as well as PDF documents and a range of other forms of data. As a result, the process of finding data can be time-consuming, and results in a lack of a clear view of the data gaps affecting the transport sector. Although the time-consuming nature of finding data is in some cases a one-off and could decrease in the long-term, it can nevertheless pose a significant challenge particularly to new/smaller data users.

Linked to this, responses to our wider stakeholder consultation suggest that new standards may also be needed to improve the collection and anonymisation of data with high spatial and temporal resolution (discussed further in 3.2.2). In particular, new standards are needed for metadata (e.g. the type and format of the information) in order to reduce the high levels of fragmentation of datasets, specifically in the transport industry. These standards are considered important to improve the integration between different datasets. Box 3.2 summarises the role of metadata and provides details of the importance of metadata for developing open data portals.

#### Box 3.2: The role of metadata in open government data initiatives

A key feature of open data government initiatives is to open public sector information with the purpose of maximising its reuse. Opening up datasets requires collecting and publishing datasets on data portals or data catalogues. The main function of data portals is the management of metadata of datasets and the harmonisation of metadata itself. Metadata, which are structured descriptions about the actual data, are a key feature to make datasets more discoverable and accessible. Discoverability and accessibility improves proportionally to the extent to which metadata's quality increase.

Source: (Judie Attard, 2015)

Improving the awareness and use of appropriate, consistent, global standards and common data frameworks is a key priority to help facilitate improvements in the coverage, identification and accessibility of key datasets. This should be a priority for both governments interested in further pursuing open data initiatives, as well as for concerned industry sectors.

An additional issue is the 'Signal Problem', as identified also in (TSC, 2015). Since data is captured from devices connected to wireless networks (mobile phones, sensors), care must be taken when extrapolating observations to the wider population. In particular, communities with limited access to the technologies that make Big Data possible are less likely to be captured (or at least less representative). Provision of services based on these datasets therefore runs the risk of discriminating against certain segments of the population (e.g. those in less affluent or accessible areas).

### 3.1.3 The role of sensors

Transport data is generated from a multitude of sources. These include both data originating from new technologies and more traditional datasets generated by conventional data collection techniques such as surveys, traffic count data and fare and schedule databases. A number of studies cite the importance of sensors in filling data gaps (Rusitschka & Curry, 2016) (IoTWorld, 2015) and a useful illustration of their role can be seen in Figure 3.2 below.

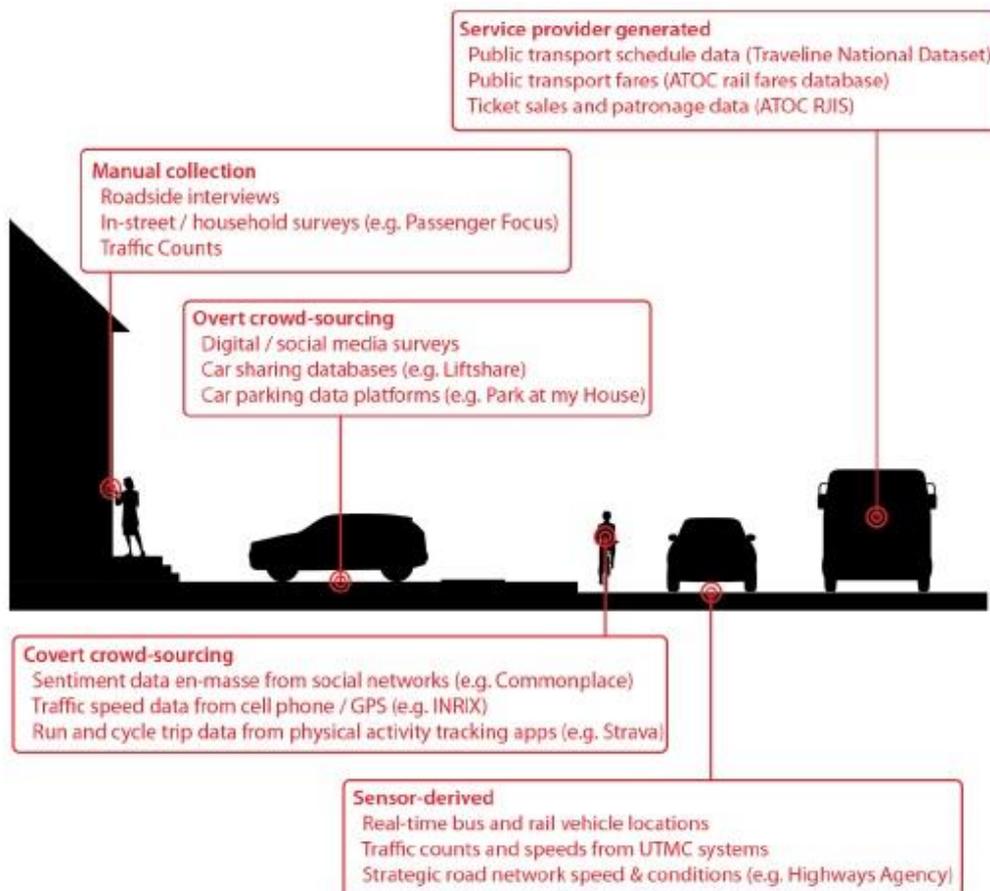
Low power sensors are already being, or will in the near future be, used in the following ways:

- **Infrastructure:** sensors can be embedded into road infrastructure to capture and wirelessly transmit data related to variables such as temperature, humidity, and traffic volumes. Real-time information about the condition of the road can also be provided to facilitate timely maintenance.

Sensors can be applied to other assets, such bridges, e.g. the low power vibration energy harvesting sensors trialled on the Forth Road Bridge (Cambridge University, 2015).

- **Parking:** significant traffic is generated on roads due to the drivers looking for available parking. Sensors in devices like ParkSight (IoTTechWorld, 2015), can be embedded into, or on top of the pavement to collect and transmit data to drivers and parking facility managers, alongside other information such as parking fees. Open datasets with sensor-informed parking availability are also available for a number of UK cities (Venturebeat, 2015a), as well as international examples such as San Francisco (Schneider Electric et al, 2014) and Barcelona (see later Box 6.3).
- **Vehicles:** both OEM-installed and aftermarket devices, such as Delphi Connect, can be installed to provide a range of information, such as the status of the battery, fuel and engine, and also allow for remote actions such as unlocking, or even preheating the vehicle or starting charging sessions (for electric vehicles). GPS enabled devices enable drivers to remotely track their journeys and vehicle location. A range of sensors are also utilised on commercial vehicles to improve supply-chain efficiency and to remotely monitor status and plan maintenance. Vehicles are, in effect, becoming sensors themselves. The Society of Motor Manufacturers and Traders (SMMT) has recently set out guidance for sharing data generated by CAVs to address safety, privacy and security concerns (SMMT, 2017).
- **Public transport:** There are many bus, train and metro/light rail services that use electronic sensors to monitor status, improve transit efficiency and to schedule appropriate maintenance. In addition, such real-time feeds are also used to track the location and provide this information to customers either on electronic display boards or via apps.

Figure 3.2: Mechanisms for transport data collection



Source: Illustration provided in (TSC, 2015).

Many cities in the UK and internationally have introduced smart transport initiatives including increased use of sensor-driven Big Data to optimise their public transport routes, create safer roads, reduce infrastructure costs, and alleviate traffic congestion as more people move into cities. For example, Autolib, an electric-car sharing programme launched in Paris in 2011, uses GPS-enabled sensors inside the vehicles to track their locations. To join the scheme drivers are asked to subscribe in order to obtain a badge which is needed to rent the car from the charging station/rental points available in the city. Subscribers can access a car's dashboard through a mobile app and reserve cars and public parking spaces in the city (Business Insider UK, 2016).

There was some disagreement among the stakeholders interviewed for this project on how necessary the extensive use of sensors might be. Some suggested that much of the data was already available in some form, or could be derived by other means. One example given was the potential for real-time analysis of CCTV data as a proxy for parking availability. However, some stakeholders indicated that, based on their practical experience in the field, there were very significant technological and legal hurdles to overcome with respect to using proxy datasets in place of sensors - particularly due to privacy considerations, which also increased costs. In the end it has often been found to be simpler and more cost-effective to implement a sensor-based solution. Nevertheless, it may be that technological and legislative solutions could be found that would enable easier use of certain proxy data.

For infrastructure, some of the experts interviewed also suggested that some caution should be given to the potential adoption of a widespread programme of sensor deployment, since other more cost-effective (and more quickly deployable) options are likely to be available. For example, it may be possible to utilise information from the sensors present in commercial vehicles (e.g. via on-board smart weight scale technologies) to monitor infrastructure condition. However, at the same time, there could be ownership, confidentiality and cost issues that would preclude the use of such data, whereas data produced by sensors embedded in infrastructure would be owned by the infrastructure operator and therefore be immediately available. In other cases, it is clear that sensors will be necessary to gather specific information (e.g. for analysing air quality pollutant emissions from vehicles) or may have a role in the cross-corroboration of datasets at key points.

Further research is necessary to answer other key questions, such as: how important are sensors for helping to fill different gaps? Is there information on sensor coverage in UK?

### 3.1.4 What are the key gaps identified in relevant datasets?

A number of transport-related data gaps have been previously identified in (TSC, 2015), which we have further expanded upon using stakeholder input as part of the consultation for this project. We asked experts to rank which gaps were the most important to fill, in order to better understand how actions to do so might be prioritised. The findings are summarised in Table 3.2.

**Table 3.2: Transport-related data gaps assessed in this project (red = most important; yellow = least important)**

Theme	Dataset	Services enabled	Ranking*
<b>People, Things &amp; Movement</b>	Automated cycle count data from National Cycle Route network	Infrastructure usage, further investment business cases, maintenance prioritisation	15 (51%)
	Mobile geo-spatially coded data**	Optimised service and capacity planning across networks	14 (59%)
<b>Disruption and event-related data</b>	Database of all UK major events compiled from local authority event licensing and sporting schedules	Transport capacity service planning, smarter traveller information services	12 (63%)
	Real-time condition of roads/bridges/rail assets**	Improved asset management leading to reduced disruption	5 (69%)

Theme	Dataset	Services enabled	Ranking*
<b>Personal Automobility</b>	Traffic speed and count data from urban traffic management control systems and non-strategic roads	Real-time journey planning and updates across all roads	<b>2 (77%)</b>
	Real-time parking space availability for on and off-street car parks	Real-time journey planning and updates for drivers	<b>3 (75%)</b>
	Real-time parking charge data for off-street and on-street car parks	Enhanced journey planning and true journey price comparisons across modes	=10 (65%)
<b>Freight connections</b>	Real-time vehicle (road, rail, air, sea) locations and capacity information	Optimised fleet utilisation across logistics providers	<b>6 (68%)</b>
<b>International Connections</b>	Real-time air and Eurostar departure and arrival time information data feeds	Ground transport and handling optimisation, reduced passenger waiting	<b>13 (60%)</b>
<b>Public and other Transport Services</b>	UK-wide local bus fare data	True journey price comparisons across modes	<b>9 (65%)</b>
	Bus and rail vehicle occupancy data derived from passenger ticket purchase, barriers, electronic ticket machines.	Smarter traveller information services to reduce passenger waiting and avoid overcrowding	<b>1 (79%)</b>
	Real-time bus and rail vehicle locations across the UK	Smarter traveller information services to reduce passenger waiting	<b>7 (66%)</b>
	Data on taxi locations and availability**	Smarter traveller information services to reduce passenger waiting	<b>16 (50%)</b>
<b>Consumption &amp; transaction data</b>	Aggregated historic passenger ticketing data for bus, rail, air, sea modes of travel	Optimised service and capacity planning across networks, smarter journey planning for travellers	=10 (65%)
<b>Important non- transport datasets</b>	Local weather data**	Optimised service and capacity planning across networks	<b>4 (70%)</b>
	Local air quality data**	Optimised service and capacity planning across networks	<b>8 (66%)</b>

*Sources:* The initial list was based on (TSC, 2015), and was expanded with additional information from this study.

*Notes:* \* Ranking is based on a Delphi survey conducted with stakeholders as part of this project. The % figures indicated the average scoring of importance given to different options. \*\* Additional important data gaps identified during this project's stakeholder consultation process.

The level of importance of each data gap is driven in part by the objectives set at city or regional level; all of the identified gaps are important to delivering a fully intelligent transport system; however, concerned entities, such as local or central government and industry, should give priority to those that could deliver 'quick wins', rather than filling gaps that will take a long time to address (e.g. where sensor networks must be extensively deployed).

Our discussions with the different stakeholder groups consulted for this project confirmed that the main gaps identified above occur at different levels. In some cases, datasets are only available locally, without national consolidation (e.g. major events information held by local authorities and transport operators). In other cases, data may exist only in silos (e.g. automated cycle count data in PDF documents, Urban Traffic Management Control traffic flow data), or are not open/available (historic passenger ticketing data). Finally, in some cases the datasets may essentially not exist at all, e.g. on-

street parking bay availability. (In the latter example, some data is becoming available in a limited number of areas through a combination of sensors and crowd-sourcing approaches, such as by Parkopedia (Venturebeat, 2015a)). The solutions/methods needed to address the different gaps will be clearly closely tied to these differences, and so there are no obvious 'silver-bullets'.

Our discussion with stakeholder experts for this study has revealed that real-time data, such as real-time vehicle (road, rail, air, sea) locations and capacity information, is judged to be particularly valuable (see Table 3.2). The lack of availability of key real-time datasets are therefore considered a key barrier to intelligent mobility service provision in Smart Cities (and more widely). For instance, bus occupancy data (one of the top-ranked data gaps) is currently only collected in a few cases for dynamic management of services. There are no strong drivers to collect such data, except in particularly busy regions, as bus operators are generally aware of passenger numbers through boarding data and don't see a real business benefit in knowing specific occupancy levels to inform capacity. Except on particularly busy routes, this level of information is not deemed essential to run the service cost-effectively for operators. However, such information can be helpful to reduce a risk-averse traveller's uncertainty when making decisions on alternative travel options. Conversely, some areas are showing progress; real-time location data availability is improving, but patchy for buses (except in major cities).

A number of data gaps are expected to be closed in the next few years for buses due to provisions in the Bus Services Act (UK Parliament, 2017a) – for example the industry is already working on a solution to make available UK-wide local bus fare data. For passenger rail, real-time location data is already available from (Network Rail, 2017), but is not a fully open dataset (as there are specific conditions placed on its use).

The lack of real-time data is often noted as being a difficult gap to close. On this point, InnovateUK's IoT demonstrator projects (e.g. i-Move and Smartstreets) have previously demonstrated that such real-time data feeds could be released in useful formats. However, these feeds have now been discontinued following the completion of the projects. Further support may therefore be needed in order to take such pioneering projects from pilot phase to wider ongoing application and commercialisation.

There are also difficulties in archiving data to provide a historical record, mainly due to the extremely large sizes of the datasets of interest. Even with the falling cost of storage, there are likely to be budget-related restrictions here unless an appropriate business case can be found. The use of smart automatic data archiving routines will likely be necessary to aggregate older data over increasingly long time-horizons as the data becomes less relevant. This would still allow for useful historical analysis and data mining to derive insight from such data.

Other gaps that were not present in the initial list that were identified in the consultation included:

- **Mobile geo-spatially coded data** are seen as critical for establishing end to end journey patterns, which is normally hard to capture even in an integrated transport system. For example, mobile data would allow real-world walking patterns to be mapped which would be beneficial for travellers. Such mobile location data on the majority of journeys should already exist, but is not currently made available. We found that both the experts consulted and the literature agree that the main challenges to filling the majority of the gaps are not due to technical issues, but rather due to other considerations (e.g. commercial business models, security and privacy concerns).
- **Environmental / weather data** (although some information is available from the Met Office): these datasets are generally considered useful to provide context for analysis of other datasets, for example to enable more meaningful analysis of data generated through infrastructure assets (e.g. via sensors). Moreover, weather data can influence travellers' behaviours, choice of transport modes and traffic flows (e.g. speeds).
- **General traffic flows data<sup>9</sup> (e.g. real-time congestion data), particularly in relation to pedestrians and for taxis private hire data**, where currently even basic data is poor. The recent announcement by Uber that they will be opening up their dataset will contribute to understanding in this latter area (Uber, 2017). There are already international examples of effective systems for the collection of real-time taxi data, for example, the Government of

---

<sup>9</sup> Whilst Google and some other organisations collect such information derived from speeds based on mobile location data, or Sat Nav devices, such datasets are generally not openly available.

Singapore's open data portal (Data.gov.sg, 2017). By collecting and analysing such data, it is possible to predict at what time and locations taxis should go at certain times in the day.

In the second round of the Delphi survey for this study, stakeholders were asked to categorise the nature of the identified data gaps – i.e. whether the gap was due to lack of any data, data being available but not collected, or data being collected but not used/released. Our analysis of the results of the survey are presented below, in Figure 3.3. This seems to suggest that certain real-time data like parking data, bus occupancy and infrastructure condition are either not currently available at all, or alternatively those responding to the survey were not aware of its availability. However, most of the other gaps appear to be due to data not being collected, or collected data not being used (e.g. because it is not openly available). This highlights that there is a lack of awareness about what data is available and/or collected.

### Box 3.3: Key reasons for current data gaps identified by stakeholders

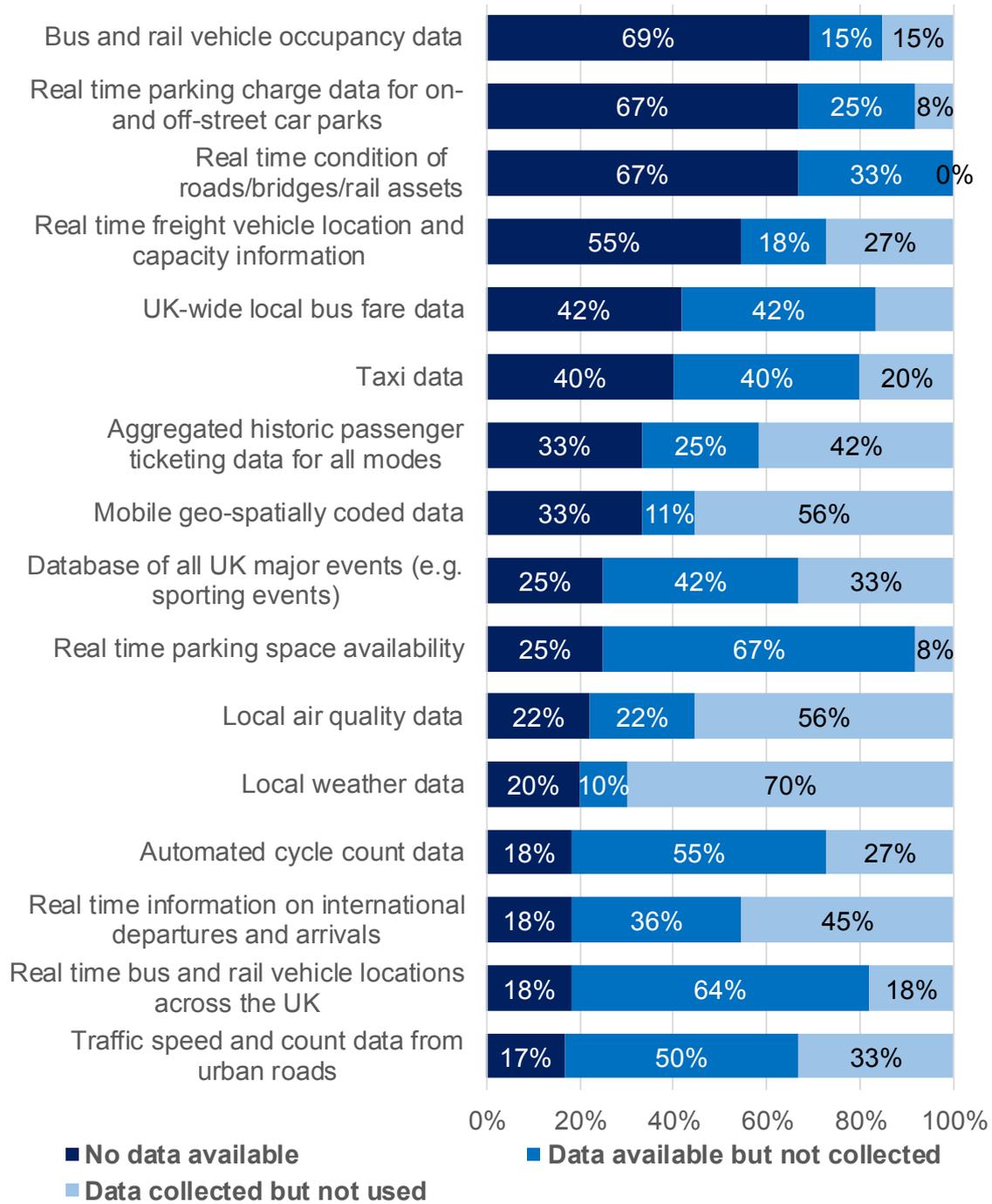
- Lack of a clear business model and commercial sensitivities associated with data use were the most important challenges to collecting/opening up missing data;
- The cost of establishing sensor networks and a lack of knowledge on data ownership were the next most significant challenges identified;
- Privacy concerns (sometime unfounded due to lack of understanding of Data Protection Act legislation) were also noted as an important barrier.

The elements identified in Box 3.3 are discussed in more detail in Sections 4 and 6, and are consistent with the findings from the literature. The views of those consulted were more mixed on the importance of technical issues and the lack of knowledge on data licencing, which suggests that these issues may be more specific to certain types of data.

Stakeholders recognised that there is a general lack of knowledge on issues regarding data ownership and data licencing. Both require further contextualised understanding in order to avoid a 'tendency to make incorrect assumptions'. Stakeholders suggested that addressing such issues would require a change in organisational culture, changes to the Government's stance with respect to supporting data collection, and an improved governance and legal framework on the processes needed to collect open data and commercially sensitive data. Generally, the perception is that there is poor understanding in the transport industry of what Big Data is and how to realise the benefits from it. Similarly, the value of data is poorly understood and systematic methodologies to value data do not exist, leading to data not being shared at all or to it being over-valued.

The options for addressing these issues range from more direct regulatory or contractual requirements through to provision of information and guidance to enable relevant actors to make more informed decisions. It is likely that a range of such more 'hands-on' and 'hands-off' approaches will be needed to address different types of gaps. For example, certain open data requirements could be included in public transport contracts/franchises, whilst at the same time there is a need for active awareness-raising activities. Such activities could provide information from case studies to help demonstrate the business case and potential 'win-wins' and provide guidance on the application of appropriate standards (in particular to address safety, security and privacy concerns). Here, public and private sector transport-related big data/smart city networks could have a key role to play in sharing knowledge and experience – such as the Transport Data Initiative (TDI).

Figure 3.3: Stakeholder assessment on the nature of the identified data gaps



Source: Results of the Delphi survey conducted for this study.

## 3.2 The role of open architectures and innovation platforms

### 3.2.1 What are open architectures and innovation platforms?

At the city scale, the next frontier of Big Data applications will comprise the fusion of different datasets within the transport sector as well as wider urban datasets, collaboration among public, private and community organisations, and ultimately, the creation of predictive capabilities for planning the future of cities. The literature and stakeholder inputs for this project agree on the importance of developing open data architectures and innovation platforms to meet this vision:

- **Open architectures** are software or hardware infrastructures which are designed to allow for adding, upgrading and swapping components in an open and accessible way, so that potential users can see and operate with all part of the architecture without property constraints. The aim is then to foster a virtuous cycle of opening up data access, developing new analytics, creating new business models, implementing practical applications, showcasing best practices, enhancing data access and use standards, etc.
- **Innovation platforms** are open application programming interfaces, which can be created for public bodies and regulated industries to enable innovative sharing and use of real-time public and private data, to inform smart-cities strategies and innovative mobility solutions. Depending on the type of data and expected solutions, innovation platforms can be developed in different ways. They can also be seen as opportunities to attract R&D funds through developing robust business cases and fostering trust, under clear legal safeguards.

There are different methods available for creating platforms that enable the exchange of data. Experts recognise the notion of a 'data infrastructure' as being key to allow for access to data (Royal Academy of Engineering, 2015). Specifically, in the context of Connected and Autonomous Vehicles (CAV), innovation platforms largely consist of online places where anonymised vehicle-generated data can be aggregated and shared between contributing parties (e.g. vehicle manufacturers and public authorities) that have defined reciprocal agreements concerning their ability to access particular types of data. According to (SMMT, 2017), innovation platforms that are designed to share data to improve traffic management and safety should be public-held neutral platforms.

Although the process of opening up data is important, it is the effective fusion of data from different sources and interpreting such data in the context of the transport problem in question that creates real value. In cities, there are multiple creators and users of data (e.g. business, citizens and local government) who need to collaborate in a network (Policy Exchange, 2016); (ARUP, 2016). This allows different actors to unlock the potential of opening up data while preserving a competitive and attractive data economy ecosystem (RTIG, UTMC, 2015).

There are strong arguments for developing open architecture and innovation platforms following a city-based approach (Policy Exchange, 2016), illustrated by both the literature and our stakeholder consultation findings. Promoting an open approach to data architecture and platforms also appears to be beneficial for addressing challenges faced by smaller cities, since the use of common open architectures across wider areas facilitates the linking of datasets to generate critical mass - making them more attractive for app developers and other commercial organisations.

At the same time, it has been suggested in the wider literature that investments to promote the use of Big Data through city-level open data architectures may put small-medium local authorities at a disadvantage. Smaller UK cities have less in-house capacity to open up and maintain the datasets they hold, moreover their smaller geographical scale and resultant smaller datasets make them less attractive propositions for third party developers. For example, while the London Datastore has been successful in attracting third party development and use of the open data, in other cities such as York there has so far been less use of the data made available. For smaller towns and cities, this situation would likely be worse but this could be mitigated by coalescing these smaller datasets into larger more attractive ones based around the emerging city regions or sub-national transport bodies, (such as Transport for the North).

As a result, there are now cases where open architecture data initiatives in smaller cities evolve to regional-level initiatives. For instance, the ‘Leeds Data Mill’, evolved into “Data Mill North”<sup>10</sup>, reflecting a desire to expand the data store across the region and break down traditional local authority boundaries. Therefore, using open data architecture to consolidate datasets at a regional, and ultimately national level, could be the most effective approach to unlock the potential of data sharing platforms, which is ultimately beneficial to transport and other urban planning related operations.

While open architectures and innovation platforms seem to hold much promise, how feasible are they in practice? The following sections explore the different dimensions of feasibility and applicability in different contexts: technical, economic, and legal/political.

### 3.2.2 Technical feasibility

The **technical feasibility** of open architectures and platforms concerns the specification of interfaces and definition of standards across multiple areas (including data taxonomies, data integration, privacy, security and accessibility). Given the large number of actors that need to be involved in open platforms, standardisation is a key factor that influences the extent to which data can be shared and linked – and consequently the level of innovation that can result (ARUP, 2016), including potential uses well beyond the applications and sectors that data were collected for (Royal Academy of Engineering, 2015).

As the IoT develops, new networks and standards are appearing, making it difficult to judge what connectivity standard should be selected for IoT devices (Royal Academy of Engineering, 2015).

The international initiative OneM2M<sup>11</sup>, which promotes the use of global open standards for IoT and machine-to-machine technology, has made the following recommendations in consultation with city managers worldwide (OneM2M, 2017), which are also consistent with our broader findings:

- **Make use of existing horizontal platforms for new deployments in different areas:** rather than rolling out a dedicated platform for each use case, cities should leverage existing networks to support multiple applications and enable data re-use. This enables cities to sidestep issues of lack of scale – for example, experience shows it is much more efficient to use a platform to support multiple uses such as traffic management, street parking, waste management and street lighting.
- **Adopt global standards:** this would give city managers more freedom in choosing the appropriate vendors according to their needs.
- **Promote access to ‘open’ data in standard formats:** the potential of smart cities open architecture platforms relies on the synergies and integration between different datasets and applications – this requires developers to have access to open data in standard formats. Use of open standards is also thought to accelerate growth and reduce deployment costs by around 30%.

The importance of global standards was also confirmed by the results of the Delphi survey, with around a third of respondents saying that global standards are very important for developing open architecture and innovation platforms. However, the majority of those consulted were not sure whether the UK standards already available were sufficient, which suggests a lack of awareness on the topic.

Indeed, the research conducted for this study suggests that the topic of data standards is still being researched and that there is a lack of clarity regarding the data standards currently available, which sectors should they be applied to, and whether there are differences in the types of standards to be used in different sectors. For example, it was not possible to find a comprehensive data standards summary. Research by the Urban Transport Group (Urban Transport Group, 2016) provides some examples of data standards which are specific to the transport sector, including:

- ATCO-CIF standard, a notable early example, which has been subsequently superseded by the TransXChange format, TXC.

<sup>10</sup> <https://datamillnorth.org/>

<sup>11</sup> OneM2M is the global standards initiative that covers requirements, architecture, API specifications, security solutions and interoperability for Machine-to-Machine and IoT technologies. OneM2M was formed in 2012 and consists of eight world preeminent ICT standards development organisations.

- The European SIRI standard for real time information.
- Google's GTFS (General Transit Feed Specification), which has become the de facto standard for public transport network data in the US and many other parts of the world.

However, given the importance of standards in facilitating enhanced data sharing, further research and progress in this area are expected. For example, the British Standards Institute (BSI) recently published a study on Smart Cities' standards, which provides a guide to establish decision-making frameworks for sharing data and information services (BSI, 2017).

The implementation of standards and the management of the platform requires an understanding of the data needs and potential uses, who are the players and how they can benefit (ARUP, 2016). Strong data governance is also important to address issues related to data quality, data ownership and interoperability. The case study on FIWARE (Box 3.5) further highlights the role of data governance in ensuring success, and in being able to address concerns over the commercial sensitivity of the data they manage.

### 3.2.3 Economic feasibility

Securing the **economic feasibility** of open architecture platforms means clarifying the value of Big Data through the monetisation of commercial data. This can act as an incentive to data sharing and would, as a result, provide the conditions for encouraging the development of innovative services and further competition among data providers.

Research conducted on Smart City opportunities for London (ARUP, 2016) demonstrates how the potential benefits of data are poorly understood, in both the private and public sectors. The Greater London Authority was able to recognise the relevance of attributing value to data while ensuring the success of new business models through the monetisation of data. In doing so, a financial framework for the city data economy can be set in order to define which investments, costs, risks and revenues are associated with the exploitation of both open and commercial city data. In addition to the commercial value of data, there is an understanding that data can also be exploited to release their social value, which results in better strategic choices.

In line with our discussions with stakeholders, competition between services providers (i.e. fear of losing competitive advantage) has been identified as a crucial barrier in the development of open architectures and platforms. Nevertheless, in some cases it is demonstrably possible for private companies to open their data and work on open systems, while retaining and even gaining competitive advantage – concrete examples of this include Thomson Reuters, Arup and Syngenta (Open Data Institute, 2016). These companies found that opening up their data and providing the necessary licences could allow for other benefits, such as:

- Maintaining leadership in competitive data markets.
- Keeping costs low for the business and for customers.
- Staying agile in shifting environments.
- Collaborating with clients and competitors to address big challenges.

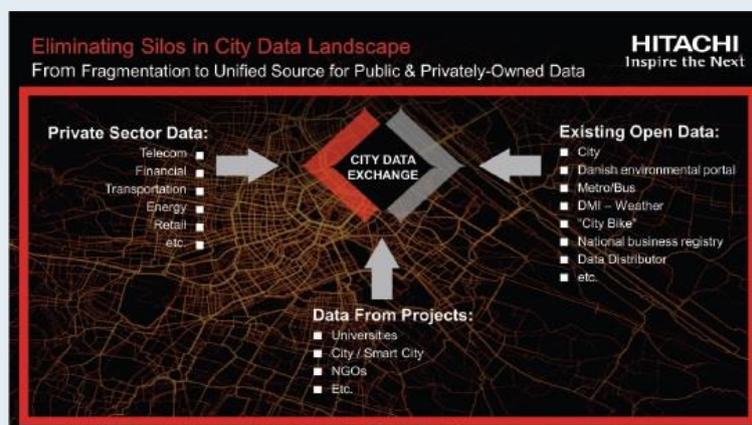
Addressing and solving issues regarding concerns over losing competitive advantage can be addressed through a combination of raising awareness of commercial case study examples where the net benefits have been demonstrably proved (i.e. there is a win-win), and developing suitable guidance on open data frameworks, applying appropriate common standards, to address key concerns.

Private data owners also do not normally share data because there is not a clear way to control how the data can be used. However, the Copenhagen City Data Exchange (see Box 3.4) is an example of a trading platform working on developing a framework to enable this (CDE, 2017).

**Box 3.4: International Examples of Smart Cities: Copenhagen****The City Data Exchange**

Copenhagen has the objective of becoming a smart, carbon-neutral city by 2025. As part of the overall approach to achieve this, a range of smart city programmes have been initiated across all sectors, including transport. One of the key steps taken in Copenhagen was the introduction of an integrated data service – The City Data Exchange. The objective of this novel solution with a new business model, is to help eliminate the data silos that have historically been present in such smart city programmes. A core component of the data exchange is the integration of data from private companies and open data from public authorities and the creation of a marketplace to allow data suppliers to find data customers. Key to the development and functioning of such marketplace was the establishment of an organisation external to the city council<sup>12</sup>, which is responsible for providing data analytics and support to third-party developers to use data in their applications. Moreover, the organisation is responsible for facilitating processes to allow businesses and researchers in to access data available on the platform.

The City Data Exchange was launched in May 2016 and it still appears to be still under an implementation phase, which means that not enough data is available yet to assess the extent to which the initiative is being successful. Following to the launch of the City Data Exchange initiative, the city of Copenhagen has also become engaged in further R&D activities by being one of the consortium partners in the current EU-funded project SELECT for Cities (2015-2018) (SELECT for Cities, 2017), which aims to design, research and develop a data-driven, Internet-of-Everything (IoE) platform for European cities to enable large-scale co-creation, testing and validation of urban IoE apps and services. The project funding is being used to implement pre-commercial procurement (PCP) processes, which will procure the research and development of new innovative solutions capable of realising the goal of having the cities as a large IoE Lab and putting it into practice.



Source: (Hitachi, 2017), (CDE, 2017).

Specifically regarding the transport sector, research has explored new mobility solutions adopting data sharing platforms, open architecture, and what the benefits resulting from applications using both the free and paid version of data are (TSC, 2015a). The identified challenges and issues can be addressed by creating a carefully designed and managed transport data exchange. To do so, platforms ought to define licencing permission and reward mechanisms. This, for example, could encourage, on the one hand, companies to increase the value of data by providing enriched, analysed data while providing a service and, on the other hand, organisations holding data, would benefit from the service and learning achieved in return for the data being made available.

Only these conditions can develop a market place that is beneficial and would allow for the development of services, which provide analysed and enriched data that help achieve transport benefits (e.g. preferred traffic routing).

In such a market place, the exchange between players - some offering data components (for free) in return for knowledge being developed - would encourage new participants to contribute and allow transport operators to benefit from the learning and insights generated by the development community (TSC, 2015a). In other words, the authors argue that a rewarding mechanism for data exchange should

<sup>12</sup> Established and run by Hitachi Insight Group.

be based on the unique value of the asset held (e.g. skills, IP, access to travellers, type of data), which should define the value distribution.

### 3.2.4 Case studies

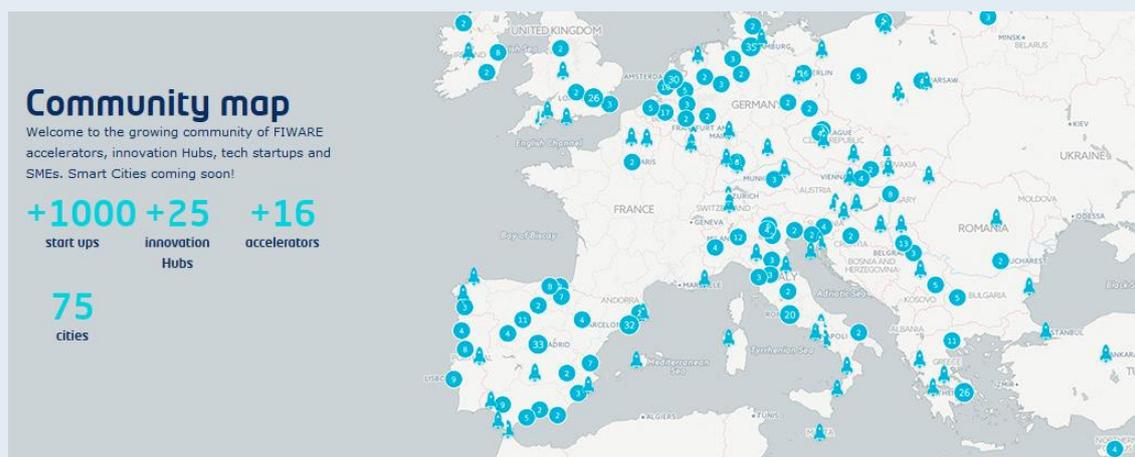
The importance of technical, economic and legal feasibility, along with the associated challenges, can sound daunting – but the following case studies demonstrate that such issues can be successfully overcome. Both case studies were developed on the basis of in-depth interviews with key stakeholders<sup>13</sup>.

The first example, FIWARE, is an open platform that provides a set of public and royalty-free APIs (Application Programming Interfaces) that ease the development of smart applications across multiple sectors. Its services span a wide range of areas, such as making it easier to connect to, process and analyse Big Data and real-time media, as well as allowing incorporation of advanced features for user interaction and many others. Key lessons and success factors are summarised in Box 3.5.

#### Box 3.5: FIWARE – lessons learned and key success factors

##### FIWARE ([www.fiware.org](http://www.fiware.org))

FIWARE started as an initiative supported by the European Commission's FP7-ICT framework, launched in 2011 and completed in 2014. It continued providing services to start-ups and, at the end of 2016, the FIWARE Foundation was established to oversee the governance of the components and of the technology. Their work is currently being extended outside Europe.



Source: Image provided by FIWARE ([www.fiware.org](http://www.fiware.org))

#### TECHNICAL FEASIBILITY

##### “Strong data governance of the platform is crucial for success”

- **The openness of architectures and of components provides the freedom to work in the most effective and efficient way:** By implementing open and transparent use of technology and data, the platform eases concerns regarding the data use.
- **Adopting universally recognised standards is key to the success of the platform:** FIWARE uses OMA NGSI (Next Generation Service Interface) standards, in line with ETSI (European Telecommunications Standards Institute) guidance. The use of standards is also relevant for security, as it enables the definition of the role management of the user.
- **Role-based access control has enabled FIWARE to address concerns over the commercial sensitiveness of the data they manage:** The exchange of commercially sensitivity data can be a major hurdle for platforms like FIWARE. In their case, FIWARE did not encounter any issues regarding the commercial sensitivity of data used. The platform's role-based access control – a data

<sup>13</sup> Both case studies provide limited evidence, especially in regard of financial feasibility, due to the limitations imposed by the data gathering exercise (e.g. the bulk of information was gathered through interviews but financial data were not made available) and the timeframe of this study.

management setting which defines data access depending on the user status (e.g. administrator, external user, data owner) – provides a high level of security in data sharing that minimises these concerns. This element is considered very valuable when dealing with public sector organisations (i.e. especially those that publish information related to public services, fleet control, private information associated to services).

### ECONOMIC FEASIBILITY

#### *“R&D Funding enabled FIWARE to test and develop tools and services”*

- **The success of FIWARE’s business model was the ability to address a need in the market and the quality of its service provision:** The FIWARE platform provides cloud-based resources for free for organisations in the embryonic stage when they are not generating profit. This results in a clear reduction (if not a reduction to zero) of CAPEX for these organisations. When companies start generating a profit, they are required to use a commercial solution. FIWARE is registering a significant number of companies asking to be moved out from the cloud-based service into a commercial solution (1,000 start-ups across two years).
- **Funding for R&D was a key enabler:** FIWARE was initially funded by the European Commission’s FP7-ICT framework. This allowed the development of a commercial product at very low costs, which can now be made available as an open service.

### LESSON LEARNED – Summary

The availability of R&D funding was identified as a key success factor in developing successful open data architecture platforms as it allows projects and products to be piloted before they become commercially viable. Investing in R&D means providing the right environment that favours start-ups testing and developing new solutions. Such investment is expected to have a positive economic return for the wider economy (e.g. creating jobs, promoting innovation and competition).

*Source: developed in consultation with the lead organisations.*

Box 3.6 outlines the lessons learned in the case of Immense Simulations, an SME that builds large scale simulation models to understand and support mobility systems. It is still operating in a R&D environment but has also done commercial pilots and its business model is evolving to be simulation as a service. Their business model adjusts to the situation: it can be based on direct commercial transactions, or shared creation of open resources.

### Box 3.6: Immense Simulations – lessons learned and key success factors

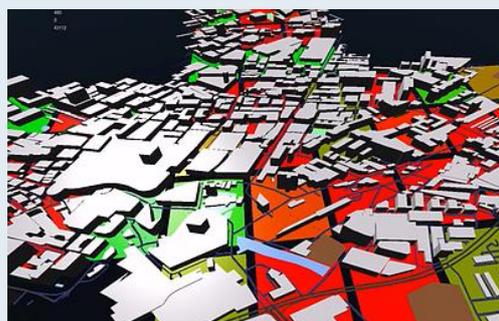
#### Immense Simulations

Immense Simulations was set up in September 2015 as a spin-off from the Transport Systems Catapult. It has been operating in the UK but it is considering expanding internationally.

### TECHNICAL FEASIBILITY

#### *“Open and collaborative approach is the key enabling factors for Immense Simulations to tackle technical issues”*

- **Information on data sources and working closely with data owners are key enablers:** it is crucial to have comprehensive information on the owner and location of the datasets. Immense Simulations find that metadata is missing in many cases and so they have worked closely with data owners to minimise this issue. The reliability of API feeds and detailed API data description are also vital to facilitate work on the platform. There is a quality assurance issue associated with this, but there are already several organisations working on cleaning and standardising data feeds (e.g. TransportAPI and ITO world). Data owners are also expected to contribute to an improvement of quality assurance over time.



*Source: Image provided by Immense Simulations*

- **The available standards should be used where possible, but more work is needed to develop them:** the organisation uses existing API standards whenever possible (GTFS (General Transit Feed Specification), TfL-based standards, working with ODI). In their view, the available standards are suitable in part - some transport data is standardised, but usually only on a regional basis.
- **Security concerns are a barrier:** the organisation aggregates data and builds digital simulation assets, which requires combining different data sources that can expose sensitive details. On the other hand, the organisation would like to have access to data that is sensitive to improve its models but there are privacy considerations and pure data security issues that constitute a barrier.

#### ECONOMIC FEASIBILITY

*“Immense Simulations adopts a revenue sharing model to clarify the monetary value of data”*

- **The clear identification of gaps in datasets and the implementation of a revenue sharing model has helped to overcome uncertainty regarding the monetisation of data:** many organisations have valuable data but do not have a clear idea of its value or how to monetise it which makes them less likely to share them. This can limit Immense Simulations’ access to datasets. They have addressed this issue by blending open and private data to narrow down the datasets they actually need as well as proposing revenue share models to their partners.
- **Opening up and standardising data can reduce the costs of using data:** on-boarding datasets constitutes a large overhead and any tools and processes that standardise and streamline that process are important. Immense Simulations believes that as data becomes open and interfaces are standardised the costs should decline.
- **The R&D phase is important for experimenting, engaging with other organisations and being flexible:** innovation grants and collaborative R&D initiatives can help to create data sharing mechanisms.
- **Having fair and contractual mechanisms in place with various suppliers will be key to ensure the long-term success of the organisation.**

#### LESSON LEARNED – Summary

Although it is difficult to provide an exact indication, in financial terms, Immense Simulations’ approach of identifying revenue sharing model between different parties wanting to share and access data appears to be successful. Key to the development and functioning of such a model is the willingness to work collaboratively. However, it is worth noting that R&D resources were key to the development of such an approach.

*Source: developed in consultation with the lead organisations.*

### 3.2.5 Public acceptability

**Public acceptability** relates to the need to demonstrate the benefits of the use of Big Data to the wider society.

Increasing public acceptance regarding the protection and privacy of data is seen as a key step to build and maintain trust in open data architecture and platforms. Several of those interviewed during this project commented on the sensitivity the public have towards sharing their data, though it is thought that this is less so for the ‘new millennial’ population (see Appendix 1). Still, a large share of the British public is uncomfortable with the idea of sharing personal data, even though many do so (perhaps unknowingly) via social media sites and internet-based services (ARUP, 2016). Cyber security (in relation to both datasets and IoT devices) is also an important consideration, however this topic was only raised by very few of the stakeholders interviewed. There is likely to be a trade-off here between privacy and security and more bespoke functionality that will need to be understood and managed.

Data collected in cities may include sensitive information, such as: transport historical data which may reveal user personal details and transport usage patterns, CCTV footage data (videos and images), mobile geo-spatially coded data (collected via Wi-Fi city networks), and smart metering data (which might reveal information such as whether a household is occupied or not). Even non-personal datasets may be sensitive – for example, data on air quality or noise could potentially have a negative impact on house prices.

Associated with the public acceptance of data protection, privacy and security issues, there are also concerns (actual and perceived) by organisations that hold or own data which mean they are reluctant in opening or sharing such data. The majority stakeholders consulted indicated the key reasons why data is not shared, or is deemed sensitive, included concerns relating to privacy, confidentiality, legal implications (real or perceived) and a variety of potential impacts on competitiveness or reputation. In other words, the fear of misusing data which could in turn create liabilities is considered an important challenge to the opening and sharing of data, which could benefit from the development of open data architecture and innovation platforms.

The solution to these challenges lies in demonstrating the benefits of the use of data to society, while ensuring the privacy of such data is protected (ARUP, 2016). In this context, city authorities need to provide the right leadership by addressing issues relating to legal, regulatory, organisational and technical challenges (TfL, 2015); (Bulger et al., 2014). Some stakeholders suggested that this could be achieved through using experience from pilot studies to inform best practice. They also acknowledged that universities have a role to play in some areas (e.g. open data pilot systems), and governments should set a clear strategy but industry should take the lead. Nevertheless, it is clear from the comments received that collaboration is needed amongst all interested parties, at the local or national level, depending on the initiative being developed.

There are already some examples from the freight logistics sector illustrating how private data can be shared. For example, Nestlé and United Biscuits collaborate on distribution, sharing vehicles to deliver their products (Robinson, 2015); this usually entails sharing data on load factors and destinations to facilitate shared use of trucks and depots, thus enabling a higher load factor and more efficient resource use (more detail on how this works in practice is not currently available).

### 3.3 Research gaps and potential solutions for Big datasets and open architecture and innovation platforms

The key research gaps and possible actions/solutions for the availability and use of Big Datasets and on open architecture and innovation platforms are summarised in the following Table 3.3.

**Table 3.3: Key understanding/research gaps and potential solutions to address Big Data gaps**

#	Key gaps and potential solutions	Source
	<b>Gaps in understanding requiring further research</b>	
1.	A more <b>detailed analysis of the degree/nature and importance of different data gaps</b> is needed, to build on the preliminary analysis in this scoping study. This work should also seek to understand the different facets of the individual gaps (i.e. format, geography, data quality, commercial or privacy issues) so that more specific solutions can be proposed to address them. The first step in exercise should identify the specific intelligent mobility solutions the data gaps will enable (building on previous work in this area by Transport Systems Catapult), to help prioritise them.	(1)
2.	More <b>research on sensor deployment</b> potential is needed. A better understanding of the importance of sensors in helping fill particular data gaps (e.g. vs possible alternative means) and obtaining information on the level of sensor coverage in the UK would be useful. Other relevant questions include the technology issues around sensors, identification of regulatory challenges and proposals to overcome them.	(1)
3.	The transport-related and general Smart City benefits associated with the development and provision of open data architectures and platforms, in particular at city level, are clear in the literature. However, <b>a clear gap is recognised in the extent to which small-medium cities can benefit from the open-by-default model</b> . Due to austerity and geographical scale reasons, small-medium cities fail to recognise the benefit of opening up data. More research on how <b>open data architecture and platforms should also focus on enabling and unlocking potential at the regional and national level, not just a city level</b> , is recommended.	(1), (2)

#	Key gaps and potential solutions	Source
4.	There is a need to review and assess if the existing standards and guidance available for Big Data, IoT and Smart Cities are fit for purpose and whether they are sufficiently visible, and applied consistently. Understanding whether better guidance or a regulatory approach is preferable in different areas to ensure compatibility and to break down challenges to sharing, indexing and aggregating datasets is also required. Standards for meta-data have been highlighted as particularly important to overcome transport data fragmentation, which should also be investigated further.	(1), (3)
	<b>Potential solutions</b>	
5.	In the UK, a substantial amount of survey-based data is still collected manually, and the approaches to collation and digitisation of such hard-copy based datasets are not ideal. Other countries, such as Finland (and in particular Helsinki), have moved almost entirely to digital collection of data, which speeds up collation of this information. A <b>UK-wide move to digital forms of data collection, by both government and industry</b> (e.g. as is already done by the commercial freight distribution sector) would facilitate the collection, collation and sharing of important datasets.	(1)
6.	<b>Promote the value and importance of Big Data and IoT</b> in the public sector and commercial spheres through collaboration between government and industry, in particular to ensure key decision-makers understand the potential benefits of Big Data and open data, the business case and business models, i.e. "Growing the pie" rather than "growing the size of the slice" in collaboration.	(1), (2)
7.	Provide <b>improved coordination between, and advice to, public sector organisations</b> , along with appropriate data collection and publishing tools, to open up, maintain and exploit core datasets. This should build on the work of existing networks, such as the Transport Data Initiative (TDI) with its network of local authorities and industry specialists. This should include the promotion of open data based on ODI good practice.	(1), (2)
8.	Building on the previous action, work with local and regional authorities to develop interconnected (or aggregated) <b>open data portals using consistent open architectures and appropriate data standards</b> (e.g. IoT Hypercat, see also Section 3.2). The aim here is to facilitate integration of current (and potential future) data silos and generate sufficient critical mass to attract innovative use of these datasets to deliver intelligent mobility solutions.	(1)
9.	Building on the previous action, <b>incubate, publicise and maintain an online transport data catalogue for the whole of the UK</b> (and potentially include international links), grouping local/regional feeds. This should include both open (free) and available (licensed/paid-for) datasets. Potentially this could build off the IM Data Index ( <a href="http://imdata.co.uk">http://imdata.co.uk</a> ) developed by TSC.	(1), (2)
10.	Ensure all public sector contracts (e.g. with transport operators) that generate relevant data include a <b>mandated requirement for the collection, and open sharing of key datasets according to ODI best practice</b> .	(1)
11.	Work with UK transport operators, agencies, and local authorities to <b>broker open releases of as many core transport datasets as possible</b> , specifically focusing on the data gaps identified in this report.	(1), (2)
12.	Once the identified data gaps have been better characterised and potential strategies to address the gaps have been developed (see research gaps above), support should be provided to <b>conduct suitable pilot studies aiming to develop and implement/test solutions to address the gaps</b> . Business case reports should be developed for successful deployments and pilot projects with further bridging support provided where necessary for wider deployment and commercialisation.	(1)

#	Key gaps and potential solutions	Source
13.	To improve the <b>technical feasibility of data architecture and platforms</b> , further clarification on standardisation of data is needed. Initiatives aimed at raising awareness of these issues, by both government and industry, are seen as useful to improve the development of open architecture platforms (e.g. promotion of global standards by the OneM2M initiative in consultation with city managers and industry). Moreover, as demonstrated by the FIWARE case study (see Box 3.5) action should be taken to improve data governance.	(1)
14.	To enhance the <b>economic feasibility of data architecture and platforms</b> , further research and pilot projects, with the aim of testing new solutions and approaches of revenue sharing models between different organisations sharing data are recommended. R&D opportunities should be defined by donors in consultation with both academia and industry. As demonstrated by FIWARE (see Box 3.5) and Immense Simulations' work (see Box 3.6), R&D investment, on the one hand, and new revenue sharing models on the other, are considered successful approaches to enhance sharing of commercially sensitive data between different organisations while promoting collaboration and competition.	(1)
15.	To improve public acceptability of data sharing to enhance the <b>feasibility of data architecture and platforms</b> , actions should be taken to strengthen the leadership, organisation and collaboration between interested parties. Government, industry and academia play equal roles in this and should further collaborate. Specifically, government is thought to have a key role in addressing regulatory and legal issues concerning data opening and sharing. These issues must be addressed in order to improve public acceptance.	(1)

Sources: (1) Based on consultation with experts and analysis of other literature by Ricardo for this study; (2) identified in (TSC, 2015), (3) identified in (Department for Transport, 2017).

## 4 Possible business models to accelerate the Big Data applications in the transport sector

### Box 4.1: Summary of key findings on business models

- The business models that derive from Big Data are very diverse and depend on the type of data, service and end user. Part of the complexity in this area is because there is no 'one size fits all' plan - instead, each organisation must determine how to use Big Data for their greatest benefit.
- Although some roles – such as those for app developers – are highly visible and attract great attention, there are multiple other roles for businesses and governments, and opportunities continue to grow as technology advances.
- The challenges in developing new business models are the same as the general challenges associated with using Big Data discussed earlier – principally around commercial sensitivities, lack of understanding of the value of data, privacy concerns and lack of skills. Addressing these challenges will smooth the way for innovation.
- The fact is that the emergence of new business models is a messy process that is difficult to predict. However, there are some steps that can be taken to facilitate and guide the process:
  - The potential for development of innovative new products and services depends on the availability of diverse, reliable and relevant datasets (with real-time data on public transport being highly utilised). Governments are already helping to kick-start the innovation process by making their data open and guaranteeing its future availability – this work is important and needs to continue, along with activities to raise awareness of the availability of the data and benefits of its use.
  - Few cities take an active role to maximise the value of the services that private actors are creating. To do this, collaboration and consultation with a broad range of actors – including industry and academia – is recommended in order to identify unmet needs and better understand what datasets should be prioritised for publication/collection. The findings could also be used to inform strategic investment in sensors and operational technologies needed to fill identified data gaps.

### 4.1 Taxonomy of business models

The traditional “value chain” concept treats information as a supporting element, rather than a source of value itself. In the context of new business models made possible by Big Data and the IoT, it is more relevant to focus on the key roles that different organisations can play in data-driven value chains.

The literature varies very widely in its classifications, but there seems to be broad agreement on several main types of activity, as shown in Table 4.1. Such distinctions are useful to structure the discussion and to identify the potential opportunities and the pathways to their implementation.

**Table 4.1: Key roles and business models**

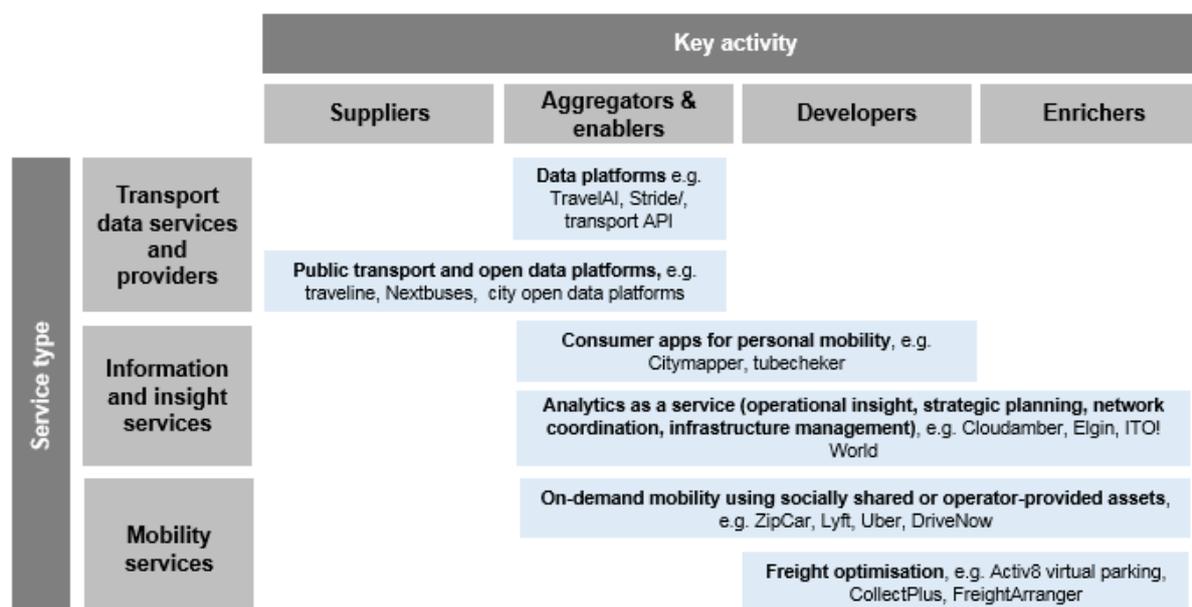
Role / activity	Description	Typical revenue model
<b>Data suppliers</b>	Organisations that publish their data for other organisations to use – this may be done freely (e.g. government open data stores), or the data can be sold to third parties (e.g. telecoms operators selling their data to Sat Nav firms). Data may be collected as part of day-to-day activities, as well as for specific purposes.	Often, data are freely published (with a basic model financed by investors or advertisers). There may be additional charges for premium access or large enterprises. Benefits to public and private sector bodies also arise indirectly, e.g. through new / improved relationships, customer loyalty and access to a wider range of datasets that can in turn support new services.
<b>Aggregators and enablers</b>	Aggregators create value by collecting and aggregating data from a number of different sources (free and proprietary), and package it for resale. Enablers have a role in assisting with analysis providing added value insights (enablers). These are organisations that facilitate the supply, publication, and sharing of data. Most often, companies will play a role in both aggregation and enabling (e.g. aggregating data and hosting it for use or providing processing/analysis services)	Some data marketplaces charge suppliers to publish their data, or earn revenue from advertising. “Freemium” pricing can be used (free basic access, charging for premium data/services). Insights are provided as value-added services to businesses, consumers and governments through subscription fees or pay-per-use pricing.
<b>Developers</b>	Organisations that design, build and sell web, tablet or smartphone applications. Target users can be private consumers (e.g. personal transport planning services such as CityMapper) or commercial users (e.g. freight exchanges, such as FreightArranger). Data sourced for these apps can be open, crowdsourced or proprietary.	Apps are usually provided to consumers free-of-charge, (potentially with freemium features or supported by advertising). Often, user trade their personal usage/location data in exchange for the free service.
<b>Enrichers</b>	Make use of data (internal or external) to inform strategic decisions, improve services and/or refine processes. For example, using data generated from sensors embedded in products to improve quality of service. For example, RAC are starting to use telematics to improve interaction with customers and exploring opportunities for predictive maintenance <sup>14</sup>	Revenues do not come directly from data, but through cost savings, increased sales or premium rates for high quality products/services.

Source: Adapted from (Deloitte and ODI, 2012); (TSC, 2015); (Schroeder, 2016); (Hartmann et al, 2014); (Oxford Internet Institute, 2014).

There is a large degree of interconnectedness, and many successful applications (elaborated further below) are characterised by some level of overlap between the different roles as shown in Figure 4.1.

<sup>14</sup> <http://www.computerweekly.com/news/450403252/Case-study-How-the-RAC-is-digitising-roadside-assistance>

Figure 4.1: Examples of service types and overlaps between different roles



Source: Adapted from (Deloitte and ODI, 2012); (TSC, 2015)

The business models that derive from Big Data are very diverse and depend on the type of data, service and end user. Part of the complexity in this area is because there is no 'one size fits all' plan - instead, each organisation must determine how to use Big Data for their greatest benefit. The literature and opinions from stakeholders on business models offer a starting point to understand the potential opportunities and gaps, but the information is highly fragmented and has led to different understandings – especially because few real-world examples fit neatly into definitions.

It is notoriously difficult to predict the next disruptive business model that will change an industry. The following sections therefore aim to give a flavour of the diverse range of possible roles for organisations in the transport sector, trends, challenges and opportunities. Despite the degree of overlap often seen between the roles, the example firms have been chosen based on their primary activities, since it is usually the case that firms are particularly active in certain areas.

## 4.1.1 Data suppliers

### 4.1.1.1 Examples and applications

Many organisations in the public and private sectors hold data that has value to third parties, who find interesting and unforeseen ways to analyse the data and extract valuable insights. A number of actors operate in this space:

- **The public sector** has an incredibly important role here, as they have a wealth of static and real-time data that could be released openly. Examples of well-curated datasets include: Glasgow Open Data store and Nottingham Open Data.
- **Private firms that collect data as part of their daily operations** are also active in this space; for instance, telecoms operators sell anonymised raw data on their customers' movements to Sat Nav operators and/or retailers (Schroeder, 2016). More broadly, social media firms sell access to their data (such as Twitter – where outputs have relevant applications in public transport) (Schneider Electric et al, 2014).
- **Private firms that collect data – through sensor networks, tracking software, crowdsourcing, etc. - specifically to monetise it.** Anonymised pedestrian and vehicle location sensor data from Motionloft, a private operator of location sensors, has been used in cities to improve urban planning and traffic management. The service is charged to commercial clients for a monthly fee, as well as shared freely with city authorities in San Francisco via the city's open data platform (Coursimault, 2013).

#### 4.1.1.2 Trends and driving forces

There are two main trends that are driving developments in this area: firstly, the scope of potentially useful data sources has increased in line with improvements in analytical techniques. Secondly, the number of business activities that are digitally monitored and recorded has also grown rapidly (DHL, 2013); (Oxford Internet Institute, 2014).

Developed cities are already seeing Intelligent Transport Systems (ITS)<sup>15</sup> as a cost effective way to improve the management of existing infrastructure, and they are increasingly incorporating such assets as part of system maintenance and renewals (Schneider Electric et al, 2014) (HBR, 2014). The raw data generated by city investments in ITS and other intelligent infrastructure provides a rich source of raw data that can encourage private enterprise to develop services.

#### 4.1.1.3 Gaps and challenges

The private sector is a big generator of data, and their role as suppliers may be an untapped source of significant future value. Yet, many organisations do not necessarily understand the value of releasing it (Bonina, 2013). Since governments and public agencies do not control all of the relevant information, getting private companies to share relevant data is highly important.

Business models in this area are rather underdeveloped due to the general challenges associated with sharing data, that are also discussed in Section 6. Respondents to the Delphi study highlight the need to better understand the value of data and the possible role of developing illustrative case studies as a way to encourage more sharing of private data. In this area, it would be beneficial for more research to provide a deeper exploration of the indirect benefits that are harder to capture (i.e. where companies do not provide the data as a good/service in return for a fee).

#### 4.1.1.4 Lessons learned and opportunities

The availability of reliable data from multiple sources creates a fertile environment for new, smart urban transport solutions to emerge. This is an unpredictable process, as it is not always possible to foresee what solutions could be developed. The focus should rather be on making diverse datasets more widely available, as well as encouraging wide-ranging discussion with stakeholders to uncover needs and build demand for datasets.

It is important to increase awareness among private firms of the scope to exploit data that they routinely collect – through supplying it to third parties, or making use of it themselves as enrichers (see below). There are a multitude of opportunities here, although the process is hampered by concerns over privacy and commercial sensitivities (see Section 6). Despite these challenges, success stories from businesses that share their data are gradually accumulating, and such businesses reportedly enjoy benefits from reduced friction in trade, increased market reach and improved commercial resilience (Deloitte and ODI, 2012).

Several examples of different models used in practice to capture value have been identified in the course of research for this report:

- **Improved customer service:** The survey conducted for this study highlighted the experience of the West Midlands Transport Authority and National Express, where a data-sharing agreement opening up their datasets facilitated an improved customer experience, drove an increase in usage and consequently increased revenues. The availability of real-time data on buses has also been cited as one of the key reasons for a “bus renaissance” in some UK towns and cities and bucking a previous longer-term decline (LowCVP, 2017).
- **Network effects:** collaborating with other organisations in order to reduce the costs of storing and maintaining data. The added value comes from the extended use of data, or the extended reach of data for a broader community – for example, open source mapping software.

For public agencies, there is already clear value from publishing data – allowing insights to be delivered back to the agencies that would otherwise not have been available, or providing services to citizens and

---

<sup>15</sup> ITS encompasses the range of operational technologies used for transport management, including sensors, payment systems, ticketing infrastructure, surveillance, remote controls, display equipment etc. that are employed to monitor and manage travel conditions. ITS also increasingly uses mobile monitoring technologies installed in vehicles or carried by individuals (Schneider Electric et al, 2014)

businesses that would otherwise have been publically funded. Although open data enhances the value of information services by expanding its reach, it is by no means the only way to supply data (as seen in the examples above). Nevertheless, city governments can kick-start the process by releasing their own data on open platforms, as seen in Nottingham, Glasgow and London.

This process need not be expensive, e.g., the original London Datastore cost £15,000 (Schneider Electric et al, 2014). Motivated by the belief that governments need to utilise the power of user-generated content in order to remain relevant, productive and effective, the GLA sought input from developers in an open workshop. The messages received were clear: “go ugly early”. Developers would prefer access to data, even if it is not perfect (Coleman, 2013). The project started scoping in October 2009 and launched on the 5th January 2010 with 50 datasets and a further 150 datasets were published to the website by the end of January 2010, with no additional staff training required and no additional permanent staff recruited for the development and deployment of the project (O'Reilly, 2014).

## 4.1.2 Aggregators and enablers

### 4.1.2.1 Examples and applications

Aggregators collect and publish data from open and proprietary sources. Examples include platforms that cover multiple industries - such as Datahub and data.gov.uk - as well as more specialised transport data sources, such as the Traveline National Dataset, or Highways Agency's National Transport Information Service (TSC, 2015).

Additional value can be unlocked when data from multiple sources is combined, and as such, aggregators often provide analysis and insights on top of data provision services (Deloitte and ODI, 2012); (Oxford Internet Institute, 2014). These added-value services are used to generate revenue, commonly using a “freemium” model, charging subscription fees or pay-per-use pricing. For example, transportAPI re-process most of the national open public transport datasets using automated data batch processes, which is then sold via monthly subscription or a ‘Price Per Hit’ plan to developers like Citymapper and the Greater London Authority (Schneider Electric et al, 2014). Similarly, TravelAI collects and aggregates user travel patterns from accelerometer, GPS and smartphone sensors to infer travel mode from transport service data. This is used to generate APIs for developers to insert into their apps. Internet giants such as Amazon, Google, Facebook and Apple have business models based on the aggregation of data and the provision of cloud services – placing them in prime position for competing in sectors such as autonomous vehicles and Smart Cities (Royal Academy of Engineering, 2015).

### 4.1.2.2 Trends and driving forces

At present, the lack of standards adds significantly to the expense of using Big Data (Oxford Internet Institute, 2014); therefore, software developers are willing to pay for access to formats that allow them to avoid or reduce processing costs. In future, the overhead costs involved in processing inconsistent data sources may be partially mitigated by development of (and adherence to) appropriate standards (Schroeder, 2016). Even in this situation, aggregators are expected to remain important as the spectrum of available data increases and combining data from disparate sources becomes more common (and necessary) for analytics. Under these circumstances it also becomes more efficient and practical for third parties to supply that data rather than duplicating effort (Schroeder, 2016).

### 4.1.2.3 Gaps and challenges

The need for standardisation of formats between separate systems, as well as ensuring high data quality, is a frequently cited challenge (IET, 2013). Another key challenge lies in the development of systems that not only bring together the contributions of both private and public data providers, but also removes the disincentives to data sharing (Schneider Electric et al, 2014).

### 4.1.2.4 Lessons learned and opportunities

Aggregators can add considerable value by assimilating and maintaining datasets in standardised formats, which reduces the frustration for innovators working to create new sources of transport information and insight. They can also help to unlock additional insights through the combination of multiple data sources. For example, the problem of sample bias in mobile phone data (where some

ages or social groups are more or less represented) can inhibit meaningful analysis – but aggregation with other datasets can overcome some of these limitations and create useful inputs (IET, 2013).

In addition, several respondents to the survey conducted for this project cited the presence of trusted third party aggregators as an important element to help overcome potential commercial challenges associated with sharing data. For example, aggregators were considered important in the context of shared ticketing schemes, e.g. where third-parties are trusted to analyse data and distribute revenues to the different commercial parties.

Wider social benefits can also arise as a result of aggregation – for example, reduced congestion made possible by information on road works published by Elgin (Deloitte and ODI, 2012). Elgin spent a decade brokering and crowd-sourcing roadworks data from over 175 systems used by local authorities to create an industry standard dataset. Local authorities pay a low subscription fee to access Elgin's software tools (which cover a range of services, such as helping to plan temporary road closures, and monitoring the impact of roadworks on their networks in real time) and in addition they receive free access to the aggregated national roadworks dataset (TSC, 2015).

### 4.1.3 Developers

#### 4.1.3.1 Examples and applications

Developers typically create apps from dynamic types of data - often derived from/using public open datasets (e.g. timetables and fares), and/or crowd-sourced datasets (e.g. live traffic conditions), which may be derived from data shared by users of the app itself (e.g. Waze and Parkopedia).

The range of services offered by developers can cover multiple areas, such as (TSC, 2015):

- **Making it easier/cheaper to use existing transport systems:** by enhancing use of existing forms of mobility, for example, through smart payments provided by the technology firm ACT. There are also many examples of personal transport planning apps such as: Citymapper; Transit App, OTP, Hopstop. Such services can maximise the use of alternative travel modes in order to reduce congestion and overcrowding.
- **Allowing use of socially shared assets:** respondents to the survey for this study pointed to the growing impact of sharing economy business models e.g. car-sharing apps such as Liftshare (using GPS data for geolocation and feedback systems that help to build trust between users). By-the-minute vehicle hire schemes such as Zipcar and cycle hire schemes such as Velib (analysing usage data to inform where cars/bikes are placed at what times of day, enabling payments and accessing credit card details).
- **Maximising the use of operator-provided assets:** this includes “mobility as a service” models – highlighted by respondents to the survey for this study as an emerging group that is disruptively changing urban mobility. Includes personal mobility services such as Lyft and Uber (holding vast databases of drivers to match with passengers and fares being calculated using GPS and street data, as well as algorithms that monitor traffic conditions in real time), as well as those offering demand-responsive collective transport systems such as Bridi.
- **Optimising commercial transport:** for example, through virtual kerb-space booking for loading and unloading (Activ8 Virtual Parking Solutions); last mile delivery and collection of parcels (e.g. CollectPlus, MYWAYS); online brokerage for intermodal freight (FreightArranger).

A characteristic revenue model for many consumer-facing apps is to provide them free-of-charge in exchange for personal usage/location data. This model has emerged due to several features of the market: in particular consumers have been conditioned to expect such services for free (due to integrated offerings from global companies such as Google and automotive OEMs). In addition, crowdsourced tools need to have a large and varied user base in order to generate valid data, making it difficult to charge users for the service.

Other apps are based on strategies to introduce a product or service for free initially in order to achieve a viable user base, and sustaining its development through advertising revenue or data sales. This model is common in personal information and insight services that are offered to consumers for free, in exchange for data on location or usage. For example, apps like CityMapper have data on user's

intended and completed journey GPS traces, which could open up opportunities for a step change in understanding travel behaviour (TSC, 2015).

The release of public data by governments has helped to generate a thriving ecosystem of app developers for private users. Through their use of certain apps, consumers have demonstrated some willingness to opt-in to (anonymous) data collection in order to receive/improve services, thereby enabling the development of crowdsourced services. A good example of this is Waze, a community-based traffic and navigation app, which relies on its users to share real-time traffic and road information and to report incidents while driving – thereby allowing others to save time and fuel. Such crowdsourced tools rely on having a varied user base – whereas similar services could not be provided directly by individual car manufacturers since they have limited coverage that would affect the performance of the service.

In the area of freight and logistics, there is considerable interest in data exploitation to improve efficiency and customer service, as well as looking to develop new business models based on the provision of proprietary data, which is being driven by the main freight operators (DHL, 2013).

#### 4.1.3.2 Gaps and challenges

To date, the focus for on-demand mobility services such as Uber has been on single modes, and services in this space are mainly focussing on gaining market share. Integration with other mobility services (e.g. multi-modal interchange) is lagging, but presents an interesting growth opportunity. Bringing users into participatory forums that involve multiple stakeholders (e.g. the Open Data User Group) can also help to identify opportunities for value creation (Bonina, 2013).

More generally, the current lack of standards adds significantly to the workload and therefore, expense of using Big Data (Oxford Internet Institute, 2014).

#### 4.1.3.3 Lessons learned and opportunities

Individual developers are already taking advantage of smart technologies and newly available data feeds to offer innovative services in many cities. This is one of the more dynamic areas in the mobility market; however, it tends to be formed of small private actors and fragmented offerings. Cities rarely make conscious and active attempts to maximise the value of the services that private actors are creating, and their role in supporting software innovation is only just starting to be recognised. Yet there could be considerable benefits in terms of advancement of urban sustainability and optimisation of assets, not to mention the economic growth of a sector built up around the technology and data.

Most of the above examples of emerging products and services are strategically building new user-based datasets that have not previously existed at such a granular level. For example, knowing an individual's end-to-end journey patterns (by retaining and examining user's intended and completed journey GPS traces, e.g. in CityMapper) creates the potential for a step change in understanding travel behaviour (TSC, 2015).

Indeed, as tools become more sophisticated, more options are opened up to optimise the networks. For example, the city of Paris is already seeking to use predictive modelling (using variables such as weather, current events, network disruptions and day of the week) in order to influence people's journey planning, and in San Francisco the price of parking bays change in response to demand (Schneider Electric et al, 2014).

- **Augmenting products to generate data:** and using the data to improve services or support new business models. For the former, an example is Rolls Royce, who have long data collected via remote sensors to detect maintenance issues, allowing them to provide a higher quality of service (Oxford Internet Institute, 2014). For the latter, Progressive Insurance offers customers a device that they can plug into their car, which monitors how safely they drive and adjusts the price of insurance accordingly (HBR, 2014). Another example is the use of sensors in physical infrastructure that enable real-time asset condition monitoring, enabling predictive analytical techniques to optimise maintenance cycles and prolong asset lifetimes. For example, a European railway company implemented a new system that monitors the status of its physical and IT assets. This allows them to recognise and resolve 50% of issues before they affect train operations, and the availability of the rail network has increased by 33 hours per month, saving approximately US\$2.3 million per year (IBM, 2014b).

- **Digitalising physical assets:** digitalisation has transformed many manufacturing industries and dramatically reduces distribution costs. For example, digital payments and e-tickets have already been used in some modes and there is a significant opportunity around seamlessly integrated cashless payment systems such as those being developed by Droplet (TSC, 2015). Another example are companies in the e-commerce space, like Amazon and Alibaba, who have circumvented the need for physical stores in which to sell consumer goods.
- **Combining data across industries:** for example, a smart-city initiative in Rio de Janeiro uses data from private utilities, transport companies, and city agencies so that they can deal with natural disasters more effectively (see box below). GLA has set up the Agile Urban Logistics project, which combines data from retailers and on traffic conditions in order to optimise delivery services and manage congestion in the city (HBR, 2014).
- **Codifying a service capability:** allows a company to take any process in which it is best-in-class and sell it to other companies, using cloud computing. Internet powerhouses (such as Google, Microsoft, Apple etc.) have successfully established information-driven business models based on the provision of standardised service and in the management of ever-increasing amounts of data. Network orchestrators, like Uber and Airbnb, specialise in providing platforms that connect supply and demand, without having to own significant assets of their own. Uber is replicating its expertise in providing the platform in other areas, such as Uber for freight (competing with traditional freight exchanges), and Uber Eats (restaurant delivery).

#### Box 4.2: International Examples of Smart Cities: Rio de Janeiro

##### Smart City Rio de Janeiro

The example of Rio de Janeiro demonstrates how smart cities can combine data across industries to improve services for its citizens. Rio built its integrated control centre in 2010, which allows city agencies to monitor events across the city in real time. Key examples of the benefits include:

- As a partner in the Waze Connected Cities program, Rio is reportedly using the navigation app's crowd-sourced traffic data to help monitor road conditions and improve garbage routes.
- Real-time monitoring has allowed for a 30% reduction in emergency response times.

Cities often need dynamic solutions that are tailored to the specific context, and this in turn relies on effective dialogue to communicate their needs. Rio provides good insight into the difficulty of applying “out of the box” products. Even though the smart city initiative was widely lauded as a success – winning the World Smart City Award in 2013 - Rio's working group reportedly found that IBM's standard software was unsuitable for their Operations Centre. Moreover, collaboration to solve the issues was difficult, highlighting that a lack of dialogue between city authorities and technology developers is an important challenge to implementing new systems.

Sources: (Jaffe, 2016); (Berst, 2013).

#### 4.1.3.4 Trends and driving forces

Due to the explosion in digital data, cloud computing, creative companies are able to find ways to expand their existing services or forge into new areas (HBR, 2014). Developments in sensor technologies and wireless communication make it more feasible to augment products to generate data.

#### 4.1.3.5 Gaps and challenges

Many companies are in the early stages of understanding how to benefit from Big Data, and furthermore do not currently have the funding or expertise needed to extract the value. Even large organisations may not possess all of the required skills to manage Big Data, often needing to collaborate with other firms that can offer specific expertise (Royal Academy of Engineering, 2015).

#### 4.1.3.6 Lessons learned and opportunities

Most companies currently have access to information that could be used to expand existing businesses or build new ones, and opportunities continue to grow as technology advances (HBR, 2014). The

difficulty often lies in understanding and extracting the value, in which case it can be useful to gather perspectives from stakeholders external to the firm. Cities and governments could play a role in facilitating this search process in collating examples, building networks and implementing digital data platforms, such as that introduced by Copenhagen (see earlier Box 3.4).

## 4.2 The role of government

According to Schroeder (2016), a common theme among industry stakeholders is that government should provide the minimum regulatory infrastructure to facilitate business models. This would allow businesses to innovate and compete on their own merits, rather than picking winners (Oxford Internet Institute, 2014). That is not to say there is no role for government; indeed, there is great demand for government support to overcome key challenges – lack of understanding of the value of data, standardisation, and addressing concerns over privacy, security and the use of personal data. These challenges are general to the use of Big Data, but overcoming them will help to smooth the way for new business models.

The public availability of data provides greater certainty and allows new products and services to grow up around it. Respondents to the Delphi survey emphasised the importance of open data and use of common standards as facilitators for development of business models – including disruptive new services that have not been identified yet. Many city authorities have already taken the first important steps by making their data open – the next step is to raise awareness of the availability of data and the benefits associated with its use, and guarantee its future availability. Those agencies that still need to make their data open should prioritise releasing their data in some form, even if this is very simplistic – as noted above, developers would prefer access to data, even if it is not perfect (Coleman, 2013). For public bodies wishing to do more, respondents to the Delphi survey suggested that funding could be beneficial, either in terms of R&D/scale up investment in return for adoption by the industry of open platforms & standard, or to focus on research towards commercial applications (rather than commercial pilots, as is typically the focus).

To ensure continued progress, city authorities and governments should make a concerted effort to engage in discussions with citizens and industry in order to better understand and source missing data that could be used to overcome challenges.

Collaboration should be encouraged between diverse local actors, including government, transport operators, infrastructure providers, telecoms, academia, etc., in order to explore the opportunities that Big Data can bring. This process can also be facilitated by compiling market monitoring and intelligence reports that can help to send commercial signals to the UK industry (TSC, 2015). New data-sharing partnership models may become necessary that go beyond today's procurement models or client-provider relationships. Research establishments that are capable of maintaining a long term memory of the databases, metadata and associated analytics will also have an important role to play to ensure continuity and innovation.

The insights generated through these activities will in turn help to inform decisions over strategic investments in sensors and operational technologies, or input into the design of fiscal incentives to ease the entry of new players. It is important to recognise that there is no universal approach for city governments investing in smart technology – hence, the input from stakeholders who are informed about the local context is key to ensure that the right decisions are made (Cosgrave et al, 2014). However, one opportunity to enhance the stores of open data available is to require that external service providers and infrastructure operators should supply the output data, so that this can be used to generate greater benefits.

A lack of awareness among decision-makers can inhibit the development of new business models, indicating a potential role for awareness-raising activities. In particular, target groups could include small businesses that may not be aware of the opportunities and resources available – or alternatively, targeting citizens to promote user acceptance and trust. There may also be demand for capacity-building and support in terms of the skillsets needed to access and use large datasets (Oxford Internet Institute, 2014).

### 4.3 Research gaps and potential solutions for the business models for transport Big Data applications

The identified key gaps and possible actions/solutions are summarised in Table 4.2 below.

**Table 4.2: Key understanding/research gaps and recommended actions on the business models for transport Big Data applications**

# Key research gaps and potential solutions	
	<b>Gaps in research / understanding</b>
1.	While there is some general information available on the roles that different organisations can play within data-driven business models, there is a lack of in-depth understanding and detail as to the opportunities of new business models within the transport sector specifically. More research is needed to better understand the structure and value of business models, as well as what actions are needed to foster their development.
	<b>Potential solutions</b>
2.	Governments and city authorities should continue to encourage innovation in new business models in the private sector by making their data open, further raising awareness of the data and guaranteeing its future availability. As also indicated earlier, it is also important to ensure consistency through common standards and architectures facilitate the collation/aggregation of datasets from different cities and regions to provide sufficient critical mass for developers/users.
3.	Few cities take an active role to maximise the value of the services that private actors are creating. To do this, collaboration and consultation with a broad range of actors – including industry and academia – would help to identify unmet needs and better understand the opportunities that Big Data could bring. The findings could also be used to inform strategic investment in sensors and operational technologies needed to fill identified data gaps
4.	The challenges in developing new business models are the same as the general challenges associated with the use of Big Data discussed earlier – principally around commercial sensitivities, lack of understanding of the value of data, privacy and security concerns, and lack of skills. Addressing these challenges will smooth the way for innovation.

## 5 The skills and capability needed for Big Data services

### Box 5.1: Summary of key findings on skills and capability

- The skills needed for Big Data services are general and not specific to transport; however, transport-specific knowledge is equally important to be able to understand the issues that Big Data is employed to tackle and know how to interpret the results.
- Transport is likely to have difficulty attracting and keeping individuals with the relevant skills for Big Data and IoT compared to some other sectors due to less competitive salaries or a perceived lack of innovative and stimulating workplaces.
- A range of key actions have been identified to help bridge the skills gap. Particularly important government/public sector-led actions identified include:
  - Earlier introduction of programming and data analytics in the school curriculum, plus general improvement to uptake, and earlier introduction, of STEM subjects.
  - Develop options to help attract skills from outside of the UK and/or keep them within the UK.
  - Industry to help develop Big Data and data analytics skills by up-skilling existing workforces.
  - Development of new courses (or improvement of existing courses) at university level, including multi-disciplinary training.
  - Improvement of the use/uptake or design of existing schemes such as the Apprenticeship Levy to make them work for Big Data analytics.

### 5.1 Identification of key skills and capabilities

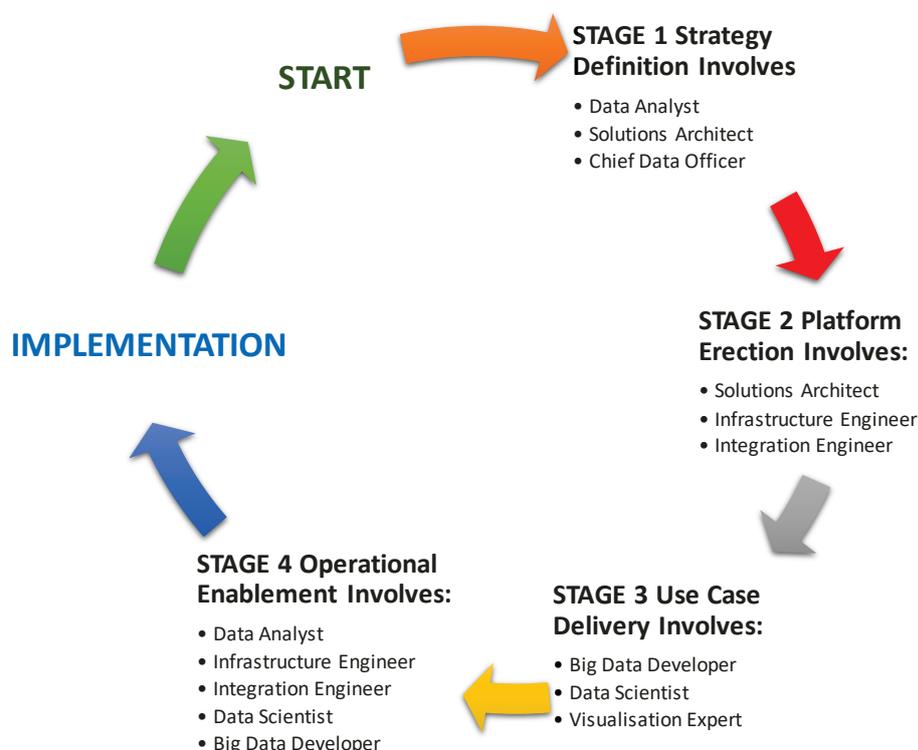
There is growing concern over a lack of appropriately skilled people as a major barrier to exploiting opportunities in Big Data – both in general (Development Economics, 2013), and in the transport sector (TSC, 2015). Indeed, a lack of Big Data skills is already proving to be a constraint that costs the UK an estimated £2 billion per year (Development Economics, 2013).

There are many skills that could be needed in a Big Data project to varying degrees, as illustrated in Figure 5.1 below, and these are not necessarily specific to the transport sector. Specific roles identified as being in very high demand in the UK, with large numbers of vacancies are (TechUK, 2016):

- **Data analysts:** required to complete a range of highly analytical tasks such as identifying trends, investigating correlations, understanding drivers of customer experience, creating management account and information, study key metrics etc.
- **Data infrastructure engineers:** needed to build reliable, scalable, usable big data platforms and install and manage tools such as Hadoop, Storm, and Spark.
- **Solutions architects:** who work with the whole big data team to select the right technologies and design for new systems, and have strong knowledge across a range of technologies and practices.
- **Data scientists:** that have programming, analytical, statistical, mathematical, predictive modelling skills as well as business strategy skills to build algorithms which answer business needs.

As well as the need for skillsets related to Big Data, sector-specific knowledge is also important for individuals to be able to understand the issues at hand, as well as how to interpret the analysis. It is not clear whether the need for transport-specific knowledge will affect the relative skills gaps for the different capabilities required. Currently, expertise in data analytics predominantly lies within service providers or consulting companies. However, some of these companies tend to have analysts focus on certain industry sectors.

Figure 5.1: Roles needed at different stages to implement a Big Data project or strategy



Source: Reproduced from (TechUK, 2016).

Moreover, senior decision makers will need to have sufficient knowledge and understanding to make strategic decisions based on intelligence derived from the analysis of multiple complex connected datasets. Stakeholders consulted for this study felt that much of the transport industry's organisational culture and leadership is highly conservative and does not have a good awareness or understanding of the potential Big Data's potential.

Addressing these issues will be highly important for other transport-sector organisations for a number of reasons, for example: (i) to help open up existing datasets vital for the delivery of intelligent mobility (ii) to ensure the value of data is understood and appropriate commercial models are applied, and (iii) also to ensure that organisations see (and realise) the value of the huge quantities of data that are generated, but currently not stored due to the organisation capturing the data not understanding its potential value.

## 5.2 Assessment of the current situation and future potential

To understand issues in a transport-specific context, we asked stakeholders about gaps in the following broad skill areas:

- Dataset handling and manipulation** – collection, mining, cleaning and aggregating data.
- Computational analysis** - algorithm design, statistical analysis, modelling and visualisation.
- Software development/programming** – design of ICT technologies and services to meet customer needs.
- Sensor technology design and deployment.**
- Management/leadership with effective understanding and vision.**

The views from the experts consulted were generally quite mixed on the relative importance of each of the skills gaps in the UK. Several of the stakeholders interviewed indicated that there were some particular skills gaps in dataset handling, visualisation, or transport-specific knowledge, e.g. transport "data scientists".

The gaps are thought to be relevant to all sectors, not just transport; however, many of the consulted experts agreed that the transport sector struggles to retain professionals, due to higher salaries in other sectors, and needs to avoid a perception of a lack of innovative and stimulating workplaces. It was also noted by a number of stakeholders consulted in this project that the transport industry has a particularly conservative and siloed culture in both the private and public sector, which contributes to a lack of sufficient resources to utilise the available skills. To combat this, there is a need for both awareness raising, as well as education at a senior level. These findings are also consistent with those from other studies in the literature.

The DfT's Smart City workshops have indicated that in-house skills in local authorities have declined, and budget cuts have resulted in a reduction of intelligent client roles within the authorities. This has led to concern over the potentially negative consequences for procurers of increasingly out-sourced service delivery in this area. Conversely, (Schneider Electric et al, 2014) has identified how strategic skill development can empower local and national governments. The study also indicated that internationally, *"Leading cities in this field are already appointing Chief Information Officers and Chief Technology Officers to own and coordinate the city's response across all sectors"*.

There is a lack of formal university-taught data science courses, and data scientists are currently mainly self-taught. A recent report by (Royal Academy of Engineering, 2015) suggested that the creation of the Alan Turing Institute will help the UK strengthen its workforce's skills in data science and thereby become more competitive. However, the introduction of additional specific academic and workplace training would be valuable to help bridge the skills gap.

### 5.3 Potential solutions to address skills and capability gaps

A number of potential solutions /possible actions have been outlined in the literature review and during the consultation with experts for this project (see Table 5.1). Broadly speaking, all actions are considered to be of similar importance, indicating a need for coordinated action on multiple fronts, including from the earliest stages of education. Ensuring collaboration between industry and government is considered highly relevant so that courses are relevant and individuals are better able to understand the data and their skills align with the industry needs. As indicated earlier, the majority of these are general actions/solutions, rather than specific to the transport sector.

**Table 5.1: Key gaps and possible actions/solutions identified in the literature and consultation for this study with regards to ensure the skills needed are available in the UK in the future**

#	Key gaps and potential solutions/actions to fill skills and capability gaps	Source
<b>Schools and colleges</b>		
1.	Earlier introduction of programming and analytics in the school curriculum.	(2), (5)
2.	General improvements in STEM (Science, Technology, Engineering and Maths) subjects, including application-oriented teaching of mathematics and statistics	(1), (5)
3.	Provide more and better information about analytical career prospects and role models	(1), (4)
4.	Embed data analytics in other subjects	(1), (5)
5.	Support development of extracurricular data activities	(1)
<b>Universities and Vocational Education</b>		
6.	Increase the visibility of strong data analytics courses	(1),
7.	Embed quantitative analysis across disciplines	(1), (4)
8.	Boost the business and soft skills of graduates from data analytics courses	(1)
9.	Foster interdisciplinary research and skills development programmes	(1), (5)
10.	Development of courses on Data Science, which should also include (e.g. as a module) its application to transport systems and personal mobility challenges.	(3), (2), (5)

#	Key gaps and potential solutions/actions to fill skills and capability gaps	Source
11.	Improved input of businesses/employers into digital skills educational courses and the provision work-based learning experiences	(4)
12.	Accelerate the roll-out of digital apprenticeships in non-ICT companies	(4)
<b>Labour Market and Industry</b>		
13.	Create a cross-cutting taskforce around data analytics	(1)
14.	Ensure top management fully understands and supports the use of Big Data to inform decision-making is also highlighted by improving awareness and education by: <ul style="list-style-type: none"> <li>Actively convening industry and community analytics networks.</li> <li>Raising awareness of the value of data for business and local public services.</li> </ul>	(1), (5)
15.	Support innovative interventions enabling local authorities to boost local analytical skills (e.g. increased support to encourage greater numbers of digital start-up enterprises at a local level, mentoring from established companies, local finance initiatives, and stronger graduate-SME digital knowledge transfer partnerships).	(1), (4)
16.	Develop innovative solutions for data analytics training, for example: <ul style="list-style-type: none"> <li>A model of reverse mentoring in industry, whereby young skilled IT professionals teach existing staff new skills.</li> <li>Development of a new standard aimed at re-training people in highly technical data skills, where re-training requires skills that fall out of the scope of existing apprenticeship standards.</li> </ul>	(1), (2), (5)
<b>Government and Local Authorities</b>		
17.	Develop specific options to help attract skills from outside of the UK and/or keep them within the UK. For example, Data Analysts, Data Infrastructure Engineers and Solution Architects should feature prominently on any future Government preferred shortage occupation list, joining Data Scientists, to ensure the best international talent can also be recruited to supplement that developed by/already in the UK.	(2), (3), (5)
18.	The Department for Education could work to ensure there is more uptake of higher-level apprenticeships from a younger age in the data analytics area with the forthcoming Apprenticeship Levy.	(2), (4)
19.	Promote the value & importance of Big Data and Data Analytics in the education sector, the public sector and industry, in particular to ensure key decision-makers understand the potential benefits of Big Data and open data.	(2), (5)
20.	Local authorities need to ensure they gather the appropriate skills and knowledge to be a 'smart customer' in procuring support to implement Big Data projects. This should include building on the work of existing networks, such as the Transport Data Initiative (TDI) with its network of local authorities and industry specialists aiming to improve the way their transport data is utilised and in turn improve the cost of delivery. Ideally local authorities should consider appointing Chief Information Officers and Chief Technology Officers to own and coordinate the response across all sectors.	(5)
21.	Review and support the development of appropriate computing power, data storage and connectivity infrastructure is sufficient to meet anticipated needs to deliver big-data /IoT driven developments. In particular, provide resources where necessary to ensure communities with limited access to the new technologies that make Big Data possible are better supported, to reduce the risk of discrimination in this area.	(5)

Sources: (1) (Nesta, 2015a), (2) (TechUK, 2016), (3) (TSC, 2015), (4) (Development Economics, 2013) and (5) consultation with experts and analysis of other literature by Ricardo for this study.

## 6 Challenges and enablers to applying Big Data in the transport sector

### Box 6.1: Summary of key findings on challenges and enablers

#### Challenges

- The identified challenges tend to be general across different sectors, rather than specific to transport. The main challenges are thought to be:
  - **Technical:** lack of appropriate skills and a lack of common standards and challenges in connecting data silos.
  - **Commercial:** lack of understanding of the value of data, a lack of clear business models, commercial sensitivities that prevent sharing of data and the need for upfront investment to develop and store data.
  - **Institutional:** uncertainties over data privacy, security and a need for strong leadership.
- The findings from the Delphi Survey suggest that experts think that the commercial sensitivities associated with data and the lack of clear business models are the most important obstacles, whereas technical challenges were thought to be less relevant and more easily overcome.

#### Enablers

- **There is a need for a strong vision on the potential of Big Data, and leadership to define strategic data-driven actions.** These are considered to be essential to overcome cultural, legal and organisational challenges, as well as uncertainty over digital rights and data protection laws.
- It is thought that technical challenges and concerns over privacy will naturally decline as the Big Data and IoT sectors develop and data sharing becomes more common; however, closing the skills gap requires coordinated actions at multiple levels of education (i.e. primary, secondary, university, vocational, etc.). The emergence of new analytical techniques that can be deployed to overcome issues of fragmentation, data silos and (some) concerns over privacy, but these need to be supported by the provision of new datasets and rigorous use of data standards.
- Commercial and institutional challenges are considered more difficult to address. Stakeholders also emphasised that minimal, reliable and consistent rules are needed to provide the right environment for innovation. Strong leadership and governance is also needed here. This should be accompanied by more research into understanding the value of data assets and issues of commercial sensitivities around sharing data, learning also from experience from other sectors, such as banking and retail.

### 6.1 Challenges

The challenges associated with using Big Data fall into three groups: technical, commercial and legal/institutional. The analysis in Table 6.1 was informed by both the Delphi survey undertaken for this project, and the literature review and consolidates the information identified in the earlier chapters of this report. Appendix 0 provides a full summary of stakeholder consultation results in this area.

Table 6.1: Challenges to applying Big Data in the transport sector

Challenges identified	
<b>Technical</b>	<ul style="list-style-type: none"> <li>• The <b>lack of in-house skills needed to deal with Big Data</b>. For example, skills to create data feeds by using common data formats (TSC, 2015); (Rusitschka &amp; Curry, 2016).</li> <li>• <b>Lack of common standards</b>, which are key elements to allow for coherent collection, linking, sharing and analysis of different datasets (Schroeder, 2016). Analysts can spend as much as 75-90% of their time cleaning data, which is often provided in non-machine readable or non-standardised formats requiring manual re-entry (Bulger et al., 2014); (Schroeder, 2016). There also seems to be significant uncertainty and lack of awareness among experts about the state of the art and specific future needs regarding data standards.</li> <li>• <b>Failure to connect data silos</b> (both within an organisation and between different organisations), which is one of the main issue preventing sharing and opening up of data (Policy Exchange, 2016). This is often an issue when ownership of data by separate organisations means that data are stored in a variety of formats and data standards (linked to the previous barrier). This problem ultimately prevents successful sharing and opening-up of data (Rusitschka &amp; Curry, 2016).</li> </ul>
<b>Commercial</b>	<ul style="list-style-type: none"> <li>• <b>Lack of understanding of the value of data</b>. The value of data is still an intangible asset in many companies, and the lack of recognition of this value hinders the opportunity to assess and measure the contribution of data. As a result, decision-making on how best to develop, trade, protect and exploit such assets is potentially compromised (Royal Academy of Engineering, 2015). Ultimately, this means some organisations are reluctant to make data available that may be commercially valuable at a later stage (TSC, 2015).</li> <li>• <b>Commercial sensitivities and vested interests</b> in the data that organisations own, which is often considered key to maintaining competitive advantage. This, together with vested commercial interest are considered the two biggest impediments to an open-by-default and viable business model. Fear of losing competitive advantage is considered by the Delphi participants as the main reason for not sharing data in the case of companies that use data specifically to operate their business.</li> <li>• <b>The need for upfront investment and a lack of financial resources</b> to maintain and curate reliable data feeds (e.g. on an open, or commoditised basis) due to uncertainty over the business model/value created for the data owner (TSC, 2015); (Rusitschka &amp; Curry, 2016). The cost of opening up data is a particularly relevant barrier for cities and local governments (Policy Exchange, 2016). This is also an important factor that companies that generate data as a by-product of their main activities tend not to share data.</li> </ul>

Challenges identified	
<b>Institutional/Legal</b>	<ul style="list-style-type: none"> <li>• <b>Concerns/uncertainties over data privacy and related data protection laws</b> are seen as the main challenges that limit both the collection and the sharing of data (TSC, 2015); (Rusitschka &amp; Curry, 2016). Respondents to the Delphi survey indicated that this was the most important issue. Privacy concerns also create <b>ethical and legal uncertainties</b> – for example through the availability of location data (Royal Academy of Engineering, 2015).</li> <li>• Most participants in the Delphi survey did not consider regulatory challenges (e.g. banking, telecoms, compliance with anti-competitive activities) to be an important impediment to data sharing (i.e. only a small number of respondents consider it to be an important reason for not sharing data). However, in some cases, regulatory challenges associated with sharing certain datasets were recognised. For example, one respondent indicated that because the sharing of data was not mandated or regulated properly, this also provided a strong disincentive for sharing data. Other examples include financial penalties linked to key performance indicators in contracts for local bus services and for rail franchises.</li> <li>• These issues are found to be linked to the <b>institutional and organisational cultures of organisations</b> – for example, the extent to which people/organisations are protective of ‘their’ data. This often leads to a reluctance to share data within an organisation as much as to an unwillingness to share data between organisations (which links back to the technical and commercial challenges already described), which influences the need for an organisation-wide policy and strategy for data use (Bulger et al., 2014).</li> </ul>

Different approaches are required to overcome the different types of challenges. In general, technical challenges are considered to be easier to overcome organically as the IoT develops, and service provider’s demands for existing public datasets grows. Conversely, challenges related to market forces are considered more difficult to address (TSC, 2015). Stakeholders commented that concerns over privacy, confidentiality and security including legal implications will decline over time as data sharing becomes more common and business models provide a strong reward to share data.

At the same time, there also seems to be agreement that in order to address economic challenges, policy-oriented actions ought to be taken. Approaches to addressing these challenges, along with consideration of other enabling conditions, are discussed more detail in the following section.

## 6.2 Enablers

The identified enablers relate to: technology, education and the political/organisational environment. Table 6.1 provides an overview of the enablers identified, and consolidating also the information identified in the earlier chapters of this report. A full summary of the findings in this area from the stakeholder consultation results is also provided in Appendix 0.

Table 6.2: Enablers for Big Data and IoT applications in the transport sector

Enablers identified	
<b>Technical</b>	<ul style="list-style-type: none"> <li>• <b>Further development of ICT and broadband communication</b> technologies (e.g. 5G technology allowing for improved cloud computing and Big Data analytics). Such developments would improve data storage capacity and ability to conduct (real-time) data collection and analysis (Rusitschka &amp; Curry, 2016); (TSC, 2015), (SMMT, 2017).</li> <li>• <b>Improving the standardisation of open and shared data</b> is considered to be a general technical enabler that can improve the extent to which Big Data are collected, shared, linked and analysed (regardless of sector) (Royal Academy of Engineering, 2015); (Rusitschka &amp; Curry, 2016) (OneM2M, 2017). Stand-alone cities, with relatively large budgets (e.g. London/TfL), have successfully opened their data and demonstrated the value of transport apps, but the adoption of common global standards is needed support national interoperability and export potential. Standards are needed to maximise Big data and IoT technology applications. In the smart city context, they help in <b>making IoT services interoperable</b>, which enables the use of networks, devices, and gathered data for multiple purposes, allowing for greater cross-departmental synergies, and, on the other hand, <b>addressing security issues</b> rising from IoT technology implementation.</li> <li>• Stakeholders noted that standards are needed to improve the collection and anonymisation of data that have high spatial and temporal resolution, and to improve metadata (e.g. the type and format of the information) in order to decrease the high level of fragmentation/data siloes, specifically in the transport industry. The latter requires further technical standards. As a result of improved standardisation, it will be possible to implement <b>emerging techniques such as differential privacy</b>, which is a new statistical analysis method that improves privacy in real-time data collection practices (Rusitschka &amp; Curry, 2016). The method is used to add a particular pattern of statistical perturbation to a dataset so that it is possible to identify the source of inappropriate use of data, if a leak occurs. In turn, such methods are seen to motivate enhanced data protection practices by the user. Such techniques open up a new and promising frontier, although much R&amp;D will still be needed to formulate rigorous data use standards and legislation.</li> <li>• In addition, stakeholders recognise the need for innovative datasets, new technologies and methodologies for data analytics (e.g. new tools for data visualisation techniques).</li> <li>• All of the technical enabling factors above are considered relevant because if implemented they would improve the <b>connectivity of data</b>. The latter is considered crucial in a data-driven economy, where data need to be transferred between multiple different organisations (Royal Academy of Engineering, 2015). Stakeholders were in agreement, recognising the need for improved standards and protocols to facilitate higher interoperability of datasets. <b>Data discoverability</b> requires an open market space where data and information can be found. This is necessary for Smart Cities as well as other industry areas, due to the need to have access to different datasets beyond just transport – e.g. environmental, health, etc.</li> </ul>
<b>Education</b>	<ul style="list-style-type: none"> <li>• <b>Address the skills gap by adapting school curricula</b> and diversify educational choices to reflect new capabilities. This is important so that the results of Big Data analytics are accessible not only to data experts but also to business users (Rusitschka &amp; Curry, 2016). Enabling factors concerning the skills gaps are further analysed in section 5 of this report.</li> <li>• We recommend setting up partnerships with local universities in order <b>to improve public sector data skills</b> (Policy Exchange, 2016). Stakeholders also recognise the need for providing specialised training.</li> </ul>

Enablers identified	
<b>Political/ organisational environment</b>	<ul style="list-style-type: none"> <li>• <b>There is a need for a strong vision on the potential of Big Data, and leadership to define strategic data-driven actions.</b> These are considered to be essential to overcome cultural, legal and organisational challenges (Policy Exchange, 2016), as well as uncertainty over digital rights and data protection laws. Stakeholders called for clearer leadership and governance regarding the different stages of Big Data use, and across different sectors and organisations. This could be achieved through improved policy and the development of strategic roadmaps from national and local government to implement smart cities visions.</li> <li>• <b>Provision of reliable, minimal, and consistent rules and regulations.</b> Once the issues of privacy and data protection are addressed, energy and mobility start-ups require freedom for exploration and experimentation with data if innovative services are to emerge (Rusitschka &amp; Curry, 2016). As highlighted in the literature – data-driven service providers across Europe (especially start-up companies looking to scale up their business models) are struggling to grow in Europe due to different data-related regulations. By contrast the US market is thought to be more appealing and attractive due to its more consistent regulations regarding digital rights (Rusitschka &amp; Curry, 2016); To tackle this issue in the context of CAV technology development, the EU has recently taken action in defining a joint strategy set out in the Amsterdam Declaration of April 2016, which emphasises the importance of coherent international, European and national regulatory frameworks (SMMT, 2017).</li> <li>• Lessons should also be drawn from other sectors (e.g. banking) to address privacy and security issues. An initiative stemming from a government and industry collaboration - The Open Banking Working Group – addresses regulatory, legal and safety concerns regarding the sharing of data by defining Open Banking Standards.</li> <li>• Finally, <b>developing a trusting, collaborative approach</b> was identified as particularly important in the stakeholder consultation, and in particular recognising the end shared benefit of “growing the pie” rather than “growing the size of the slice”.</li> </ul>

### 6.2.1 Actions for industry

In the Smart City context, relevant industries, both established and start-ups active in providing data-driven services, have a role in taking action. Below, we list some type of actions in relation to specific challenges and enablers identified above.

- **Address the lack of understanding of the value of data assets as well as commercial sensitivity issues**

The lack of awareness about the value of opening up and sharing data is a key challenge that tends to precede all the challenges in sharing data (privacy, platforms, commercial value, etc.) (Open Data Institute, 2016). A recommended action for industry (e.g. individual organisations, both established and start-ups) is to document and share case studies or success stories that arise from sharing data, in order to promote awareness. In Section 3, the economic feasibility of opening up data and creating or feeding into open data architecture, especially from private companies' perspectives, is analysed in detail.

- **Contribute to the development and improvement of open data standardisation**

Our findings from the literature review and stakeholder discussions indicated that industries should play a key role in collaborating with government to develop open standards. This should be a market-led process, in order to ensure that the requirements of industry are considered and included. Moreover, the rate of evolution of standards should depend on market feedback, and should be updated as necessary to meet user requirements (Royal Academy of Engineering, 2015). A good practice example is the initiative led by the Information Technology Standards Committee (ITSC) in Singapore, which set

up an industry-led working group to identify relevant open standards in line with the Smart Nation plan<sup>16</sup> (SPRING Singapore, 2016).

Moreover, some stakeholders indicated that there was uncertainty over whether new standards are needed, and suggested that current standards across different industry sectors should be analysed in order to clearly identify gaps, if any. If new standards are proposed, these should be linked to existing ones to ensure coherence and rationalisation of approaches.

## 6.2.2 Actions for Government

- **Take further actions towards adopting Open Government policies and initiatives**

The literature on Open Government Data (OGD) initiatives and policy framework (Ubaldi, 2013), (Judie Attard, 2015) identify possible solutions to the challenges and issues that hinder open government initiatives from reaching their full potential, summarised in Figure 6.1.

In the UK, in the last three years there has been a shift from the transparency agenda (i.e. open data) towards the development and implementation of the shared data agenda. The latter represents a step further in terms of facilitating Open Government Data (OGD) initiatives as it enables to use and benefit from IoT technology further, especially in the smart city context.

To enable such shift in the data-based services ecosystems (both public and private), different legal requirements and regulations are needed. To serve this purpose, the Digital Economy Bill (UK Parliament, 2017) proposed a range of measures to support the development of the digital economy in the UK, including improved connectivity and better digital infrastructure, provision of public services using digital technologies and ensuring protection of citizens' activities online.

In particular, Part 5 – Digital Government concerns data sharing between public bodies; it focuses on improving the provision of public services through better use of data, assisting in fraud detection and facilitating researchers and statisticians' access to public data.

However, the bill has generated some controversy and concerns regarding the terms for sharing data (i.e. particularly in the area of data protection and on which data would be shared and how).

Following agreement by both Houses on the text of the Bill, the Digital Economy Act received Royal Assent on 27 April 2017 (UK Parliament, 2017a).

---

<sup>16</sup> The ITSC is supported by SPRING Singapore, which is an agency under the Ministry of Trade and Industry responsible for helping Singapore enterprises grow and building trust in Singapore products and services, and Singapore IDA, which is the Infocomm Development Authority, under the Ministry of Communication and Information.

Figure 6.1: Overview of possible solutions in open government data initiatives

Challenges		Possible Solutions
<b>Technical</b>	<ul style="list-style-type: none"> <li>• <b>Formats</b></li> <li>• <b>Ambiguity</b></li> <li>• <b>Discoverability</b></li> <li>• <b>Representation</b></li> <li>• <b>Capacity</b></li> </ul>	<ul style="list-style-type: none"> <li>• Use a machine-processable, non-proprietary format;</li> <li>• Use descriptive format; add documentation Metadata;</li> <li>• Use good quality metadata; adopt advanced tools on portals;</li> <li>• Define and use standardised representation;</li> <li>• Apply standards; run large-scale training;</li> </ul>
<b>Policy /Legal</b>	<ul style="list-style-type: none"> <li>• <b>Copyright/Licencing</b></li> <li>• <b>Conflicting Regulations</b></li> <li>• <b>Privacy/Data Protection</b></li> <li>• <b>Liability</b></li> </ul>	<ul style="list-style-type: none"> <li>• Define standard data policies;</li> <li>• Define open government data initiative policies and legal frameworks;</li> <li>• Define privacy regulations;</li> <li>• Implement access control mechanisms (to limit openness of the data);</li> <li>• Raise awareness through stakeholder engagement;</li> </ul>
<b>Economic</b>	<ul style="list-style-type: none"> <li>• <b>Budget Provision</b></li> </ul>	<ul style="list-style-type: none"> <li>• Provide budget targeted at promoting open data initiatives</li> </ul>
<b>Organisa-tional</b>	<ul style="list-style-type: none"> <li>• <b>Institutionalisation</b></li> <li>• <b>Overlapping Scope</b></li> <li>• <b>Technical Support</b></li> </ul>	<ul style="list-style-type: none"> <li>• Define open government initiative policies; re-organise unfit organisational structure;</li> <li>• Use provenance metadata;</li> <li>• Provide support to public entities with the executing of an open data initiative; Use provenance metadata;</li> </ul>
<b>Cultural</b>	<ul style="list-style-type: none"> <li>• <b>Motivation</b></li> <li>• <b>Awareness</b></li> <li>• <b>Public Participation</b></li> <li>• <b>Competition</b></li> </ul>	<ul style="list-style-type: none"> <li>• Raise awareness on the reuse of open data and its benefits;</li> <li>• Highlight the value and potential of open data;</li> <li>• Raise awareness providing incentives;</li> <li>• Provide specific data at a nominal fee (this limit the openness of the data);</li> </ul>

Source: adapted from (Judie Attard, 2015)

- **Create the conditions for fostering innovation and business development**

The evolution of the Big Data landscape has made it necessary to leverage all the benefits of Big Data through public-private partnership. An increasing amount of the data and analyses (e.g. on travel conditions, road safety, traffic management and travel behaviour) are now held by the private sector. Yet central and local governments and their agencies are still, and will continue to be the provider of essential services as well as the regulatory bodies.

The case study on Singapore's Smart Nation programme', and Barcelona's 'Digital City' programme (see Box 6.2 and Box 6.3 respectively) represent good examples of government-led open data initiatives. The latter requires city or national governments to facilitate and attract businesses and citizens to participate, collaborate and re-use open government data.

#### Box 6.2: International Examples of Smart Cities: Singapore

##### Singapore Smart City

Singapore is considered a prime example of a Smart City. The [Smart Nation programme](#) was introduced in Singapore in 2014 with the aim of fostering collaboration between citizens, public and private sectors to develop innovative solutions that improve living standards and create better opportunities. The following examples show how the city uses big data and the IoT in Smart City applications:



- **Smart Transport:** transport users only require one Radio Frequency Identification (RFID) card to use both the metro and the bus, thereby enabling user numbers to be monitored in real-time across the whole network. The data collected is then used by the Future City Laboratory to assess the impacts of updating the network (such as adding new bus routes) in terms of congestion, traffic patterns and customer behaviour. A partnership between the Land Transport Authority of Singapore, Singapore's Public Transport Operator, IBM and Starhub (a major mobile operator) is using the data collected from the cards and combining it with footage from security cameras and anonymised mobile phone data. Real-time taxi data is also collected and available through the Government of Singapore's open data portal. These datasets are used to assess traffic patterns and apply predictive analytics to the network.
- **Smart Healthcare:** to address the challenge of its ageing population, Singapore has put in place the Smart Elderly Monitoring and Alert System. This plan is voluntary and entails the use of home sensors which alert the caregiver of any unusual movements that might indicate a situation of distress.

The collection of vast amounts of data may have privacy implications. However, to Singapore's citizens, this does not seem to be a significant issue, since they are willing to trade their privacy for the benefits associated with using their data. The collection of data without the user's permission is widely accepted by Singaporeans. Laws in Singapore allow the government to gather data without the court's permission, and the increased surveillance is accepted by most Singaporeans in exchange for higher security. This must be considered in the context of a country where strict controls on personal behaviour are already the norm. Regarding the Smart Nation Programme, the government has said it is considering the privacy issue and has vouched to ensure privacy by anonymising citizens' data. The programme seems to be popular among Singaporeans despite the privacy issues.

*Sources:* (CeBIT, 2017), (Cibils, 2017), (Smart Nation Singapore, 2016), (Watts & Purnell, 2016), (Wen, 2016) and (Data.gov.sg, 2017), (Purnell, 2016), (Lee, 2013).

**Box 6.3: International Examples of Smart Cities: Barcelona**

**Barcelona Smart City**

The **Barcelona Smart City team**, set up in 2011, identified 12 areas of intervention and implemented 22 programmes across the city in areas such as transport, energy, waste and water. Some examples of applications in these sectors are provided below:

- **Transport:** Barcelona invested in a multi-modal strategy. This included the provision of new **digital bus stops**, which are equipped with USB charging stations, free Wi-Fi, tools to facilitate the download of city apps and real-time updates on bus location. In addition, Barcelona has also developed a **smart parking application**, ApparkB, that guides drivers to available parking spaces. This application is based on a sensor network and allows drivers to pay for their parking space online.
- **Energy:** Barcelona has installed **smart meters** to optimise energy use. For street lighting, Barcelona has devised the **Barcelona Lighting Masterplan** which fosters the installation of LED lampposts and the use of sensor technology to regulate the intensity of the light (light is dimmed when there are no pedestrians around) and collect data on air quality. The lampposts are also Wi-Fi hotspots.



Barcelona has also provided additional support to its local technology industry to capitalise on the city's IoT projects. An industrial area was redeveloped using €220 million in public funds to provide a location where start-ups are now putting public data to use in new applications and developing IoT technologies.

The city's published roadmap - [Barcelona Digital City 2017-2020](#) - covers a wide range of the different elements, including the development of digital city infrastructures and actions aimed at reducing the digital divide in the different neighbourhoods, as well as promoting research and innovation in the urban environment.

Source: (Harvard, 2016), (Barcelona City Council, 2016).

## 6.3 Research gaps and potential solutions for the challenges and enablers of transport-related Big Data applications

The key research gaps and recommended actions are summarised in the following Table 6.3.

**Table 6.3: Key understanding/research gaps and recommended actions for the challenges and enablers for transport Big Data applications**

# Key research gaps and potential solutions	
Gaps in research / understanding	
1.	Challenges to the full potential of Big Data are assessed in depth in the literature, but <b>the degree to which they affect organisations of different sizes</b> is not clear. The findings from both the literature review and stakeholder consultation suggest that smaller local authorities and companies might be at a disadvantage, but further research is required to understand the full impacts of these challenges.

# Key research gaps and potential solutions	
2.	The literature on challenges tends to highlight issues that are general to all sectors (not just transport). It would be helpful to develop more detailed insights on the <b>specific challenges for the transport sector</b> and to what extent they prevent the sector from fully benefiting from Big Data. The particular characteristics of the sector and how they affect these challenges could reveal important information on how to address them.
3.	Concerns over <b>data privacy</b> were identified to be one of the key challenges. The literature and stakeholder consultation highlighted that there are several factors at play here; there seems to be a <b>weak legal framework, but also uncertainty regarding the legal requirements and concern over reputational impacts</b> . It would be useful to identify <b>what is at stake and how these different issues affect data sharing</b> in order to be able to address data protection and improve digital rights.
<b>Potential solutions</b>	
4.	The industry could on the one hand, contribute to the improvement of <b>open data standardisation</b> , and on the other hand, <b>address the lack of understanding of the value of data</b> assets as well as the commercially sensitivity issue. Both these enabling actions are thought to address key technical and economic challenges. Further details on industry actions are presented in Section 3 and Section 8.
5.	The Government could take further actions towards adapting <b>Open Government policies and initiatives</b> , as well as creating the conditions for <b>fostering innovation and business development (Open Innovation)</b> .
6.	<b>Improvements to data standardisation</b> are needed to make services and datasets inter-operable, and to address privacy and security concerns. Such improvements will enable and maximise opportunities for data sharing between different organisations, across different sectors.

## 7 Potential links and co-benefits of transport Big Data to other policy areas

### Box 7.1: Summary of key points on connections to other policy areas

- Although there are expected to be a range of co-benefits, no significant literature was found on the potential links between transport-related Big Data and other policy areas.
- A scoping-level assessment was made, which suggested possible co-benefits in the areas of planning, asset management, budgets, air quality, noise, energy, trade, business, skills and employment. Further research is required to examine the potential in more detail.
- Recommended actions include the fostering of strong interdepartmental collaboration, encouraging the development of cross-sectoral solutions, and ensuring effective ex-post analysis of pilot projects.

Big Data, IoT and the Smart Cities context are thought to have enormous cross-sectoral potential for implementation and are expected to have a range of anticipated co-benefits encompassing a range of policy areas.

However, the literature mainly focusses on targeted services and their benefits within specific sectors such as transport, energy and healthcare independently. It appears that no assessment has been made of the cross-linkages between the different areas. Very little information was identified in the literature that explored these potential links and co-benefits, with the exception of a high-level assessment of potential MaaS (mobility-as-a-service) alignment with wider Governmental department policy in (TSC, 2016). Although a range of general energy and emissions benefits were identified in the consultation, the assessment was only at a high level (see Section 2.2 and Appendix 0 and A2).

In the absence of other evidence, a scoping-level inventory of potential impacts/co-benefits resulting from the application of Big Data and IoT-enabled services in Smart Cities is summarised in Table 7.1.

**Table 7.1: Summary of potential cross-departmental / policy agenda impacts resulting from transport-related Big Data and IoT development in Smart Cities**

Government Department	Impact Type	Potential Benefits or Costs
Department for Transport	Planning	<ul style="list-style-type: none"> <li>• Improved information for modelling and urban planning purposes</li> </ul>
	Traffic	<ul style="list-style-type: none"> <li>• Improved traffic management and reduced congestion</li> <li>• Better justification/evidence for new infrastructure projects</li> </ul>
	Assets	<ul style="list-style-type: none"> <li>• Better vehicle and infrastructure asset management (lower costs, higher level of service)</li> </ul>
	Emissions	<ul style="list-style-type: none"> <li>• Reduction in air quality pollutant and GHG emissions</li> </ul>
Treasury	Budget	<ul style="list-style-type: none"> <li>• Improved cost-effectiveness of provision of mobility / end-to-end journeys from local authority and general transport budgets</li> </ul>
Department for Environment, Food and Rural Affairs	Air Quality	<ul style="list-style-type: none"> <li>• Reduction in air quality pollutant emissions from: (a) reduced congestion, (b) potential a shift to public transport and active modes*.</li> </ul>
	Noise	<ul style="list-style-type: none"> <li>• Reduced ambient noise</li> </ul>
Department of Business, Energy and Trade	Energy & Climate Change	<ul style="list-style-type: none"> <li>• Reduction in energy and GHG emissions* due to: (a) more end-end efficient journeys, (b) improved traffic flows, (c) potential shift to lower energy/GHG intensive modes.</li> </ul>
	Trade	<ul style="list-style-type: none"> <li>• Increased attractiveness of UK for businesses</li> </ul>

Government Department	Impact Type	Potential Benefits or Costs
<b>Industrial Strategy</b>	Business	<ul style="list-style-type: none"> <li>Reduction in transport costs and improved productivity through improved network efficiency and better optimising the management of freight movements</li> </ul>
	Skills / Employment	<ul style="list-style-type: none"> <li>Employment in the transport sector</li> <li>Improved travel to work</li> <li>Supporting development of UK skills and innovation, and attraction of skilled overseas workers where necessary</li> </ul>
<b>Department of Health</b>	Health	<ul style="list-style-type: none"> <li>Reduced air quality and noise related health impacts*</li> <li>Support active lifestyle objectives *</li> <li>Reduced stress resulting from congestion</li> <li>Improved social cohesion (ride/car-sharing)</li> </ul>
	Access	<ul style="list-style-type: none"> <li>Improved links to hospitals and other healthcare providers</li> </ul>
<b>Department for Education</b>	Access	<ul style="list-style-type: none"> <li>Improved access to educational establishments</li> <li>Improved application-oriented teaching of STEM, computational and analytics courses at all stages</li> </ul>
<b>Department for Communities and Local Government</b>	Budgets	<ul style="list-style-type: none"> <li>More cost-effective delivery of mobility-based services</li> </ul>
	Mobility	<ul style="list-style-type: none"> <li>Potentially improved access to mobility solutions for vulnerable, low income groups and the older population</li> </ul>
	Services	<ul style="list-style-type: none"> <li>Optimising the management of road-based or road-dependent services such as emergency services and waste collections</li> </ul>
	Business	<ul style="list-style-type: none"> <li>Stimulation of local innovation and other benefits to SMEs</li> </ul>
	Planning	<ul style="list-style-type: none"> <li>Reduced impacts of transport on new developments</li> <li><i>Potential increase in planning requirements for installation of sensors and other equipment</i></li> </ul>

Notes: \* there is a risk that improvements in end-to-end journey planning and the enabling/growth of MaaS (mobility as a service) could result in unintended negative impacts, e.g. greater use of car based transport and a reduction in demand for public transport services, with negative consequences for air quality, GHG emissions.

A summary of the key gaps identified and a recommendation for action to improve understanding in this area are presented in Table 7.2.

**Table 7.2: Key understanding/research gaps and potential solutions for the potential links and co-benefits of transport Big Data to other policy areas**

# Key research gaps and potential solutions	
	<b>Gaps in research / understanding</b>
1.	No significant research into the potential links between transport-related Big Data and other policy areas was identified as part of this scoping study. Further research is needed to elaborate on the potential here, to build on the preliminary analysis conducted for this project.
2.	Further research is required to examine the potential for cross-government collaboration on Big Data and IoT across policy agendas.

# Key research gaps and potential solutions	
Potential solutions	
3.	Development of strong interdepartmental groups to foster cooperative and interconnected strategic development, research and implementation activities, to ensure the full benefits of Big Data and IoT exploitation are achieved as rapidly and as cost-effectively as possible.
4.	To better understand the wider effects of, and opportunities for transport applications, as well as encouraging the development of cross-sectoral solutions, impact assessment and ex-post evaluation analysis of relevant pilot projects could be commissioned and conducted for existing projects. Impact assessment and ex-post evaluation could also become a standard requirement for newly commissioned work. Government, business and academia should work collaboratively in taking actions in this area.

## 8 Summary of findings and potential solutions

This report provides a summary of the findings of the literature review and consultation activities for this scoping study on deriving transport benefits from Big Data and IoT in Smart Cities. In this final section the findings of this work have been used to identify provide suggestions on possible actions / potential solutions that could be taken by the Department for Transport to accelerate the unlocking of the potential of Big Data and IoT for transport.

Currently the evidence base is rather weak with regards to the potential fiscal benefits that could result from more effective utilisation of big data and IoT by the transport sector. However, based on our engagement with stakeholders in this project it is clear that the anticipated benefits could be substantial. Related to this, there do not appear to be many good examples, where well-documented ex-post analysis of project impacts have been carried out to draw out and understand the key lessons learned, to provide effective case studies and to enable the work to be effectively built upon.

At the moment there are significant gaps in the availability of key big datasets that are important for unlocking the full potential of intelligent mobility; in general, there are significant challenges to overcome issues of fragmentation/data silos in transport datasets. However, the literature review and stakeholder consultation both highlighted that the main challenges to filling the majority of the identified gaps are not due to technical challenges, but rather due to other considerations (e.g. commercial business models, security and privacy concerns). For example, competition between service providers (i.e. fear of losing competitive advantage) has been identified as a crucial barrier in the development of open architectures and platforms. There is also likely to be a trade-off between privacy and security and opening up more bespoke functionality for intelligent mobility solutions in some cases that will need to be understood and managed.

Another important consideration is that communities with limited access to the new technologies that make Big Data possible are less likely to be captured (or at least less representative). Provision of services based on these datasets therefore runs the risk of discriminating against certain segments of the population (e.g. those in less affluent or accessible areas) unless this issue is addressed.

It is important to emphasise that there are no 'silver bullets' for unlocking the potential of big data, as the challenges are often varied and complex. Nevertheless, there are certain actions (such as raising awareness of the value of big data, publishing a roadmap, working to address key commercial sensitivities, and addressing uncertainties/gaps in standards) that could help to unlock the potential in the short-term. To deliver a step-change in this area, a sustained work programme would be required over a number of years, to include these elements, as well as further research, pilot projects and support to bridge the gap to develop and deploy innovative and commercially sustainable solutions.

Overall, the key role for Government and city authorities should be to set the direction, identify the key problems to be solved, and to provide appropriate guidance and support to allow the market to develop and deliver the solutions. Establishing an expert panel to advise the Department for Transport on this work programme could be an important step in the process of developing a strategy and prioritising actions in this area. A joined-up approach by the different actors in the public and private sectors could also reduce the likelihood of unnecessary duplication of efforts in this area, so that the maximum benefits can be obtained from the available resources.

A summary of the key potential solutions we have identified in this study is provided in Table 8.1 below, and these are based upon the longer list of identified research gaps and possible actions provided at the end of the earlier report chapters. These actions are also broadly consistent with, and additive to, the actions proposed for DfT at recent Smart Cities and Transport workshops (Department for Transport, 2017) (also included in Appendix A4 of this report).

**Table 8.1: Key recommended actions for the Department for Transport to facilitate the deriving of transport benefits from Big Data and IoT in a Smart Cities context**

Area	#	Potential solutions
<b>Raising awareness, buy-in and coordination</b>	1.	<p><b>DfT could promote the Government's vision, the value and importance of Big Data and Data Analytics, IoT and the Smart City agenda for transport.</b> In particular, ensure key decision-makers in public and private sector organisations fully understand the availability of existing open data, the benefits of its use, possible future needs/applications and the value of sharing their own data by:</p> <ul style="list-style-type: none"> <li>• Convening an external Strategy Panel of experts from Government, business, and academia to provide guidance on strategy and prioritisation of effort.</li> <li>• Actively convening industry and local authority analytics networks to share knowledge/experience and also best practice / case study examples.</li> <li>• Raising awareness and understanding of the value of data for business, end users and local public services.</li> <li>• Publishing a roadmap for the application of Big Data and IoT in transport.</li> <li>• Generate public awareness to create acceptance of the use of big data and to create a demand for services.</li> <li>• Developing strong interdepartmental groups to foster cooperative and interconnected strategic development, research and implementation activities.</li> </ul> <p>This work could build on the work of existing networks, such as the Transport Data Initiative (TDI) with its network of local authorities and industry specialists, and should include the promotion of open data based on ODI (Open Data Institute) good practice.</p>
<b>Standards</b>	2.	<p><b>DfT could address uncertainties and potential gaps in standards for big data, IoT and open architectures.</b> There is a need to review and assess if the existing standards and guidance are fit for purpose, and to implement open data policies. This could involve appropriate networks and working groups, and build upon experiences in other sectors.</p> <p>Particular emphasis could be placed on standards that impact on privacy, cyber security (i.e. of data and IoT devices) and the protection of commercially sensitive data where relevant, to help address concerns in this area.</p> <p>This work could also include establishing a consistent framework to apply for open architectures for public data platforms, which will facilitate the linking and aggregation of city-/regional-level datasets. Standards for meta-data have also been highlighted as particularly important to overcome transport data fragmentation/silos, which should be investigated further.</p> <p>By the end of the process, DfT could provide clear guidance on the appropriate standards that should be applied, and raise awareness among stakeholders in order to ensure these standards are sufficiently visible and applied consistently.</p>

Area	#	Potential solutions
Connecting datasets	3.	<p><b>DfT could ensure the development and maintenance of a publically available national transport data catalogue, and facilitate the standardisation of open architectures for public datasets.</b> Improving connectivity of data is considered crucial to the development of intelligent mobility services and obtaining the most important potential benefits. Applying a consistent framework for public shared data will enable potential to be unlocked at the regional and national levels, and not just at the city level. This should also aim to avoid a focus on individual transport modes; instead the framework should aim to better address complete end-to-end journey considerations.</p> <p>Currently it is believed that city-level datasets may not reach the critical mass needed for intelligent mobility service development, except in larger cities. This work should be supported by further research to better understand the issues that are preventing local authorities from delivering on the big data agenda.</p>
Closing data gaps	4.	<p><b>DfT could conduct further detailed analysis of the degree, nature and importance of different data gaps and test innovative solutions for closing them.</b> The first (short-term) step in this exercise should be to conduct further research that will identify the specific intelligent mobility solutions the data gaps will enable (building on existing work), to help prioritise them. This work should also seek to understand the nature and key causes of the individual data gaps (i.e. format, geography, data quality, commercial or privacy issues) so that more specific solutions can be proposed to address them (e.g. the role of existing or new sensors). The second (medium-term) step should include developing suitable pilot/discovery projects to:</p> <ul style="list-style-type: none"> <li>• Explore, test and implement innovative solutions to address data gaps.</li> <li>• Enhance the economic feasibility of data architecture and innovation platforms by testing new approaches of revenue sharing models between different organisations sharing data.</li> <li>• Develop business case reports for successful pilots/deployments, to emphasise also the wider/non-financial benefits.</li> <li>• Provide further bridging support where necessary for wider commercialisation.</li> </ul>
Privacy, security and commercial sensitivities	5.	<p><b>DfT should continue to work with public commercial organisations to identify ways to improve certainty and overcome concerns over data privacy, security and other commercial sensitivities.</b> This could include clarification and raising awareness on the regulatory/legal requirements (e.g. on privacy, security and competition), to clearly identify what is at stake and how these different issues affect data sharing, in order to be able to address data protection and improve digital rights. In some cases, it may be necessary to address contractual terms for public transport service providers to reduce the potential challenges to collection and sharing of key datasets. These are all potentially quick-win actions.</p>

The above actions should be conducted in a fully collaborative and cross-sectoral manor with both industry and academia, guided by the proposed Strategy Panel. In addition, action needs to be taken more broadly in educational establishments and in the workplace to ensure the sufficient availability of individuals with skills necessary for big data applications are available in the UK. Particularly important government/public sector-led actions identified include:

- Earlier introduction of programming and data analytics in the school curriculum, plus general improvement to uptake, and earlier introduction, of STEM subjects.
- Develop options to help attract skills from outside of the UK, where necessary, and/or keep them within the UK.
- Industry to help develop Big Data and data analytics skills by up-skilling existing workforces.

- 
- Development of new courses (or improvement of existing courses) at university level, including multi-disciplinary training.
  - Improvements in the use or design of existing schemes such as the Apprenticeship Levy to make them work for Big Data analytics.

It will also be important to reverse the reduction of intelligent client roles within local authorities identified in the recent DfT Smart City workshops.

Without taking the above actions for the application of Big Data and IoT to develop intelligent mobility solutions (and to improve logistics efficiency), it is likely that:

- Overall progress will continue but at a much slower pace.
- Applications (and benefits) will remain geographically fragmented (with only the larger cities seeing significant benefits in the short-medium term).
- Certain datasets, and the services that they would unlock, will remain unavailable (or will only be partially available).
- The cost-effectiveness of the application of Big Data and IoT will on average be significantly poorer.
- The UK will be less able to attract overseas investment and key skills in this area.

## 9 References

- Abellá-García, A., de-Urbina-Criado, M. O., & De-Pablos-Heredero, C. (2015). The ecosystem of services around Smart cities: An exploratory analysis. *Procedia Computer Science*, 1075-1080. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1877050915026897>
- Angiuli et al. (2015). How to De-identify Your Data - Balancing statistical accuracy and subject privacy in large social-science data sets. *acmqueue*, 13(8). Retrieved from <http://queue.acm.org/detail.cfm?id=2838930>
- Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, 45-53. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0143622813002178>
- ARUP. (2016). *Smart Cities Opportunities for London*. Greater London Authority (GLA).
- Atkins. (2015). *Journeys of the Future: Introducing Mobility as a Service*.
- Barcelona City Council. (2016). *Barcelona Digital City 2017-2020 plan*. Retrieved from [www.barcelona.cat](http://www.barcelona.cat): <http://ajuntament.barcelona.cat/estrategiadigital/en>
- Batty, M. (2013). *The New Science of Cities*. Retrieved from <https://mitpress.mit.edu/books/new-science-cities>
- Berst, J. (2013). *Why Rio's citywide control center has become world famous*. Retrieved from SmartCitiesCouncil.com: <http://smartcitiescouncil.com/article/why-rios-citywide-control-center-has-become-world-famous>
- Birmingham City Council. (2014). *The Roadmap to a Smarter Birmingham*.
- BITRE. (2014). *New traffic data sources – An overview*. Retrieved from <https://bitre.gov.au/events/2014/files/NewDataSources-BackgroundPaper-April%202014.pdf>
- Bonina, C. (2013). *New business models and the value of open data: definitions, challenges and opportunities*. NEMODE.
- Bowers, K., Buscher, V., Dentten, R., Edwards, M., England, J., Enzer, M., . . . Schooling, J. (n.d.). *Smart Infrastructure. Getting more from strategic assets*. Retrieved from <http://www-smartinfrastucture.eng.cam.ac.uk/files/the-smart-infrastructure-paper>
- BSI. (2017). *Smart cities – Guide to establishing a decision-making framework for sharing data and information services*. Published by BSI (British Standards Institute), developed in partnership with Future Cities Catapult (FCC) and the Department for Business, Energy and Industrial Strategy (BEIS). Retrieved from <http://futurecities.catapult.org.uk/wp-content/uploads/2017/04/PAS-183.pdf>
- Bulger et al. (2014). *Data-Driven Business Models: Challenges and Opportunities of Big data*. A report by Monica Bulger, Greg Taylor and Ralph Schroeder of the Oxford Internet Institute. Retrieved from [http://www.nemode.ac.uk/wp-content/uploads/2014/09/nemode\\_business\\_models\\_for\\_bigdata\\_2014\\_oxford.pdf](http://www.nemode.ac.uk/wp-content/uploads/2014/09/nemode_business_models_for_bigdata_2014_oxford.pdf)

- Burrows, A. B. (2015). *Journeys of the Future: Introducing Mobility as a Service*. Atkins.
- Buscher, V. D. (2014). *Urban Mobility in the Smart City Age*. Arup, The Climate Group, Schneider Electric.
- Business Insider UK. (2016). *How the Internet of Things will transform private and public transportation*. Retrieved from Business Insider UK:  
<http://uk.businessinsider.com/internet-of-things-connected-transportation-2016-10?r=US&IR=T>
- Cabinet Office. (2013). *National Information Infrastructure*. Retrieved from .GOV.UK:  
<https://www.gov.uk/government/publications/national-information-infrastructure>
- Cambridge Centre for Smart Infrastructure & Construction. (2016). *Promoting Adoption of Smart Infrastructure in the Transport Sector - Recommendation to the Department for Transport Chief Scientific Advisor*. Retrieved from  
[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/559770/dft-science-advisory-council-condition-monitoring-and-intelligent-infrastructure-report.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/559770/dft-science-advisory-council-condition-monitoring-and-intelligent-infrastructure-report.pdf)
- Cambridge University. (2015). *New vibration energy harvester powers-up on Forth Road Bridge*. Retrieved from Cambridge University: <http://www-smartinfrasturcture.eng.cam.ac.uk/what-we-do-and-why/focus-areas/sensors-data-collection/projects-and-deployments-case-studies/Novel-Vibration-Energy-Harvester-Deployed>
- CDE. (2017). *Copenhagen City Data Exchange*. Retrieved from Copenhagen City Data Exchange: <https://www.citydataexchange.com/#/home>
- CeBIT. (2017). *Singapore: smart city, smarter nation*. Retrieved from CeBIT Australia:  
<http://blog.cebit.com.au/singapore-smart-city>
- Cebr. (2016). *The Value of Big Data and the Internet of Things to the UK Economy*. A report by Cebr for SAS. Retrieved from  
[https://www.sas.com/content/dam/SAS/en\\_gb/doc/analystreport/cebr-value-of-big-data.pdf](https://www.sas.com/content/dam/SAS/en_gb/doc/analystreport/cebr-value-of-big-data.pdf)
- Chan, C. (2013). *From open data to open innovation strategies: Creating e-services using open government data*. System Sciences (HICSS) 2013 46th Hawaii International Conference.
- Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C*, 285-299. Retrieved from  
<http://www.sciencedirect.com/science/article/pii/S0968090X16300092>
- Cibils, T. (2017). *SINGAPORE: ALREADY OVER BIG DATA WITH SMART DATA*. (s. S. Written by Thomas Cibils, Producer) Retrieved from  
<https://globalstatement2015.wordpress.com/2015/06/03/singapore-already-over-big-data-with-smart-data/>
- CIO. (2016). *BOB BROWN - MANCHESTER CITY COUNCIL*. Retrieved from CIO:  
<http://www.cio.co.uk/cio100/2016/bob-brown/>

- Coleman, E. (2013). *Beyond Transparency. CHAPTER 4 - Lessons from the London Datastore*. Code for America. Retrieved from <http://beyondtransparency.org/chapters/part-1/lessons-from-the-london-datastore/>
- Cosgrave et al. (2014). *Delivering the smart city*.
- Coursimault, C. (2013). *Motionloft analyzes people and vehicle traffic for commercial and civic purposes*. Retrieved from L'Atelier: [http://www.atelier.net/en/trends/articles/motionloft-analyzes-people-and-vehicle-traffic-commercial-and-civic-purposes\\_421214](http://www.atelier.net/en/trends/articles/motionloft-analyzes-people-and-vehicle-traffic-commercial-and-civic-purposes_421214)
- Daries et al. (2014). Privacy, anonymity, and big data in the social sciences. *acmqueue*, 12(7), 56-63. Retrieved from <http://queue.acm.org/detail.cfm?id=2661641>
- Data.gov.sg. (2017). *Taxi Availability*. Retrieved from Data.gov.sg: <https://data.gov.sg/dataset/taxi-availability>
- Dataflog. (n.d.). *How Big Data Can Create A Smarter Transportation Industry*.
- Deloitte & Touche (M.E.). (2015). *Smart cities...Not just the sum of its parts*. Deloitte & Touche (M.E.).
- Deloitte. (2013). *Market Assessment of Public Sector Information*. Department for Business, Innovation & Skills and Cabinet Office. Retrieved from <https://www.gov.uk/government/publications/public-sector-information-market-assessment>
- Deloitte and ODI. (2012). *Open growth: Stimulating demand for open data in the UK*.
- Department for Business Innovation & Skills. (2013). *The Smart City Market: Opportunities for the UK*. Retrieved from [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/249423/bis-13-1217-smart-city-market-opportunities-uk.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/249423/bis-13-1217-smart-city-market-opportunities-uk.pdf)
- Department for Transport. (2017). *Smart Cities and Transport Workshop Report*. Department for Transport.
- DeVan, A. (2016, April). *The 7 V's of Big Data*. Retrieved from Impact Radius: <https://www.impactradius.com/blog/7-vs-big-data/>
- Development Economics. (2013). *The Future Digital Skills Needs of the UK Economy*. A report for O2 by Development Economics. Retrieved from <http://news.o2.co.uk/wp-content/uploads/2013/09/The-Future-Digital-Skills-Needs-of-the-UK-Economy1.pdf>
- DHL. (2013). *Big data in logistics*.
- Douglass, J. (2016). *How Our Existing Data Can Benefit Individuals And Local Communities*.
- Economist. (2010). *Data, data everywhere*. Retrieved from The Economist: <http://www.economist.com/node/15557443>
- Elgin. (n.d.). *Case Study. Transforming traffic management at Cheshire East Council*.
- ENISA. (2015). *Cyber security for Smart Cities: An Architecture model for public transport*.

- Environmental Industries Commission. (2014). *Getting the green light: Will smart technology clean up city environments?*
- European Innovation Partnership on Smart Cities and Communities. (2013). *Strategic Implementation Plan*.
- eurotransport. (2016). *How TfL uses 'big data' to plan transport services*. Retrieved from eurotransport: <http://www.eurotransportmagazine.com/19635/past-issues/issue-3-2016/tfl-big-data-transport-services/>
- EY. (2015). *Routes to prosperity. How smart transport infrastructure can help cities to thrive*. Retrieved from [http://www.ey.com/Publication/vwLUAssets/EY-routes-to-prosperity-via-smart-transport/\\$FILE/EY-routes-to-prosperity-via-smart-transport.pdf](http://www.ey.com/Publication/vwLUAssets/EY-routes-to-prosperity-via-smart-transport/$FILE/EY-routes-to-prosperity-via-smart-transport.pdf)
- GLA. (2017). *London DataStore*. Retrieved from London DataStore: <https://data.london.gov.uk/>
- Glasgow City Council. (2012). *Glasgow City Management System*. Retrieved from <https://connect.innovateuk.org/documents/3130726/3794125/Feasibility+Study+-+Glasgow+City+Council.pdf>
- GO Science. (2014). *The Internet of Things: making the most of the Second Digital Revolution*.
- Greater London Authority. (2016). *Data for London: A City Data Strategy*.
- Greater London Authority/Arup. (2016). *Smart City Opportunities for London*.
- Hartmann et al. (2014). *Big Data for Big Business? A Taxonomy of Data-driven Business Models used by Start-up Firms*. Cambridge Service Alliance.
- Hartmann, P. M. (2014). *Data from big business: A Taxonomy of Data-driven Business Models used by Start-up Firms*. University of Cambridge.
- Harvard. (2016). *How Smart City Barcelona Brought the Internet of Things to Life*. Retrieved from <http://datasmart.ash.harvard.edu/news/article/how-smart-city-barcelona-brought-the-internet-of-things-to-life-789>
- Hashema, I. A., Chang, V., Anuar, N. B., K. A., I. Y., Gani, A., . . . Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management*, 748-758. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0268401216302778>
- HBR. (2014). *The new patterns of innovation*.
- Hehenberger et al. (2016). Design, modelling, simulation and integration of cyber physical systems: Methods and applications. By Hehenberger, P., Vogel-Heuser, B., Bradley, D., Eynard, B., Tomiyama, T., Achiche, S. *Computers in Industry, Volume 82*, p 273-289.
- Hitachi. (2017). *Taking Cities to the Next Level - The City Data Exchange*. Retrieved from Hitachi Insight Group: <https://www.hitachiinsightgroup.com/en-us/pdf/solution-profile/hitachi-solution-profile-city-data-exchange.pdf>

- HM Government. (2013). *Seizing the data opportunity. A strategy for UK data capability*. Retrieved from [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/254136/bis-13-1250-strategy-for-uk-data-capability-v4.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254136/bis-13-1250-strategy-for-uk-data-capability-v4.pdf)
- House of Commons, Science and Technology Committee. (2016). *The big data dilemma*. Retrieved from <http://www.publications.parliament.uk/pa/cm201516/cmselect/cmsstech/468/468.pdf>
- Houses of Parliament, Parliamentary Office Of Science & Technology. (2014). *Big and Open Data in Transport*.
- Huawei/Navigant. (2016). *UK Smart Cities Index Assessment of Strategy and Execution of the UK's Leading Smart Cities*.
- IBM. (2014). *Building a smarter transportation management network*.
- IBM. (2014b). *Big data and analytics in travel and transportation*.
- IBM. (2017). *4 Vs*. Retrieved from IBM Big Data & Analytics Hub: <http://www.ibmbigdatahub.com/tag/587>
- IET. (2013). *Big data in transport*.
- IET. (2017). *The Local Authority Guide to Emerging Transport Technology 2017-2018*.
- InsideBIGDATA. (2013). *Beyond Volume, Variety and Velocity is the Issue of Big Data Veracity*. Retrieved from insideBIGDATA: <http://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/>
- IoTTechWorld. (2015). *The impact of Internet of Things at the Transportation System*. Retrieved from Internet of Things Tech World: <http://www.iotechworld.com/automotive/the-impact-of-internet-of-things-at-the-transportation-system.html>
- ITF. (2015). *Big Data and Transport. Understanding and assessing options*. International Transport Forum, OECD.
- J. Leon Zhao, S. F. (2014). Business challenges and research directions of management analytics in the big data era. *Journal of Management Analytics*, 169-174.
- Jaffe, E. (2016). *4 lessons from Rio's 'flawed' smart cities initiative*. Retrieved from Medium.com: <https://medium.com/sidewalk-talk/4-lessons-from-rios-flawed-smart-cities-initiative-31cbf4e54b72>
- Judie Attard, F. O. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*.
- Kagermann, H. W. (2013). *Recommendations for implementing the strategic initiative Industrie 4.0: Final report of the Industrie 4.0 Working Group*. German Federal Ministry of Education and Research. Retrieved from <http://www.acatech.de/de/publikationen/stellungnahmen/kooperationen/detail/artikel/r>

ecommandations-for-implementing-the-strategic-initiative-industrie-40-final-report-of-the-industr.html

- Kobielus, J. (2016, December). *Big data and analytics trends in 2017: James Kobielus's predictions*. Retrieved from IBM Bg Data & Analytics Hub: <http://www.ibmbigdatahub.com/blog/big-data-and-analytics-trends-2017-james-kobielus-s-predictions>
- Komorowski, M. (2014). *A history of storage cost (update)*. Retrieved from mkomo.com: <http://www.mkomo.com/cost-per-gigabyte-update>
- Lee, T. (2013, Novemeber). *Singapore an advanced surveillance state, but citizens don't mind*. Retrieved from TechInAsia: <https://www.techinasia.com/singapore-advanced-surveillance-state-citizens-mind>
- LowCVP. (2017). *Bus renaissance under way in some UK towns and cities*. Retrieved from Low Carbon Vehicle Partnership: [http://www.lowcvp.org.uk/news,bus-renaissance-under-way-in-some-uk-towns-and-cities\\_3572.htm](http://www.lowcvp.org.uk/news,bus-renaissance-under-way-in-some-uk-towns-and-cities_3572.htm)
- McKinsey & Company. (2015). *Big data versus big congestion: Using information to improve transport*.
- Merswolke, B. (2003). *The Internet of Things and Transportation* . Retrieved from miovision: <https://miovision.com/blog/the-internet-of-things-and-transportation/>
- MGI. (2011). *Big data: The next frontier for innovation, competition, and productivity*. A report by McKinsey Gobal Institute. Retrieved from <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>
- Nesta. (2015). *Rethinking Smart Cities from the Ground Up*.
- Nesta. (2015a). *Analytic Britain: Securing the right skills for the data-driven economy*. Policy briefing by Nesta and Universities UK. Retrieved from [http://www.nesta.org.uk/sites/default/files/analytic\\_britain.pdf](http://www.nesta.org.uk/sites/default/files/analytic_britain.pdf)
- Network Rail. (2017). *Data Feed information*. Retrieved from Network Rail Data Feeds: <https://datafeeds.networkrail.co.uk/ntrod/login>
- ODI. (2017). *Closed, shared, open data: what's in a name?* Retrieved from Open Data Institute: <https://theodi.org/blog/closed-shared-open-data-whats-in-a-name>
- ODI. (2017). *The Data Spectrum helps you understand the language of data*. Retrieved from Open Data Institute: <https://theodi.org/data-spectrum>
- OneM2M. (2017). *Smart Cities Done Smarter*.
- Open Data Institute. (2016). *Open enterprise: how three big businesses create value with open innovation*.
- Open Rail Data wiki. (2017). *Main Page*. Retrieved from Open Rail Data wiki: [http://nrodwiki.rockshore.net/index.php/Main\\_Page](http://nrodwiki.rockshore.net/index.php/Main_Page)
- O'Reilly, T. (2014). *Strip Out Complexity and Open Your Data – How Cities Can Open Their Data Quickly and Easily*. Retrieved from DSRPTN: <https://disruptionltd.com/2014/02/14/182/>

- Osborne Clarke. (2015). *Smart cities in Europe. The future of urban mobility.*
- Oxford Internet Institute. (2014). *Data-Driven Business Models: Challenges and Opportunities of Big Data.*
- Parliamentary Office of Science & Technology. (2014). *Big and Open Data in Transport, POSTnote 472.*
- Policy Exchange. (2016). *Smart Devolution: Why smarter use of technology and data are vital to the success of city devolution.*
- Purnell, N. (2016, April). *Singapore Is Taking the 'Smart City' to a Whole New Level.* Retrieved from The Wall Street Journal : <https://www.wsj.com/articles/singapore-is-taking-the-smart-city-to-a-whole-new-level-1461550026>
- RAND. (2015). *Travel in Britain in 2035.*
- Robinson, N. (2015). *Firms cut their costs by sharing trucks.* Retrieved from Food manufacture.co.uk: <http://www.foodmanufacture.co.uk/Supply-Chain/More-food-firms-should-share-logistics>
- Royal Academy of Engineering. (2015). *Connecting data. Driving productivity and innovation.* Retrieved from <http://www.raeng.org.uk/publications/reports/connecting-data-driving-productivity>
- RTIG, UTMC. (2015). *Smart Cities: a joint position paper.*
- Ruiz-Alarcon-Quintero, C. (2016). Harmonization of transport data sources according to INSPIRE data specification on transport networks. *Transportation Research Procedia*, 320-327. Retrieved from <http://www.sciencedirect.com/science/article/pii/S2352146516307980>
- Rusitschka, S., & Curry, E. (2016). *Chapter 13. Big Data in the Energy and Transport Sectors.* Retrieved from [http://link.springer.com/chapter/10.1007/978-3-319-21569-3\\_13](http://link.springer.com/chapter/10.1007/978-3-319-21569-3_13)
- Sánchez-Martínez, G. E., & Munizaga, M. (2016). Workshop 5 report: Harnessing big data. *Research in Transportation Economics*, 236-241. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0739885916301494>
- Sarma et al. (1999). *The Networked Physical World: Proposals for the Next Generation of Computing Commerce, and Automatic Identification.* A White Paper by Sanjay Sarma, David L. Brock & Kevin Ashton, Auto-ID Center, Massachusetts Institute of Technology (MIT). Retrieved from [http://cocoa.ethz.ch/downloads/2014/06/None\\_MIT-AUTOID-WH-001.pdf](http://cocoa.ethz.ch/downloads/2014/06/None_MIT-AUTOID-WH-001.pdf)
- Schneider Electric et al. (2014). *Urban mobility in the smart city age.* A report by Schneider Electric, ARUP and The Climate Group. Retrieved from [http://publications.arup.com/publications/u/urban\\_mobility\\_in\\_the\\_smart\\_city\\_age](http://publications.arup.com/publications/u/urban_mobility_in_the_smart_city_age)
- Schroeder. (2016). *Big data business models: Challenges and opportunities.* Cogent Social Sciences.

- SELECT for Cities. (2017). *Competition to create large scale city innovation labs*. Retrieved from SELECT for Cities: <http://www.select4cities.eu/>
- Smart London Board. (2013). *Smart London Plan: Using the creative power of new technologies to serve London and improve Londoners' lives*.
- Smart Nation Singapore. (2016). *Smart Nation Singapore*. Retrieved from Smart Nation Singapore: <https://www.smartnation.sg/>
- SMMT. (2017). *Connected and Autonomous Vehicles - Position Paper*. The Society of Motor Manufacturers and Traders (SMMT) Limited. Retrieved from <https://www.smmt.co.uk/wp-content/uploads/sites/2/SMMT-CAV-position-paper-final.pdf>
- Soga et al. (2016). *Infrastructure Sensing*. A paper by Soga, K., Schooling, J. published by The Royal Society in Interface Focus 6:20160023. Retrieved from <http://rsfs.royalsocietypublishing.org/content/6/4/20160023.article-info>
- SPRING Singapore. (2016, Januray). *Setting the standard worldwide: intelligent city, Smart Nation*. Retrieved from SPRING Singapore. Enabling Enterprise: <https://www.spring.gov.sg/Inspiring-Success/Enterprise-Stories/Pages/Setting-the-standard-worldwide-intelligent-city-Smart-Nation.aspx>
- Symantec. (2013). *Transformational 'smart cities': cyber security & resilience*.
- Tableu. (n.d.). *Solving the Internet of Things' Last-Mile Problem*.
- TechUK. (2016). *Understanding, Demystifying and Addressing the UK's Big Data Skills Gap*. Retrieved from <https://www.techuk.org/insights/reports/item/9469-the-uk-s-big-data-future-mind-the-gap>
- TfL. (2015). *Safety, Accessibility and Sustainability Panel*. Retrieved from <http://content.tfl.gov.uk/sasp-20151015-part-1-item12-leveraging-our-data-for-maximum-impact.pdf>
- TfL. (2017). *Taxi and private hire requirements*. Retrieved from Transport for London: <https://tfl.gov.uk/modes/driving/ultra-low-emission-zone/taxi-and-private-hire-requirements>
- (n.d.). *The best of both worlds: how one TRANSPORT improves on established models of Public-Private Partnership*.
- The Institution of Engineering and Technology. (n.d.). *Big Data in Transport*.
- Transport Systems Catapult. (2015). *The Transport Data Revolution - Investigation into the data required to support and drive intelligent mobility*. Retrieved from <https://ts.catapult.org.uk/wp-content/uploads/2016/04/The-Transport-Data-Revolution.pdf>
- TSC. (2015). *The Transport Data Revolution - Investigation into the data required to support and drive intelligent mobility*. Transport Systems Catapult (TSC). Retrieved from <https://ts.catapult.org.uk/wp-content/uploads/2016/04/The-Transport-Data-Revolution.pdf>

- TSC. (2015a). *Traveller Needs and UK Capability Study: Supporting the realisation of Intelligent Mobility in the UK*. Transport Systems Catapult.
- TSC. (2016). *Mobility as a Service: Exploring the opportunity for mobility as a service in the UK*. Transport Systems Catapult.
- TTT. (2016). *TomTom to help Moscow design an intelligent road transportation program*. Retrieved from Traffic Technology Today: <http://www.traffictechnologytoday.com/news.php?NewsID=80333>
- Ubaldi, B. (2013). *Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives*. OECD Working Papers on Public Governance, No. 22, OECD.
- Uber. (2017). *Cities*. Retrieved from Uber Movement: <https://movement.uber.com/cities>
- UK Parliament. (2017). *Bill documents — Digital Economy Bill 2016-17*. Retrieved from Uk Parliament Website: <http://services.parliament.uk/bills/2016-17/digitaleconomy/documents.html>
- UK Parliament. (2017a). *Bill documents — Bus Services Bill [HL] 2016-17*. Retrieved from UK Parliament Website: <http://services.parliament.uk/bills/2016-17/busservices/documents.html>
- UK Parliament. (2017a). *Digital Economy Act 2017*. Retrieved from [www.parliament.co.uk](http://www.parliament.co.uk): <http://services.parliament.uk/bills/2016-17/digitaleconomy.html>
- University of Cambridge. (2016). *Smart Infrastructure - Getting more from strategic assets*. University of Cambridge. Retrieved from <http://www-smartinfrasturcture.eng.cam.ac.uk/files/the-smart-infrastructure-paper>
- Urban Transport Group. (2016). *Challenges and Opportunities for*.
- Venturebeat. (2015). *Here's what IoT will do for transportation*. Retrieved from Venturebeat: <http://venturebeat.com/2015/09/19/heres-what-iot-will-do-for-transportation/>
- Venturebeat. (2015a). *Parkopedia now predicts where and when on-street parking will be available*. Retrieved from VentureBeat: <http://venturebeat.com/2015/09/17/parkopedia-now-predicts-where-and-when-on-street-parking-will-be-available/>
- Watts, J. M., & Purnell, N. (2016). *Singapore Is Taking the 'Smart City' to a Whole New Level*. Retrieved from The Wall Street Journal: <https://www.wsj.com/articles/singapore-is-taking-the-smart-city-to-a-whole-new-level-1461550026>
- Wen, C. S. (2016). *Do We Need Big Data to Create Smart Cities?* Retrieved from Fair Observer: <https://www.fairobserver.com/culture/do-we-need-big-data-to-create-smart-cities-42108/>
- Xiong, L. D. (2015). Big data analytics and business analytics. *Journal of Management Analytics*, 1-21.

Zoonen, L. v. (2016). Privacy concerns in smart cities. *Government Information Quarterly*, 472-480. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0740624X16300818>

## 10 Acknowledgements

The authors and the Department for Transport would like to thank and acknowledge the following organisations, in Table 10.1 below, that took part in the consultation activities for this study and which provided important contributions to the work.

Particular thanks and acknowledgement goes to the Transport Systems Catapult (TSC) and the authors of their report *“The Transport Data Revolution - Investigation into the data required to support and drive intelligent mobility”* (TSC, 2015), which has provided invaluable information for this scoping study’s evidence base.

**Table 10.1: Organisations contributing to the consultation and wider evidence base for the project**

Organisation	
Confederation of Passenger Transport (CPT)	London Underground
Dataflog	Newcastle University
Department of Culture, Media and Sport (DCMS)	Open Data Institute (ODI)
ELGIN	Oxford Bus Company
FIWARE	Siemens
Flying Binary Ltd	techUK
Future Cities Catapult (FCC)	Transport Data Initiative (TDI)
GSMA	Transport for Greater Manchester (TfGM)
Heriot Watt University	Transport for London
IBM	Transport Systems Catapult (TSC)
Immense Simulations	UK Transport Research Laboratory (TRL)
Imperial College London	University of Cambridge
Institution of Engineering and Technology (IET)	University of Westminster
InterDigital Europe Ltd	Vodafone UK
Knowledge Transfer Network (KTN)	Volvo Group UK

## Appendices

Appendix 1 – Final Delphi Survey Summary

Appendix 2 – Summary of Benefits, Challenges and Enablers

Appendix 3 – Examples of key datasets

Appendix 4 – DfT Actions Proposed in the Smart Cities and Transport Workshop Report

Appendix 5 – The potential of Big Data to inform new predictive analytics

## A1 Appendix 1 – Final Delphi Survey Summary

The summary of the results of the Delphi Survey activity conducted for this project will be provided to DfT as a separate summary file.

## A2 Appendix 2 – Summary of Benefits, Challenges and Enablers

**Table A1: Stakeholder feedback in regard to the benefits, challenges and enablers for transport Big Data**

	<b>BENEFITS</b>	<b>CHALLENGES</b>	<b>ENABLERS</b>
<b>Data Generation</b>	<p>Stakeholders recognised three main types of benefits associated with the generation of data.</p> <ul style="list-style-type: none"> <li>• The first one is monetary. Data generation is seen as an opportunity to stimulate the economy. For example, industries that produce sensors technology can benefit from the production of sensors that will be purchased to generate data.</li> <li>• The second type is associated to the opportunity of generating high-quality of data in a more efficient manner.</li> <li>• The third one refers to the general benefit of generating data that can inform decision and policy making.</li> </ul> <p>Finally, among stakeholders there is an agreement that data generation, in order to be beneficial, requires data to be used in a specific way.</p>	<p>Overall stakeholders see the following as being the main challenges to data generation:</p> <ul style="list-style-type: none"> <li>• Standards: generating data according to mutually recognised standard is difficult. As a result, data are generated often using different standards which decrease the level of data quality.</li> <li>• There are technical, political and economic challenges in getting sensors installed (e.g. often the process is not effective and efficient); There is a general lack of viable business model for generating data (e.g. running costs for maintaining and collecting data are high).</li> <li>• Lack of awareness and understanding on how to best share data, which limits the scope and opportunity for generating them in the first place.</li> </ul>	<p>Stakeholders have recognised some key enabling factors to improve data generation:</p> <ul style="list-style-type: none"> <li>• There is a need for improved policy strategies and strategic roadmaps by national and local government.</li> <li>• More funding, especially in the form of seed funding, is seen as necessary to support the initial and experimental stage of data generation industries, embed data generation processes in public procurement contract with major transport service providers.</li> </ul>

	BENEFITS	CHALLENGES	ENABLERS
Data analytics	<ul style="list-style-type: none"> <li>Stakeholders see the benefit of data analytics in the improved decision making, as data analysis can provide new insights.</li> </ul>	<ul style="list-style-type: none"> <li>The majority of stakeholders think that one of the main challenges to data analytics is the lack of skills in cleaning data and making the best use of it. Associated to this barrier, one stakeholder has also pointed out that some organisations might struggle to carry out useful data analytics because of lack of internal capacity (e.g. resources and staff time).</li> <li>Secondly, data quality and lack of relevant data is seen as another key barrier to data analytics.</li> <li>Third, refer to the methods, (e.g. lack of common data framework platforms, models and tools).</li> </ul>	<ul style="list-style-type: none"> <li>Stakeholders recognise that the enabling factors to improve data analytics are the following:                             <ul style="list-style-type: none"> <li>TECHNICAL: need for innovative datasets, new technology and methodology for data analytics (e.g. new tools for data visualisation).</li> <li>ECONOMIC: need to clarify the value of data.</li> <li>TIME available to carry out data analytics.</li> <li>POLICY strategy to give direction to the use of data.</li> </ul> </li> </ul>
Data Exploitation	<ul style="list-style-type: none"> <li>Some stakeholders recognised data exploitation benefits that are common to every sector. For example: improved service delivery, market efficiency and customer experience, which as a result can improve the business and organisational value.</li> <li>Data exploitation is beneficial to policy makers, government and service providers as it provides intelligence and can inform strategy and agendas.</li> <li>Specific to the transport sector, different examples were mentioned such as: real-time traffic light scheduling, improved transport scheduling, smart transport systems solutions to deal with extreme weather conditions.</li> <li>Overall data exploitation is seen as an opportunity to make transport systems more effective and efficient while reducing their environmental impact.</li> </ul>	<ul style="list-style-type: none"> <li>The majority of stakeholders think the lack of a clear business model is the key barrier in data exploitation.</li> <li>One stakeholder thinks that the lack of universally recognised / common standards can also limit data exploitation.</li> <li>Finally, it is mentioned that there is a lack of understanding on what are end-user needs in regard of data exploitation, which limits opportunities.</li> </ul>	<ul style="list-style-type: none"> <li>Stakeholders have mixed views regarding what are the enabling factors for improving data exploitation.</li> <li>Some think that there is a need for new and successful business models. Associated to this, few stakeholders seem to agree on the fact that data should be made further open and easier to access. Such approach would also require that both public and private organisations should make further investment to improve data accessibility and use.</li> <li>Others think that improved quality of data is also required in order to maximise data exploitation.</li> <li>Finally, a couple of stakeholders think that the public sector should play a bigger role as in being further involved and in investigating how to collaborate with private operators.</li> </ul>

	<b>BENEFITS</b>	<b>CHALLENGES</b>	<b>ENABLERS</b>
<b>Data Linking</b>	<p>The majority of stakeholders seem to agree that the key benefit associate to data linking and data integration is the one of improving knowledge and understandings of events. For example, being able to integrate real-time data with historical data has a great potential in developing new knowledge and provide high quality intelligence. (e.g. It allows to carry out trends analysis). This can have secondary positive effects such as creating new opportunities for business development (e.g. develop new services).</p>	<p>Stakeholders identified four key types of challenges to data linking.</p> <ul style="list-style-type: none"> <li>• One relates to organisational cultures and to the fact that there seem to be a general lack of understanding on how data can be used across different sectors in a collaborative way.</li> <li>• The second one is of technical nature and refers to the challenge of working with different standards and data forms, which decreases the interoperability of data.</li> <li>• Third, stakeholders recognise a general lack of specific skills in understanding how to best link and make data interoperable across different sectors.</li> <li>• Last but not least, data ownership and privacy are two key concerns which limit the opportunities around data linking.</li> </ul>	<p>Stakeholders have mixed views about key enabling factors for data linking.</p> <ul style="list-style-type: none"> <li>• Some stakeholders emphasised the need for improved standards and protocols to facilitate a higher interoperability of datasets.</li> <li>• However, there was also general agreement on the need for a clearer governance regarding the use of data across different sectors. For example, one respondents call for a strategic authority to facilitate required processes.</li> </ul>
<b>Data Sharing</b>	<ul style="list-style-type: none"> <li>• Stakeholders see data sharing key benefit in the opportunity of promoting new business and service opportunities which are innovative and that promote cross-sector innovation. Data sharing is also seen as benefitting SMEs growth.</li> <li>• In addition, improved data sharing is considered very important in maximising the value of data and increasing the chances of finding solutions to problems.</li> </ul>	<p>Stakeholders recognised the following challenges to data sharing:</p> <ul style="list-style-type: none"> <li>• PRIVACY and SECURITY concerns.</li> <li>• COMMERCIAL SENSITIVITY of data.</li> <li>• Silos-driven organisational culture, especially in the public sector (e.g. "local authorities tend to focus on the areas within their boundaries. Little cross-broader sharing takes place.").</li> <li>• Standards and data forms through which data can be shared are not uniform and often represent a challenge to data sharing.</li> <li>• Lack of specialised skills.</li> </ul>	<p>Stakeholder recognise the following to be enabling factors of data sharing:</p> <ul style="list-style-type: none"> <li>• Common processes and data standards.</li> <li>• Clearer leadership and governance regarding data sharing.</li> <li>• Develop platforms to share knowledge of best practices and successful experience to improve performance across different sectors.</li> <li>• Provide specialised training,</li> <li>• Develop tailored business models to data sharing.</li> </ul>

## A3 Appendix 3 – Examples of key datasets and transport data portals

### A3.1 Examples of the meta-analysis of key datasets provided in Annex B of TSC (2015)

Table A2: Examples of the meta-analysis of key datasets provided in Annex B of (TSC, 2015)

Theme	1. Place & Space	2. Environment	3. People, Things & Movement	4. Disruption and event-related data	5. Public Transport Services
<b>Example Datasets</b>	OpenStreetMap database	3 Hourly weather forecast and observational data	INRIX - Real-Time Traffic Information	National rail future engineering work	Traveline National Data Set - route and schedule information
<b>Data Assembly processes</b>	Overtly crowd sourced	Publicly collected or brokered	Privately collected	Privately collected	Collated from operators by local authorities and transport agencies by data brokers at a regional level
<b>References</b>	<a href="http://www.openstreetmap.org">http://www.openstreetmap.org</a>	<a href="http://data.gov.uk/dataset/metoffice_uklocs3hr_fc">http://data.gov.uk/dataset/metoffice_uklocs3hr_fc</a>	<a href="http://www.inrix.com/">http://www.inrix.com/</a>	<a href="http://www.nationalrail.co.uk/service_disruptions/currentAndFuture.aspx">http://www.nationalrail.co.uk/service_disruptions/currentAndFuture.aspx</a>	<a href="http://www.travelinedata.org.uk/broker.htm">http://www.travelinedata.org.uk/broker.htm</a>
<b>Maintained by</b>	Crowd-sourcing	Met Office	INRIX	National Rail	Multiple data brokers
<b>Coverage</b>	Worldwide	Great Britain	North America, Europe South America and Africa	Rail Network	Great Britain
<b>NII?*</b>	No	Yes	No	No	Yes
<b>Available publicly?</b>	Yes	Yes		Yes	Yes
<b>Data format</b>	XML DTD / PBF	HMTL			TransXchange
<b>How is it made available</b>	API, Direct Download	Download	API (XML)	Website	API (XML)
<b>Free/Paid</b>	Free	Free	Paid-for private	Free	Free
<b>Temporal Availability</b>	Minute by minute	Hourly	Historic, Real-time	Regular updates	Weekly
<b>Static/Dynamic</b>	Dynamic	Dynamic	Static	Static	Dynamic

Notes: \* Whether it is part of the UK's National Information Infrastructure (Cabinet Office, 2013)

Table A2: Examples of the meta-analysis of key datasets provided in Annex B of (TSC, 2015) (continued)

Theme	6. Personal Automobility	7. Freight connections	8. International Connections	9. Consumption & transaction data
Example Datasets	Waze	Road Freight Statistics	Airport parking data	Smartcard data for London
<b>Data Assembly processes</b>	Overtly crowd sourced	Privately collected	Parking brokers	Derived from Oyster tap on/off
<b>References</b>	<a href="http://www.waze.com">www.waze.com</a>	<a href="http://data.gov.uk/dataset/road_freight_statistics">http://data.gov.uk/dataset/road_freight_statistics</a>	<a href="http://www.airport-parking-shop.co.uk/">http://www.airport-parking-shop.co.uk/</a>	<a href="http://www.tfl.gov.uk/businessandpartners/syndication/16493.aspx">http://www.tfl.gov.uk/businessandpartners/syndication/16493.aspx</a>
<b>Maintained by</b>	Crowd sourcing	Department for Transport	Airport Parking Shop	Cubic for TfL
<b>Coverage</b>	Worldwide	UK	UK	London
<b>NII?*</b>	No	No	N	N
<b>Available publicly?</b>	Yes	Yes	Yes	Yes
<b>Data format</b>		HML	API	CSV
<b>How is it made available</b>	API (XML)	Download	API	CSV
<b>Free/Paid</b>	Free	Free	Free	Free
<b>Temporal Availability</b>	Real-time	Annually	0	Sample from Nov 2009
<b>Static/Dynamic</b>	Dynamic	Static	Dynamic	Static

## A3.2 Examples of transport-related data portals

**Table A3: Transport-related data portals**

Type	Description	Examples	Run by
Distributed data catalogues	Available data from multiple organisations are published via a searchable online catalogue. Datasets remain stored with the data owners.	<a href="#">EU Open Data Portal</a>	European Commission
		<a href="#">data.gov.uk</a>	Central Govt. (Cabinet Office)
		<a href="#">London Data Store</a>	Greater London Authority
		<a href="#">Datahub/Transport</a>	Open Knowledge Foundation Transport Open Working Group
Individual data catalogues	A single organisation's data is published via its own online catalogue. Datasets remain stored by each data owner, and may also be accessible via other data platforms.	<a href="#">Open Data Nottingham</a>	Local Government *
		<a href="#">Birmingham Data Factory</a>	
		<a href="#">Redbridge</a>	
		<a href="#">Edinburgh Open Data</a>	
		<a href="#">Glasgow Data</a>	
		<a href="#">Open Data Bristol</a>	
		<a href="#">Camden</a>	
		<a href="#">Data Mill North</a>	
		<a href="#">Oxfordshire Open Data</a>	
		<a href="#">Sheffield City Council</a>	
		<a href="#">Transport for Greater Manchester</a>	Transport Agency
		<a href="#">TfL open data</a>	
		<a href="#">ATOC</a>	Private sector
<a href="#">Network Rail Data</a>			
Consolidated 'Big' data platforms / ecosystems	Available data are consumed and stored ('warehoused') so they can be combined and served-up on-demand from a single interface.	<a href="#">Met Office</a>	Public Sector
		<a href="#">Stride</a>	Private sector (seed-funded by government)
		<a href="#">i-Move</a>	
		<a href="#">Google Maps APIs</a> Places, Directions, Distance Matrix, Geocoding, Weather	Private sector
		<a href="#">INRIX Traffic</a>	Private sector
		<a href="#">TomTom</a>	Private sector
Distributed data aggregation platforms / ecosystems	Provide a gateway into available data from multiple sources using APIs in order to allow for real-time fusion and consumption of datasets.	<a href="#">smart streets</a>	Private sector (seed-funded by government)
		<a href="#">Transport API</a>	Private sector
		<a href="#">ITO! World's 'Transport DMP'</a>	
		<a href="#">ProgrammableWeb</a>	

Notes: Reproduced from (TSC, 2015) \*Additional entries have been added by Ricardo for Local Government.

## A4 Appendix 4 – DfT Actions Proposed in the Smart Cities and Transport Workshop Report

After considering the role of Government in Smart Cities, participants of DfT's Smart Cities and Transport workshops (Department for Transport, 2017) were asked to identify what actions DfT should take forward to support uptake of the smart agenda locally in the short term.

Below presents a range of options proposed, with each potential action outlining the problem, solution, and probable impact.

### Action 1: Develop a public repository of Smart information

**Problem:** Participants identified a lack of guidance to support those wanting to deliver Smart transport innovations, and a lack of awareness around what is already happening across the UK.

**Solution:** Participants suggested that DfT should support the development of a public repository of information available on a third party platform. This repository would act as a one stop shop for all those interested in understanding Smart cities and how to deliver Smart transport innovations. This repository could include:

- a, 'Beginner's Guide to Smart cities';
- a map of Smart transport innovations across the UK;
- signposts to existing data or security standards;
- Smart procurement practises and case studies;
- examples of how to increase digital skills, and present opportunities to establish networks.

**Impact:** Participants said this would help to build expertise for all those wanting to deliver Smart transport innovations. It would also help to de-risk uptake of this agenda by, for example, providing case studies and signposting available guidance on data standards.

### Action 2: Build a portfolio of case studies

**Problem:** Participants identified a lack of case studies demonstrating how to deliver and monetise the benefits of Smart transport innovations.

**Solution:** Participants said DfT should commission new research, competition bids or a call for evidence to build-up a portfolio of case studies with clear evaluation. This would showcase the cost benefit analysis of Smart transport innovations, taking into account, wherever possible, cross-sector benefits.

**Impact:** Participants said this would help to de-risk delivery by providing detailed evaluation, and strengthen the argument to better use data and deliver Smart services.

### Action 3: Adjust funding models and appraisal procedures

**Problem:** Participants said existing funding models and appraisal procedures may need to be altered to better support the delivery of Smart transport innovations.

**Solution:** Participants suggested DfT could take the following actions:

- Ensure competition bids are focused around a particular problem;
- Allow bids to be collaborative across regions where it is more likely to deliver better overall outcomes, exploit the diversity of expertise and experience, and reduce nugatory bidding costs;
- Have more than one departmental sponsor to widen the sector focus;

- Increase funding streams and seek to provide funding for longer to support successful projects;
- Simplify bidding processes to support the involvement of SMEs and promote innovation;
- Review WebTAG to enable transport appraisals to reflect the benefits that Smart transport innovations can bring (when compared to more traditional transport interventions which WebTAG is designed for);
- Encourage evaluation of projects providing examples of what good evaluation looks like and how to identify the benefits of data-based services;
- Make evaluation a criteria of bids; and
- Externally publish outputs of projects, making it a requirement of the process.

**Impact:** Participants said a review of existing funding practices should help to de-risk and encourage more to deliver Smart transport innovations whilst establishing a portfolio of case studies from which to learn.

#### **Action 4: Build public awareness**

**Problem:** Participants said there is a need for national promotion of this agenda to raise public awareness.

**Solution:** Participants said DfT should be an active promoter politically championing this agenda, by publically stating its Smart City strategy and encouraging city-regions to do the same. To support this further, DfT should enable more demonstrators, and work across Government to feed into the wider digital agenda.

**Impact:** Participants said this will build public awareness of Smart Cities and encourage delivery of Smart transport innovations.

#### **Action 5: Review existing standards and guidance**

**Problem:** Participants said it is unclear if existing standards and guidance on Smart Cities are fit for purpose and whether they have been advertised sufficiently.

**Solution:** Participants said DfT should review standards and, where required, commission new guidance.

**Impact:** Participants said this will help to de-risk the agenda for local government decision makers.

#### **Action 6: Engage with external stakeholders**

**Problem:** Participants said there is a need for DfT to engage more with city leaders, service providers and industry to further understand and take action to unlock barriers to the delivery of Smart transport innovations.

**Solution:** Participants said DfT should continue to build its network of external stakeholders, reviewing Smart barriers and taking actions where needed, for example developing legislation to encourage open-data.

**Impact:** Participants said this would help to build expertise within the Department and de-risk the agenda externally.

## A5 Appendix 5 – The potential of Big Data to inform new predictive analytics

### A5.1 Turning data into predictive capabilities

Ultimately, appropriate access to Big Data sources is not merely about understanding what its description of the existing conditions of the transport systems or the cities, but more importantly the capabilities to predict future conditions and trends, and design appropriate interventions in a timely manner. Turning data into predictive capabilities is an important mission for data science and technology.

At the core of the predictive capabilities at the city scale is the interdisciplinary science of urban modelling. Urban modelling was first created as a coalition of research communities some 50 years ago which has since made significant contributions to integrated transport and land use planning, accessibility studies, spatial economic analyses and the assessment of the business cases of infrastructure projects. As a new frontier research community urban modelling is now at the very core of the New Science of Cities (Batty, 2013).

In the days when limited data sources were available, the focus of urban modelling tends to be placed on the predictions of the longer term effects of major interventions and cumulative causation. The emergence of Big Data sources with near-real time access has enabled urban modellers to analyse much narrower time horizons which are associated with near-term issues and interventions in cities, matters that pertain to minutes and hours as well as those that pertain to years and decades.

### A5.2 The need for new thinking in Smart City data access and predictive analytics

The different aspects of city-scale Big Data applications involve combination and fusion of diverse data sources. The assembling data sources into larger wholes are expected to bring about very large benefits for monitoring, managing and predicting transport activities in cities. When combined, the data sources can reveal hitherto unknown patterns in our urban lives. This knowledge could benefit society as well as individuals. However, it will also come up against an enormous risk that insights derived from putting the data together may open the possibilities of abuse and breach of privacy. This is because in many cases the data fusion is not anticipated by original data collectors at the time of collection, and the use of such knowledge may not have been anticipated or indicated to people or organisations who are the object of that data. Such concerns are at the root of many, if not the most of the challenges facing practical applications today.

Location and trajectory data is inherently personal and identifiable in nature and difficult to anonymise effectively. This is particularly so when various databases are assembled together. Tracking and co-locating people with other people and places, sometimes even with non-personal data such as actual public transport and taxi departure / arrival times, could expose daily patterns of activity and relationships that lead to serious breaches of privacy and trust.

Data access models as well as data protection policies are lagging well behind in this field. This is especially true for location data, and smart city applications are in most cases related to location data in one way or another.

This challenge looks even harder to tackle when it is demonstrated that existing techniques to de-identify sensitive personal or organisational data may not be effective in protecting their identity and privacy, and at the same time, a de-identified dataset could lead to significantly different analytical conclusions from those that would be drawn when the original, un-treated dataset is used, creating unacceptable biases (Daries et al, 2014). This is because statistical biases can be introduced into a de-identified dataset by inappropriate generalisation or suppression of the data records (Angiuli et al., 2015).

If the uses of a dataset are known, then the recent discoveries of the pitfalls of de-identification processes can be minimized through bespoke designs of de-identification by those who have the access to the original dataset. Nevertheless, in the Smart City context, a high-quality dataset may contain a variety of attributes that are of interest to a wide range of users. In such cases, bespoke de-identification will be problematic, because it will be hard to make assumptions regarding what statistical relationships and variables are to be prioritised e-identification.

While such issues will continue to be high priority research topics in the field, particularly regarding monitoring and management of transport systems, there are opportunities to explore new data use models for predictive analytics. This is because the model parameters that are used in predictive analytics may be estimated in closed data labs – the task can be performed by specially designed procedures that avoid the disclosure of actual data to outside the closed data labs. This is a new field of analytics building that requires careful technical and legal consideration in the near future.



Ricardo  
Energy & Environment

The Gemini Building  
Fermi Avenue  
Harwell  
Didcot  
Oxfordshire  
OX11 0QR  
United Kingdom

t: +44 (0)1235 753000  
e: [enquiry@ricardo.com](mailto:enquiry@ricardo.com)

[ee.ricardo.com](http://ee.ricardo.com)