Personal data in transport: exploring a framework for the future

Open Data Institute

About

This report has been researched and produced by the Open Data Institute, and published in June 2018. Its lead author was Libby Young, with contributions from Gillian Whitworth, Jared Robert Keller, Jamie Fawcett, Miranda Marcus, Jessica Weereratne, Caley Dewhurst, Anna Scott, Peter Wells and Jeni Tennison.

The purpose of the report is to help transport companies make better decisions about how they use personal data in their business models for the future. Our focus is journey data, a form of personal data the sector has always needed, now accessible in new ways and at new scales.

The report coincides with the implementation of the EU's General Data Protection Regulation (GDPR). It is one of the most comprehensive pieces of regulation yet to deal with personal data, and is changing the way businesses across sectors collect and use data.

The GDPR is just one part of a growing trend towards people exercising more control over data about them, so we take a long term view on this as well as other trends – the increasing collection and use of journey data in transport, and the benefits this can create for businesses who foster and earn trust by engaging with us as they use data about us ethically and equitably.

We hope businesses find the research and learning shared here helpful in exploring how to address and lead change around these trends, and as they work to find ways to make travel more accessible, affordable and sustainable for all.

If you would like to send us feedback, please get in touch by email at RandD@theodi.org.

Contents

We don't 'own' data

At the ODI, we believe that data does not belong to any one person, not least because data about us is often also about other people. Our DNA, for example, reveals information about our parents and family. Utility bills reveal who we live with. Health records contain information about medical professionals. This is why we do not refer to 'my or your data' but instead use 'data about me, you, us or them'.

Many people can have rights to – and responsibilities for – data. For instance, many people and organisations may have intellectual property rights due to their contributions to the data. In transport, people often take journeys with family or friends, so ticket or location data can be connected to or used to identify more than one person. In connected vehicles such as cars, drivers and their passengers may be connected with other traffic participants.

By avoiding possessive pronouns we hope to encourage understanding of the wider network of relationships involved in data collection and use. Put simply, neither individuals or organisations should think of themselves as owning data. Instead they should think of themselves as holding or stewarding data with the goals of both protecting the rights of people impacted by the data and ensuring that people who have the right to access and use it can get the data when they need it to make a decision.

Executive summary

Using journey data to meet a universal need

Transport is a universal need: almost everyone travels regularly, be it walking, taking a bus, using a car or going by train. These journeys often generate data that can reveal things about us – where we work, where our loved ones live, what we like to do. This journey data is personal data.

The role of journey data in transport is changing. It is being collected in real time and used on new scales, creating insights into individual and aggregated mobility. Businesses and public services can use this data to make operational and policy decisions to improve transport services, user experience or infrastructure. Innovative new mobility services are also emerging

New ways of operating transport, such as bikesharing and ridesharing, and new ways of accessing transport, like journey planners and smart ticketing, are changing the sector. These on-demand, personalised services can help people save time and money, be more (or less) active, own fewer or no means of transport, lower their carbon footprint, or travel in other ways that meet their individual or societal needs. This also opens up ways to make transport systems more user-centred for more people and for the full length of their desired journeys. In the future, travellers could complete trips door-to-door using a single planner and ticket and access varied public and private transport as their journey requires.

Through our interviews with industry, workshops and research we found that journey data can be used and shared to benefit people, organisations and our transport systems more broadly in three ways:

- to increase accessibility and personalisation for people
- to improve operational efficiency and innovation for organisations
- to tackle systemic transport issues for the benefit of everyone

We also uncovered areas that must be explored to help realise this potential equitably, using data about us ethically, and in ways that support our systems. For instance, just as open data from public transport and other sources supports new mobility services, these new services now hold journey datasets that can support better decision-making about transport infrastructure and operations. Many companies are thinking about ways to make aggregated and de-identified journey data more open for this purpose, and some of them are testing ways to do so. Further exploration is encouraged.

Our research also showed that sharing journey and other transport data more widely requires us to address technical barriers including inconsistent open standards, data silos and poor interoperability of data. Breaking down these barriers will require collaboration and investment across the sector.

Most importantly, we found that journey data is transforming relationships between companies and customers, drastically increasing the importance of trust, and raising critical questions of ethics, equity and engagement which cannot go unanswered. The impact of these trends will only increase as people grow more aware of these issues and as they gain more rights over data about them.

These shifts create both opportunities and challenges for businesses. Based on these and our other findings, we recommend transport organisations:

i) Develop a strategic and ethical approach to data by looking beyond compliance with regulation to develop governance policies and processes, investing in training and resources, and engaging with sector initiatives.

ii) Lay the foundations for and engage in data-enabled innovation by supporting open standards and APIs, practising open innovation, and exploring new ways to partner with public services.

iii) Foster trusted relationships with customers by communicating with them clearly, finding ways to build more informed consent into the user experience and explaining the benefits created by data sharing.

iv) Explore sharing more journey data safely and securely in three ways:
1. As open data – with careful and appropriate aggregation and anonymisation, and accompanied by privacy impact assessments
2. In secure environments – in particular with accredited researchers, who should openly publish their results
3. With people the data is about – through open APIs and in ways that enable innovation.

There is much at stake for us all in the future of data and sharing data in transport. For economies, the Transport Systems Catapult says data-enabled innovations in transport may be worth £900bn by 2025, and conversely, not sharing transport data may cost a country like the UK £15bn[1].

For individuals, sharing journey and other data to increase personalisation of and access to transport can improve our quality of life. Conversely, if personal data is used in ways that erode trust it can create data wastelands (where data is not collected or used), limiting the potential for transport systems to benefit everyone.

A future in transport where data is shared and used in ways that combine openness and innovation with ethics and trust has powerful potential to benefit our societies and economies.

Data is infrastructure, like transport

Infrastructure powers our societies. It provides the fundamental services and systems that enable our economies to function, allow us to communicate, and improve and support our lives.

When we think of infrastructure we first think of roads, bridges, water supplies and electrical grids. But infrastructure also takes less tangible forms, such as ideas, basic research and the internet.

Data is becoming as important as our physical infrastructure – it underpins every sector of our economies and impacts every aspect of society. Roads help us navigate to a place; data helps us navigate to a decision.

Data infrastructure consists of data assets, the organisations that operate and maintain them, and the guides describing how to use and manage the data. Trustworthy data infrastructure is sustainably funded and is directed to maximise data use and value, meeting society's needs.

Data infrastructure that is as open as possible creates the most value. Like other types of infrastructure it is invisible (when it works). Making it as open

---

[1] Transport Systems Catapult, https://ts.catapult.org.uk

as possible, while respecting privacy, helps to support a huge variety of uses and ecosystems.

Governments, businesses and communities plan essential physical infrastructure strategically – including our highways, electricity lines, water courses and broadband connections – and we should treat data infrastructure in the same way.

Why is journey data personal data?

Our journeys from location to location through time can be used to infer where we work and shop, what we earn and what we do in our free time. These are all things about us. Data collected when we make journeys is revealing more about us than ever, such as when and where we travel with friends, family or colleagues[2]. For these reasons, journey data is personal data.

Historically, public and private transport providers collected journey data as 'counts', i.e. the number of people on a train, bus tickets sold or cars using a road. These counts were usually not tied to individuals, except when people took part in customer surveys. The data helped transport organisations make decisions about things like timetables or the location of bus stops.

In the last few years, GPS-enabled devices, smartphones and NFC-enabled smart tickets and payment cards have made journey data easier and cheaper to collect. Parts of the sector have moved on from 'counts' to data collected using these technologies, at much larger scales and in real time, across modes of transport. These richer journey datasets can help decision-makers improve transport planning, personalise services and tackle systemic issues. They have increased innovation by established operators and allowed new mobility services to emerge.

Our increasing use of these technologies to plan journeys and access transport is allowing banks and smartphone makers to collect journey data. Banks can do this when providing payment services for transport. And smartphone makers can collect data each time we enable our location services during a journey, or when we access mobility services on our phones[3]. Telecom providers have historically collected customer location data, generally by triangulating mobile phone signals, but this data is not considered reliable enough to use as journey data supporting decision-making in transport.

All this journey data can be pseudonymised with identifiers like random numbers, allowing journeys by one individual to be linked without using direct identifiers like names. However, even when data is pseudonymised, there is still a risk of re-identification when it is linked to other datasets. One study found that four data points on time and location can identify 95% of people in a dataset[4]. Others have mapped a cyclist's journeys with TfL data[5], or identified trips made by taxi drivers in New York[6].

Consumers and regulators are becoming more engaged with the issues around this and personal data more broadly. How organisations develop their personal

---

[2] This report focuses on transport customers, but we note that employees of transport companies, like bus drivers and station guards, are also affected by most of these issues.
[3] While this report is on transport, we note location data raises issues for many industries
[4] Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen & Vincent D. Blonde (2013), 'Unique in the Crowd', https://www.nature.com/articles/srep01376
[5] James Siddle (2014), 'I Know Where You Were Last Summer: London's public bike data is telling everyone where you've been',
https://vartree.blogspot.com/2014/04/i-know-where-you-were-last-summer.html
[6] Vijay Pandurangan (2014), 'On Taxis and Rainbows: Lessons from NYC's improperly anonymized taxi logs', https://tech.vijayp.ca/of-taxis-and-rainbows-f6bc289679a1

data strategies could increasingly become a differentiator for transport organisations. We will look at some of the opportunities and risks this can create for businesses later in this report.

---

[7] https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/key-definitions

Transport needs data

Access to journey data is transforming the sector

The role of data in transport is changing, as the sector shifts from data-poor to data-rich. In particular, journey data, a form of personal data the sector has always needed but has historically found hard to get, is now accessible in new ways and at new scales. Companies are collecting more of this data, more frequently, and in real time – as technology makes this more feasible and viable – and using it to innovate. Customers expect more personalisation and communication, and are sharing more real time data in order to get it.

Our research and interviews with industry show transport companies are collecting data in a variety of ways. While cash-only taxis and buses collect little or no journey data, public transport operators such as trains and new private operators such as bikesharing and ridesharing companies can connect customer account details with real time data collected via smartphones, smart cards and other devices to gather journey data for single modes of transport. End-to-end mobility services like journey planners and smart ticketing services can collect journey data (and intended journey data, in the case of journey planners) across many journeys and modes of transport, creating multimodal or overarching views of transport in real time. Public transport providers who operate multiple modes of transport, such as Transport for London (TfL) can also create these powerful overviews.

This real time journey data can be used and shared to benefit businesses, people and public services, potentially in ways that meet the needs of all three groups. This highlights that transport is an ecosystem of organisations, individuals and infrastructure depending on each other to achieve their goals. Our interviews, workshops and research called attention to three key areas of potential benefit:

> **Increasing accessibility and personalisation for passengers**
> *– people can benefit from more personalised or affordable services as organisations use journey data to better understand and serve their needs.*

For example, services like Waze can crowdsource and share car journey data to help drivers choose their route based on live traffic patterns. TfL can use journey data in a personalised way, to warn passengers of relevant disruptions affecting their route and at times that allow them to adjust their journeys. Services like Moovit or Digital Matatus can help people use public transport in cities with less formal or discoverable networks (see Digital Matatus case study).

> **Improving operational efficiency and innovation for organisations**
> *– private and public organisations can benefit by using journey data to make better decisions about how to use resources to develop or improve services.*

For example, TfL can use journey data collected through its Oyster cards to improve and plan its transport interchanges in London[8]. Rideshare services can use journey planners – enabled by open data and journey data – to find and connect with customers. Bikeshare services can use journey data to help locate their bikes where customers want them, in turn leading to innovation in complementary services like Stage Intelligence which can help mobility services like bikesharing make these operational decisions[9].

---

[8]https://www.citymetric.com/transport/data-helps-us-provide-better-transport-tfl-oyster-cards-big-data-and-contactless-payments-1396
[9] https://stageintelligence.co.uk

Tackling systemic transport issues for the benefit of everyone
– public services and governments <u>can use data</u> to better meet the travel needs of underserved communities, and address issues like the 'last mile', congestion and pollution[10]. It can also help make previously expensive services more affordable for businesses to provide.

For example, Stratford-upon-Avon startup Enable ID is collaborating with UK train operators to use personal data to explore intelligent accessibility for people with disabilities (see TOC Ability case study). Private companies are working with local authorities to develop smart bus services in rural areas currently underserved by public transport (see Smart Buses in the UK case study). Finnish startup MaaS Global is using multimodal transport and journey data to help tackle the last mile problem in Helsinki and the West Midlands (see Transport in Finland case study).

Defining terms

Transport, mobility, MaaS: explaining the differences

In this report, the term **transport** is used to describe the whole **ecosystem** of moving people or goods from place to place[11]. Transport is made up of physical infrastructure, as well as the organisations and individuals who own or operate the different modes of transport, such as trains, buses, bikes and cars.

On the other hand, **mobility** refers to an individual's journey from location to location through the transport system, either using one mode of transport, such as walking, or multiple modes of transport, such as a car + train + bus trip to a destination (also referred to as **multimodal** travel).

The term mobility is often used in the context of **intelligent mobility**, which is using <u>'emerging technologies to enable the smarter, greener and more efficient movement of people and goods'[12]</u>. More user-focused or newer transport companies often call themselves **mobility services**.

Some tie these concepts together as '**mobility as a service**', or **MaaS**. MaaS is an evolving concept which aims to combine the benefits of multimodal journey planners and smart ticketing to help people more easily and efficiently plan and pay for multimodal journeys. MaaS could change the way people consume public and private transport, enabling more multimodal journeys, reducing the need for private car ownership and improving **end-to-end mobility**.

End-to-end mobility relates to the '**last mile**' problem, one of the systemic transport challenge we reference in this report. The last mile describes the movement of people to a final destination such as a home or place of work (in reverse order, this is also a first mile problem). This challenge highlights how hard it is to deliver physical transport infrastructure that meets the needs of all travellers, each with individual destinations.

---

[10] For examples from transport and other sectors, see: Ed Parkes and Mandy Costello (2018), 'Using open data for public services', https://theodi.org/article/patterns-for-using-open-data-in-the-delivery-of-public-services

[11] This report focuses on the movement of people through transport systems, not goods.

[12] https://ts.catapult.org.uk/intelligent-mobility/introduction

Companies are thinking about sharing data

Our interviews with transport companies showed a keen awareness of these trends in the collection and use of data, as well as an appreciation of the opportunities these trends create to develop services, earn the trust of customers to support data sharing, as well as support transport ecosystems overall.

> " *"We want to be able to share aggregated data to help innovation."*
> – Chris Perry, MaaS Global

Some stakeholders want to put the data they collect to wider use, and early data sharing pilots are generating compelling results. For example, Waze's data partnership with TfL has reduced delays on London's most congested route, the Blackwall Tunnel, by using journey data to suggest drivers at risk of running out of fuel reroute to a nearby petrol station[13].

> " *"Urban mobility is a universal asset."*
> – Ed Parsons, Google

There is also recognition from the sector that it needs to find safe and secure ways to make journey data it collects more open in ways that protect individual privacy and support transport ecosystems. Many businesses think better collaboration between organisations and clearer communication with people can help do this.

> " *"We need to work out how to share in a way that supports the wider system."*
> – Chris Perry, MaaS Global

A growing number of businesses believe services that meet the needs of both individuals and communities can foster and earn the trust of users, especially when they tell users about these wider benefits. They observe that users share journey data more when they know it has a communal benefit, because they value the benefits that the journey data of others has given them.

---

[13] Theo Chapple (2018), 'Waze partnership: Reducing congestion at Blackwall Tunnel', https://blog.tfl.gov.uk/2018/01/11/waze-partnership-reducing-congestion-at-blackwall-tunnel/

> 66
>
> *"Once you enable the GPS to share anonymised data, you are getting better information from us, and you are also contributing this information to other people in your community"*
> – Roy Bick, Moovit

Some mobility services explicitly communicate this message to customers, sometimes as part of the user experience.

> 66
>
> *"People are more open when it is for the greater good and if it is feeding back into their own travel experience."*
> – Max Stewart, Caution Your Blast

Open standards in transport

Open standards are an important part of our local, national and global data infrastructure[14]. They help us to use, share and publish data in consistent and easy-to-access ways, and can help to change markets. Data-enabled innovation in transport depends on open standards. In order to work, most mobility services need to link journey data with real time transport data. Even companies like Waze and Moovit, who crowdsource data, need it.

Transport has long been a leader in open data, supporting the creation of new mobility services around the world. For example, the General Transit Feed Specification (GTFS) is an open standard for public transport schedules originally started by Google and Portland, Oregon's transit agency, TriMet[15]. The Service Interface for Real Time Information (SIRI) is an open protocol to exchange real time information about public transport services and vehicles[16] (GTFS also has a real time extension). More recently, the General Bikeshare Feed Specification (GBFS) developed by the North American Bike Share Association shares real time data for docked bikeshare services[17].

---

[14] https://standards.theodi.org
[15] Bibiana McHugh (2013), 'Pioneering Open Data Standards: The GTFS Story', http://beyondtransparency.org/chapters/part-2/pioneering-open-data-standards-the-gtfs-story
[16] https://streamdata.io/blog/understanding-service-interface-real-time-information-siri-standard
[17] Adam Russell (2016), 'New bikeshare data standard opens doors for apps and analysis', https://mobilitylab.org/2016/05/04/new-bikeshare-data-standard-gbfs

[SharedStreets](#) is an open data standard for describing streets, and has been developed specifically to support transport systems[18].

Despite all this, open standards in transport are not yet universal. The UK is a case in point. For timetables, [TransXChange](#) is not used by National Rail, GTFS is not used by TfL[19]. For stations, the UK has no open address register to support [NaPTAN](#)[20]. Fare data is very fragmented (for instance, the Rail Delivery Group estimates there are currently [55 million different train fares](#)[21]). For ticketing, some find the [ITSO](#) open standard hard to implement[22].

Friction in interoperability can increase costs for data users and limit innovation, for instance by making it harder for new entrants to develop complementary services to support existing transport. In particular, poor interoperability limits innovation around MaaS.

Consumers and policy makers are also thinking about sharing data

Companies' attitudes to sharing in transport align with the findings of a recent [ODI survey](#) of UK consumer attitudes to sharing personal data[23]. The survey of over 2000 consumers found that if a consumer trusts and knows an organisation, they are more likely to share personal data about them with that organisation. 94% of respondents said trust was important in deciding whether to share personal data. 64% said they would share personal data with an organisation they know, compared to just 36% for an organisation they do not.

The survey also showed consumers are prepared to make trade-offs to share data about them if it benefits them and society. While 15% said they would share data to get insights about themselves, even more (26%) would share personal data to help decide which public services to fund. 37% would share some data to help academic research, while the most popular 'data trade-off' in the survey was similarly for the greater good: nearly 50% said they would share personal data if it helped develop new medicines and treatments. We use 'greater good' in this report to refer to these types of trade offs, when an individual puts a community's needs ahead of their own, or sees the two as connected, even inseparable.

37% of people would share data about them
to advance academic research

Recent policy [conversations](#) about transport in the UK have focused on how to share data more. The 2017 [Bus Services Act](#) requires private bus operators to make fare, timetable and route data open, to give people access to more information. The Transport Systems Catapult, working with the ODI and Deloitte, [identified](#) ways for government to incentivise data sharing to support new mobility services[24]. And a government [inquiry](#) is exploring ways to support MaaS.

---

[18] https://www.sharedstreets.io

[19] http://naptan.dft.gov.uk/transxchange

[20] https://data.gov.uk/dataset/ff93ffc1-6656-47d8-9155-85ea0b8f2251/national-public-transport-access-nodes-naptan

[21] https://www.raildeliverygroup.com/media-centre/press-releases/2018/469773920-2018-05-08.html

[22] https://www.itso.org.uk

[23] ODI (2018), 'ODI survey reveals British consumer attitudes to sharing personal data', https://theodi.org/article/odi-survey-reveals-british-consumer-attitudes-to-sharing-personal-data/

[24] See 'The Case for Government Involvement to Incentivise Data Sharing in the UK Intelligent Mobility Sector' briefing paper on https://ts.catapult.org.uk

Case studies for the future

Transport in Finland

Connecting personal data with open data infrastructure for mobility-as-a-service

Finland's groundbreaking Act on Transport Services, effective July 2018, requires every transport group, public or private, to share an open API[25]. This legislation aims to support a shift to transport as 'mobility-as-a-service', meeting the travel needs of all Finnish citizens through single ticket, end-to-end, multimodal journeys.

The Act builds on the success of Helsinki Region Infoshare (HRI), launched in 2011[26]. HRI shares hundreds of open datasets to support academic research, business innovation and public engagement. It also uses meet-ups, hackathons and social media to drive these efforts. HRI has created many benefits. For instance, the Helsinki Regional Transport Authority journey planner open API has enabled many mobility services like:

> WHIM – MaaS Global's application to help people in Helsinki plan and complete journeys easily and affordably without owning a car. The service is also in beta in Birmingham in the UK in a collaboration with Transport for West Midlands, National Express and Gett[27].

> TrafficSense – Aalto University's research project collecting data on users' regular journeys and linking it with real time and multimodal data on speeds, congestion and other data in order to give users more personalised suggestions when planning travel routes[28].

> Blindsquare – a GPS application helping visually impaired people travel independently. The voice-enabled app collects a user's location and other personal data in order to personalise and share data such as street intersections, cafés and favourite locations[29].

Alongside these initiatives, the government also supports MyData, an alliance exploring ways to give people more control over data about them while helping businesses use this data to develop services built on trust[30].

This emerging national framework for transport data infrastructure – integrating public and private services, supporting equity and ethics – could suggest a blueprint for other countries, especially those exploring MaaS.

---

[25] Ministry of Transport and Communications (2017), 'Good and flexible transport services through a new act',
https://www.lvm.fi/en/-/good-and-flexible-transport-services-through-a-new-act-933165
[26] Helsinki Region Infoshare (2018), 'What is HRI?',
https://hri.fi/en_gb/hri-service/what-is-hri
[27] https://maas.global
[28] Helsinki Region Infoshare (2017), 'TrafficSense',
https://hri.fi/data/en_GB/showcase/trafficsense
[29] http://www.blindsquare.com
[30] https://mydata.org/finland

Fostering trust in transport

Trust is a high stakes issue when using personal data

Trust is a high stakes issue for transport companies who rely on their ability to collect and use data about people, in particular journey data. While companies in many sectors can collect location data during a person's journey – from banks and smartphone makers to telecom providers and social media – it is transport companies whose relationships with customers are most defined by journey data, and whose business models are most closely aligned with it. These stakes will likely grow higher as journey data is collected and used more, as more people think about how and why they share data, and as we gain more rights over data about us.

In any sector using personal data, trust is strongest when people think, and feel, that an organisation uses data about them in ways that meet both their needs for a service and their expectations in relation to privacy and receiving a fair share of the benefits for data about them. People become more willing to share data about them and less likely to move to competitors. But when this trust goes, for instance because data is used in a misleading way, we risk a backlash taking us into a 'data wasteland'. In such cases, people may restrict their privacy settings, adding bias into what might have otherwise been useful datasets. They may also withdraw consent by changing service provider, by ceasing to use all providers of a service, or stopping sharing personal data altogether, even for the greater good. For transport – a universal asset which can benefit greatly from people's willingness to share journey data – such a loss of trust could be detrimental to both transport providers and passengers.

Current levels of trust in how location data is being used across a range of sectors should give us all pause for thought. One global consumer [study](#) by HERE showed only one in five people feel they have full control over data about their location, while 44% have shared it unintentionally and 76% feel stressed or vulnerable doing so[31]. On the other hand, using journey data about a person to provide a service for them at the right time and in the right place can create trust. For example, interviewees reported that transport services who tell people about disruptions affecting their journey can help them adjust routes in real time and build trust in the service as a result.

> " *"(If I) manage that disruption… that makes you view me as a trusted partner."*
> – Stuart Walker, Arriva

Our interviews with industry showed an interesting range of views about the role of trust. Those who think of it as an abundant value to foster and grow take a 'show and tell' approach, aiming to communicate and demonstrate to users the individual and communal benefits of sharing journey data. Others see trust more in terms of scarcity, as a fixed asset to preserve, and commit to not sharing any personal data (potentially without asking if and how users might be willing to share it). A third group, increasingly a minority, and now forced to change by tighter data

---

[31] 'Privacy and Location Data' survey, https://www.here.com/en/company/newsroom/press-releases/2018-05-03

regulation, do not proactively communicate about data at all, but may still share or sell it, according to their terms and conditions.

How tighter data protection regulation impacts transport businesses

As organisations have adopted new technologies to collect and use data in new ways and at new scales, regulators have increasingly focused on how personal data is collected, stored and used, and how people can have more control over it. The most recent culmination of this trend is the GDPR, a European Union regulation. It applies to any company processing data about EU citizens, even companies based outside the EU. The GDPR creates some new rights around data, but mostly builds on existing rights in existing data protection regulations.[32]

Right	Description	Did it exist already?	What are its key changes and elements?
The right to be informed	Organisations must give people 'fair processing information' on the use of data about them, primarily through privacy notices	Yes	The information must be concise, transparent and written in clear and plain language
The right of access	Gives people access to data an organisation holds about them	Yes	Information must be available free of charge for the first copy, unless the request is 'manifestly unfounded', and provided without undue delay, generally within one month of its receipt
The right to data portability	Lets a person obtain a copy of data about them to share with other services.	No	This is a new right under the GDPR. It applies in specific situations where data was processed automatically, and the justification for processing it was that the individual had consented to that use, such as loyalty card data.

Under the GDPR, companies must also apply the principles of data protection by design and by default, minimising the data they collect, ensuring good governance of personal data processing and demonstrating compliance. In situations deemed high-risk, they must also carry out Data Protection Impact Assessments (DPIAs). **'It is all about strengthening trust and updating the law to the 21st century.' (Garreth Cameron, The Information Commissioner's Office (ICO))**

Fines for non-compliance under the GDPR could be substantial, with businesses potentially liable for up to 4% of annual global turnover, or €20 million (whichever is higher). But fines are not the focus of the regulation: **'Thinking that GDPR is about crippling financial punishment misses the point... it's about putting the consumer and citizen first.' (Elizabeth Denham, UK Information Commissioner).**

Most transport companies welcome the GDPR, as guidelines and benchmarks that are necessary to manage long-term trends in the collection and use of personal data. **'There have to be actual laws that providers are accountable to in order for transport to remain sustainable', says Eanna Lalor of Urbo.**

Many also see it as an opportunity.

---

[32] https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr

> **"GDPR helps you think about the customer experience in more detail."**
> – Stuart Walker, Arriva

The GDPR's introduction of the right to data portability may also increase innovation and competition, as people can use their right to 'port' data about them to support the development of new services striving to meet their needs[33]. For instance, a transport startup working to fill a gap in a specific location could prompt people in that location to share historic journey data about them by changing permission settings in a journey planner service they use.

Some businesses may try to minimise the risk of liabilities created by the GDPR by collecting and using personal data less. In a sector increasingly reliant on real time data from networks and users, this could reduce customer insight, create skewed datasets and erode the quality of a company's services over time. Trying to minimise a legal risk could create a greater, more strategic business risk.

Beyond regulation – ethics, equity, engagement

The more an organisation needs personal data, the more it needs to change how it manages risk. The complex issues raised by the commercial use of personal data need to be assessed and managed proactively and consistently across a company's business model, user experience and communication. Risks which might have been considered reputational, ancillary, or both in the past become strategic and operational, changing how a company makes decisions. To manage this, organisations need decision-making frameworks that support innovation and are guided by three core principles of ethics, equity and engagement.

1. Organisations and people must use data ethically
To earn trust it is vital that organisations and people do not collect, use or share data about people in unfair, discriminatory or misleading ways.

2. Organisations must support equitable access to, and benefits from, data
This promotes fair competition, informed markets, and empowers people as citizens, creators and consumers.

3. Organisations must engage with people to make data work for all
To help meet the needs of all communities, organisations must communicate clearly and work collaboratively with them.

1. Using data ethically

Trust and ethics are inseparable – to earn trust, organisations must act ethically. A decision can be lawful but still be unethical. Ethics is knowing the difference between what you have a right to do and what is right to do (a line sometimes attributed to US Supreme Court Justice Potter Stewart). The ethical issues raised by the use of personal data are complex, and as a result some businesses are exploring the rapidly emerging discipline of data ethics. These businesses see the GDPR as a stepping stone to this.

---

[33] Jack Hardinges and Gillian Whitworth (2018), 'Will GDPR and data portability support innovation?', https://theodi.org/article/will-gdpr-and-data-portability-support-innovation

> **"**
> *"We really see GDPR as being the baseline and very much the minimum that has to be done… We want to take it a step further not just with ticking the boxes but by using our data in an ethical manner."*
> – Eanna Lalor, Urbo

In transport, the potential ethical issues raised by the use of personal data are many and highly complex. In particular, more discussion is urgently needed about what users expect from new mobility services, and the appropriate ways to use and share data about them.

For instance, one study of rideshare services analysed cancellation rates and wait times to suggest some drivers were discriminating against passengers based on the perceived race or gender of the user profiles provided by rideshare services[34]. The study also noted some female passengers were taken on longer trips. Should a driver using their own car to offer trips via a rideshare service be given information about race and gender before being given the option to cancel a trip? And if drivers' cancellations of a certain group make customer demand from that group appear lower than it is, would this bias future innovation using rideshare services' journey datasets? For instance, would it encourage investment in services for groups less discriminated against, because their customer demand appears proportionally greater than it is? (This is as much a question of equity as ethics, a reminder of how closely the two principles can relate).

There are also urgent ethical questions around the potential for data to be used in misleading ways. Commentators like the Data Society have asked how the use of crowdsourced traffic data can impact drivers in potentially misleading ways, for instance if services recommend under-tested routes to gain more data about road conditions[35]. This issue was taken to tragic extremes when a person's death in Niteroi, Brazil was attributed to misdirection caused by street names[36]. How can services accessed while people are focused on activities such as driving ensure they are not misleading? Even if a driver is not being misled, is it right to be given the option to use an under-tested route in exchange for a free service?

2. Supporting equitable access to and benefits from data

Transport is a global need and a universal asset, so it is critical that data is used in ways that meet the transport needs of everyone.

Some groups of people and modes of transport are now intensively monitored and measured, and data about them can be used to strengthen the case for investment in infrastructure, subsidies or innovation. On the other hand, because of the ways journey data is collected, there are groups we have little or no data about who are increasingly at risk of being excluded from such decision-making processes. These

---

[34] Yanbo Ge, Christopher R. Knittel, Don MacKenzie, Stephen Zoepf (2016), 'Racial and Gender Discrimination in Transportation Network Companies', http://www.nber.org/papers/w22776
[35] Alex Rosenblat and Tim Hwang (2016), 'The Wisdom of the Captured', https://datasociety.net/output/the-wisdom-of-the-captured
[36] British Academy (2018), 'Data management and use: case studies of technologies and governance' report, https://www.britac.ac.uk/sites/default/files/Data%20Governance%20-%20Case%20studies.pdf

'data wastelands' can in turn mean businesses miss opportunities to develop new services or enter new markets.

Groups of people most at risk of being overlooked include those who:

- cannot access transport (for example wheelchair users, or people in rural areas without transport)
- pay in cash for transport (including those who use cash-only taxis or buses)
- do not use monitored transport (for example walkers and runners who don't use fitness trackers and are outside connected city centres)
- are digitally and/or financially excluded (for example the very young or old)
- choose not to share data about themselves

This raises important questions about how data can harm, as well as help. For instance, how will a lack of data on walkers and runners relative to other more resource-intensive modes of transport impact the development of our transport and cities over time? Will transport be skewed to deliver more resource-intensive modes of transport because there is more data about users of those services? Will it be harder to make a case for low carbon services and support active mobility?

Potholes offer another opportunity to think about the problem of data wastelands. New mobility services for cyclists such as See.sense lights and Blubel navigators collect data on potholes from users and share it with local authorities, either for free or paid for as a service[37]. Will this make local authorities prioritise the needs of those cyclists who can afford this equipment and share pothole data over other people who use roads? Or will it help local authorities redirect the time and resources saved through this new service into collecting other data to help make better decisions about pothole repairs overall?

The example of people who choose not to share data about them raises policy issues around personal choice which need to be explored. As new mobility services integrate more with public transport, should there be options for people who do not want data about them to be stored or used? (Some smart ticketing services offer unregistered or anonymous versions which offer fewer benefits than cards which are personalised to one user). It could even be a question of free will. Can you still use transport without choosing to share data at all? If the answer is no, it is not a genuine choice, so simply asking for consent could be considered misleading[38].

Societal debates and democratic decisions help determine the choices that citizens and consumers can make or, to put it another way, the sets of individual user needs that will be met. If societies determine that they want consumers to have access to a range of transport services that can be used without sharing personal data but market forces do not lead to the emergence of these services then it may take government intervention to make it happen.

---

[37] https://seesense.cc and https://blubel.co

[38] https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/consent

Rural mobility: an overlooked opportunity?

In rural communities, a person's independence is closely linked to their access to transport, but to date new mobility services have done little to meet these critical travel needs[39]. This lack of rural mobility services relative to cities is effectively creating journey data wastelands in many areas. Left unchanged, this risks limiting innovation in rural transport even further.

To date, new mobility services have focused most on cities, which offer more of the infrastructure these services need to work, and are big, growing markets. This makes sense for a startup, but once a service is established it is important it extends outside urban areas where possible. This supports equity in transport and helps businesses access markets to meet the needs of the growing number of commuters who cross both rural and urban areas.

Journey data could play a critical role in improving rural transport. Historically, one of the obstacles to developing rural transport was the cost of running it – operations in less densely populated areas need to cover more miles for fewer people. Journey data can help organisations understand demand more, making services more feasible and viable.

Some organisations are exploring opportunities like this. For instance, in the UK, Arriva operates smart buses in Kent. Ride's smart bus service in Essex and Suffolk is a FutureGov partnership with local authorities (see Smart Buses in the UK case study). Community led groups like Western Dales Buses also play a valuable role in meeting the needs of some local communities, but most do not collect journey data.

Public sector initiatives may also help reduce rural data wastelands. For example, a government investment to introduce smart ticketing across England and Wales will give train operators the ability to collect more journey data[40]. This could be used to support better investment and planning in rail infrastructure and services. With appropriate anonymisation and consent, it could be linked with other data to help understand the overarching travel needs of people in rural areas beyond trains alone.

While some journey planners only cover cities, Google Maps can collect journey data in rural areas, and offers suggested routes across large distances, including some public transport. Could such rural journey data be aggregated to develop overarching multimodal views of rural travel needs?

MaaS services could also play a valuable role in meeting rural travel needs. For example, Navigogo is working with 16-25 year olds to design and test a new service in Dundee and North East Fife, Scotland[41]. This collaborative initiative aims to give young people personalised travel data across multiple modes of transport, such as walking, buses, trains, taxis, car clubs and bike schemes, and offer concessions and discounts in a single payment process.

---

[39] See Transport Systems Catapult research on rural transport at https://ts.catapult.org.uk/readyforinnovation

[40] https://www.gov.uk/government/news/government-plans-80-million-smart-ticketing-rail-revolution

[41] https://www.the-espgroup.com/project/navigogo

3. Engaging to make data work for all

Organisations who encourage people to share data about them must earn trust by making good on their word. They should do this by ensuring their commercial and communications strategies for how they collect and use personal data match up. Companies who let gaps grow between the two, or who are opaque about the benefits of personal data to the business itself, can create strategic risks around trust that business models using personal data cannot afford.

People currently feel poorly informed about how personal data is being used. A DotEveryone survey of people in the UK found that 70% recognise that a service they share personal data with can use it to target advertising, but only 56% know that the data can be sold to other companies. And while 94% of people said it is important to be able to choose how much data they share with companies and know how that data is used, half of them do not know how to do this[42].

For location data specifically, the findings are even more troubling. The global survey by HERE mentioned earlier in this report found that 80% of consumers are not clear about why location data about them is needed most of the time, what it is used for, where it is stored, and which third parties have access to it.

To address these issues and risks, businesses are encouraged to think more holistically about engagement, exploring ways to explain their business models, how data about people is collected, used and stored, and the benefits and trade-offs involved in clear and relevant ways. People can understand the fundamentals of how a system works when given the relevant information (systems thinker Donella Meadows used the example of a bathtub to show this[43]). This sort of communication can help users make informed choices, either as individuals, or with the support of third parties like data representatives[44].

Figure 2. A seat on the Central Line, Transport for London

---

[42] See http://attitudes.doteveryone.org.uk and http://understanding.doteveryone.org.uk
[43] http://donellameadows.org/wp-content/userfiles/bathtubs101.pdf
[44] Tom Steinberg (2018), 'Personal Data Representatives: An Idea', https://medium.com/@TomSteinberg/personal-data-representatives-an-idea-d2274d1af69b

Case studies for the future

TOC Ability

Using personal data to improve equality in transport in the UK

The spending power of households with people with disabilities in the UK is £250 billion, but businesses often miss opportunities to deliver services for this group[45]. 9% of adults have difficulty with mobility, but represent only 5% of train users, put off by poor accessibility at stations and during journeys[46].

To address this market failure, Stratford-upon-Avon intelligent mobility startup Enable ID is piloting TOC Ability[47]. The startup is using a collaborative and iterative process to design and test the service. It involves a panel of people with disabilities, Transport for London, Arriva, Loughborough University (providing customer experience expertise), the University of Surrey's cyber security experts, and other groups.

Tests will run at sites including Wembley Central and Harrow & Wealdstone (London Underground), Willesden Junction and Carpenders Park (London Overground), and other stations outside London.

In particular, the pilot will test the feasibility of a data hub which delivers better rail services to passengers with accessibility needs by sharing real time data about them with Train Operating Companies (TOCs), journey planners, bus and taxi firms, and even non-transport services like food retailers and chemists.

To provide these services viably, the hub will help TOCs track and balance the profitability of providing these services alongside the related costs of running them. To support trust in the service, the hub will include a consent management system to give users control over data about them, including how it is used and shared.

How could an approach such as the one being used by TOC Ability be applied to test ways to increase accessibility and improve user experience for other groups with specific travel needs?

---

[45] BBC News (2017), 'New voice for disabled consumers',
http://www.bbc.co.uk/news/business-39028509
[46] The Guardian (2016), 'Disabled people are still being treated as second class on public transport',
https://www.theguardian.com/public-leaders-network/2016/sep/27/disabled-people-second-class-citizens-public-transport
[47] https://enableid.com/tocability

How transport uses data

This section looks at some major land transport for people, outlining how they collect and use data, particularly journey data. We do not aim to give a comprehensive global view covering all types of transport – instead we hope to give an overview of some key stakeholders in interconnected transport systems.

We have been struck by two observations while doing this. Firstly, categories in this innovative sector are not clear cut. For this report, we look at modes of transport (often thought of as operations) and new ways of accessing transport (often called services), but the two constantly overlap. For example, buses are a mode of transport that may also use digital services to give people new ways to access their operations. A bikeshare service is a new way of accessing transport that runs physical operations to enable that new form of access. A journey planner is a new service for sharing data about multiple modes of transport that can also use that data to develop physical operations in different modes of transport.

Our second observation is that new mobility services tend to collect and use journey data far more than established public transport operations. If this trend continues it could create imbalances in our transport systems, perhaps even oligopolies or monopolies, in the form of journey data for specific locations. There are private companies and public services already exploring ways, often collaboratively, to address these imbalances and make data more open, and these efforts must continue to ensure transport systems serve the needs of everyone.

Trains

Historically, intercity rail operators collected little personal data, focusing instead on the technologies and data required to validate payments and prevent ticket fraud. However, as new mobility services and other sectors show the benefits of personal data, train operators are exploring ways to collect and use it.

In a growing numbers of places, such as Europe, North America and India, train operators are collecting personal data by asking people to create accounts before buying tickets online; third party retailers such as Trainline and GoEuro can do the same. An increasing number of train operators are also moving to mobile tickets and smart ticketing. Many train operators can also collect personal data through wifi logins offered onboard (wifi providers can collect it too). All this data can be used to improve core operations, customer service and complementary services.

Many train operators still don't know when specific people board specific trains, because cash payments and paper tickets transferable between people or trains remain common. This is changing as countries like Finland, Holland, Germany, France and the UK use more mobile tickets and smart tickets. Ticket barriers or onboard staff can scan these tickets to collect real time journey data linked to individuals. Even in these situations, it can still be hard to complete journey data, as tickets are not always scanned at the end of journeys, and passengers may get off at a station not reflected in their original ticket purchase.

The UK government's investment in rail smart ticketing across England and Wales mentioned earlier in this report will give train operators the ability to collect more accurate journey data and may encourage them to think more about how they use it and other personal data. Operators could find ways to increase engagement with customers and address service gaps, for instance by combining appropriately anonymised journey data with other data to understand overcrowding beyond simple counts. Done ethically, this could help reduce overcrowding and pressure on infrastructure by helping train operators understand what trade-offs individuals might make to adjust when they travel to help the system and themselves.

Buses

Buses, like trains, have historically collected little data and are working to change this. Bus services and their oversight are generally more fragmented than trains, particularly outside major cities. This fragmentation can make it harder for individual organisations to collect data at the scales needed to deliver benefits, and reduce investment in data as a result. Access to more open data infrastructure, perhaps including safely and securely shared journey data, could help address this issue.

The UK's highly fragmented and shrinking bus network is a case in point[48]. There is currently no consistent way for private sector operators of buses to share timetables and fare data with public sector operators of bus lanes and bus stops. This makes it hard for people to get the data they need, potentially making them less likely to trust and use buses. Bus operators' smart cards, which can collect customer data, are still relatively new and equally fragmented, and it is unclear how much they are used. Even when they are used, accurate journey data can still be difficult to collect, as passengers may only scan or tap a smart ticket when they board a bus, not when they get off. As a result of all this, there is little innovation using data, limiting the public and private sectors' ability to understand and serve people's needs.

More open transport data and engagement with local communities is not a magic solution but it could help overcome these obstacles and there are signs of positive change in these directions – the Bus Services Act requires UK bus operators to provide open data by 2020, and some companies are now testing 'smart buses'.

Case studies for the future

Smart buses in the UK

Using personal data to develop bus services for underserved locations

Formalised carpool-type services like Lyft Shuttle, Uber Pool and BlaBlaCar show people will combine sharing data about them with businesses and sharing small private spaces with drivers and passengers they don't know if it helps them make necessary journeys like work commutes or getting home late at night. Now established operators, new entrants and local councils are developing similar services for bigger groups in cities like Manchester, London and Bristol, and in rural areas like Kent, Suffolk and Essex.

Often called smart buses or shared transport, these minibuses and minivans offer routes with pre-determined or on-demand pick-up points and destinations. In particular, these flexible, adaptable vehicles could help meet rural travel needs, improving equity in transport.

The UK's first smart bus pilot was launched by public transport operator Arriva in Sittingbourne, Kent. ArrivaClick collects users' journey plans via a smartphone application and then maps a 'corner-to-corner' minibus route guaranteeing seats for people, including wheelchair users. The year-long pilot reported promising results: 12% of local residents downloaded the

---

[48] BBC News (2018), 'Britain's bus coverage hits 28-year low', http://www.bbc.co.uk/news/uk-england-42749973

application, and 52% of users switched from private car transport as a result. [49]

In Essex and Suffolk, Ride's on-demand bus service was started by FutureGov in partnership with Essex and Suffolk County Councils and the support of Transport Systems Catapult. In urban areas, Paris public transport provider RATP's international arm, RATP Dev, developed Slide Bristol to deliver shared ride-to-work services in Bristol. San Francisco mobility startup Chariot, bought by Ford in 2016, is running a TfL-approved pilot with fixed pick-up points in parts of south London[50].

Citymapper is an example of a journey planner using personal data to expand into transport operations. It began as a London journey planner using TfL open data, and Google and Apple maps and platforms. It combined this data with users' journey data to build an algorithm called 'Simcity'. Simcity analysed demand for routes, mapped it over existing public transport routes to identify gaps, and forecast how new routes might change how people travelled. A smart bus pilot tested these insights and collected customer feedback. A late-night Smart Ride service launched as a result, and Citymapper shares open data on it[51]. Now TfL, Google and Apple display Citymapper's service in their applications, as well as vice versa.

Some people argue that bus networks must remain fundamentally public to ensure infrastructure is built for all, including people less able to pay or persuade. Others say that the private sector should play a role developing business models to fill public service gaps.

Cars

Since the 1990s, most cars have automatically collected location, speed, emissions and other data via event data recorders and diagnostics systems installed directly in cars. This data could usually only be retrieved via physical access to the car. It was used by car makers to measure and improve safety and performance in car models, and by mechanics to find and fix issues in specific cars. Car owners can now also connect devices such as smartphones or telematics 'black boxes' to their cars. These devices can collect and share journey data and other data with third parties who use it to provide services such as navigation, insurance and entertainment.

As a result, cars collect large amounts of data – 25Gb of data an hour, likely to grow to 4 terabytes with self-driving cars[52]. This includes a lot of personal data linked to many people and interested parties.[53] (Connected cars help demonstrate why no one owns data[54]). In future, autonomous cars could even collect personal data that is emotional and physical, like heart rate, facial expression and voice analysis. As cars become more autonomous, people will be able to rent them as they need them for specific journeys between places, improving end-to-end mobility. This

---

[49] See https://www.arrivabus.co.uk/arrivaclick and https://www.intelligenttransport.com/transport-news/65899/arrivas-on-demand-public-transport-service-a-success

[50] See https://www.bookyourride.co.uk, http://www.slidebristol.com and https://www.chariot.com/cities/london

[51] Citymapper (2018), 'Good Bus - What happens when an app company runs a bus', https://medium.com/citymapper/good-bus-part-1-3-77d65e6f8ce3

[52] MIT Alumni (2016), 'Cars, Data, and Internet of Things', https://alum.mit.edu/slice/cars-data-and-internet-things and Networked World (2016), 'Just one autonomous car will use 4,000 GB of data/day', https://www.networkworld.com/article/3147892/internet/one-autonomous-car-will-use-4000-gb-of-dataday.html

[53] For more detail, see the Future of Privacy Forum's guide 'Personal Data in Your Car', https://fpf.org/wp-content/uploads/2017/01/consumerguide.pdf

[54] UC Hastings (2018), 'No One Owns Data', https://ssrn.com/abstract=3123957

rental model will mean cars are more likely to be owned by governments or businesses than people. It will become more important that people know how data about them is collected and used, and that they trust the organisations that do this.

Making data from cars as open as possible could help solve many shared problems. Perhaps most compellingly, it could help reduce road deaths. 94% of car crashes are due to human error[55]. Data from cars in accidents could be anonymised and shared to help researchers and public authorities analyse and, where possible, fix factors causing them. This data could also be linked with other open data on road accidents, such as UK road safety data or New York motor vehicle collision data[56]. It could also help car makers develop safer software for automated cars more quickly.

Accidents could also be reduced by sharing data including journey data between cars or with transport infrastructure. The EU's C-Roads project and the Nevada Center for Advanced Mobility's partnership with Nexar are testing ways to create these vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) networks[57]. They could also use this data to reduce traffic and improve real time road management, as well as help make roads more suitable for self-driving cars.

When an accident occurs, telematics insurers can already use a customer's location data to get help there quickly. Telematics insurers collect data from boxes installed in customers' cars or via their smartphones and use the data to personalise insurance premiums. This can make insurance more affordable for drivers in higher risk groups, such as young drivers, if they are better-than-average. Businesses can also use this data to identify insurance fraud. Many telematics algorithms are not transparent in how they track and measure driving patterns, and should explain to customers how data about them is being used and the benefits it can create. More clarity is also needed on how the data they collect can be shared[58].

There are many other ways cars sharing data could help solve problems. Flourish is exploring ways both connected and autonomous vehicles might help address the mobility needs of older people in the UK[59]. Connected cars could share anonymised location data to help maintain open mapping data which benefits all drivers, or help make open address registers more accurate by sharing missing data with them.

Given the amount of personal data cars can collect, and the number of organisations who can use it, it is critical privacy is considered and protected at all times. Major car makers like Volvo, Ford, Nissan and BMW have signed up to the Privacy Principles for Vehicle Technologies and Services[60]. One name not mentioned is Tesla, who some commentators have criticised for how it shares data about people[61]. As cars use more data themselves, car hacking will also become a growing risk and increase organisations' focus on the security of all data[62].

---

[55] https://crashstats.nhtsa.dot.gov

[56] See https://data.gov.uk and https://opendata.cityofnewyork.us

[57] See https://www.c-roads.eu/platform and https://www.getnexar.com/nevada

[58] The Telegraph (2014), 'Insurers admit 'black box' data may be handed to police', https://www.telegraph.co.uk/finance/personalfinance/insurance/motorinsurance/11217690/Insurers-admit-black-box-data-may-be-handed-to-police.html

[59] Shergold, Wilson & Parkhurst, G. (2016), 'The mobility of older people, and the future role of Connected Autonomous Vehicles', http://www.flourishmobility.com/publications

[60] https://autoalliance.org/connected-cars/automotive-privacy

[61] The Guardian (2017), 'The customer is always wrong: Tesla lets out self-driving car data – when it suits', https://www.theguardian.com/technology/2017/apr/03/the-customer-is-always-wrong-tesla-lets-out-self-driving-car-data-when-it-suits

[62] The Register (2018), 'Newsflash: Car cyber-security still sucks', https://www.theregister.co.uk/2018/01/26/car_hacking_wireless

Ridesharing

While most traditional taxis remain data-poor, ridesharing services are built on data. Companies like Uber and Lyft collect journey data, user ratings, contact and bank details for both passengers and drivers, to connect them on-demand and manage payments between them. These mobility services can also link this data with other data to understand customers more and develop adjacent services, such as food deliveries[63]. They may also sell this data to third parties.

These popular services have met travel needs and created jobs in cities around the world. They have helped people meet their last mile travel needs, and made things easier for people who can't access a car or public transport, or whose travel needs cannot be met by public transport. All this benefits people who can afford to pay for such services, but it does not assist people who cannot. It is critical that the impact of these new services does not reduce equal access to transport over time.

These fast-growing mobility services have also highlighted the need to make decisions about data which foster and earn trust at all times. Some commentators are concerned about how rideshare services store the large amounts of personal data they collect, and Uber is working to rebuild trust after concealing data hacks last year[64]. Other commentators are concerned about the demands these services put on transport infrastructure[65], and debate their net effect on economies overall[66].

As a result, there is growing pressure for rideshare businesses to engage with us more, and make their journey data more open – San Francisco even compelled it[67]. Now, Uber Movement shares aggregated, anonymised data for Paris, Sydney and other cities to help policy makers understand and meet urban mobility needs[68]. Several ridesharing services have also signed Zipcar co-founder Robin Chase's Shared Mobility Principles for Livable Cities, supporting collaboration and open data to develop transport in the interests of all[69]. Concepts such as data observatories may be one way to start turning these commitments into action and provide insights that help local areas make more informed decisions[70].

If rideshare services continue to grow, the imperative to share insights such as usage data will grow with it. The speed of adoption of these new services, combined with ongoing poor data collection in some public transport, could create journey data monopolies, for instance in cities where fragmented bus services not collecting data compete with a single ridesharing service. This asymmetry between public and private operators could limit the ability of public services to invest and innovate over time, and be exacerbated as public transport data continues to become more open. For example, the Bus Services Act in the UK mandates smart buses with a bus operating license to provide open data by 2020, whereas rideshare services which can effectively compete with them but operate under a private hire license are not required to do the same.

---

[63] https://www.ubereats.com

[64] Bloomberg (2017), 'Uber Paid Hackers to Delete Stolen Data on 57 Million People', https://www.bloomberg.com/news/articles/2017-11-21/uber-concealed-cyberattack-that-exposed-57-million-people-s-data

[65] CItyLab (2017), 'The Ride-Hailing Effect: More Cars, More Trips, More Miles', https://www.citylab.com/transportation/2017/10/the-ride-hailing-effect-more-cars-more-trips-more-miles/542592/

[66] Thor Berger, Chinchih Chen & Carl Benedikt Frey (2017), 'Drivers of Disruption? Estimating the Uber Effect', https://www.oxfordmartin.ox.ac.uk/downloads/academic/Uber_Drivers_of_Disruption.pdf

[67] BBC News (2017), 'Uber and Lyft ordered to share data with San Francisco', http://www.bbc.co.uk/news/technology-40172017

[68] https://movement.uber.com

[69] https://www.sharedmobilityprinciples.org

[70] Open Data Institute (2018), 'Understanding the impacts of peer-to-peer accommodation, the role of data and data observatories (report)', https://theodi.org/article/understanding-the-impacts-of-peer-to-peer-accommodation-the-role-of-data-and-data-observatories-report

Bikesharing

Bikesharing has great potential to support the travel needs of urban populations and has been embraced by people in cities around the world[71]. In the UK alone, 20 bikeshare operators have launched in 10 years. These services let people who can cycle travel more flexibly and actively while reducing congestion and pollution.

Docked and dockless services like Urbo, obike and ofo[72] use applications and bank cards to collect account and bank details as well as some journey data. Operators collect GPS data directly from bikes to locate and move them to match customer demand. Some also collect journey data from users, via a smartphone or connected device linking them to the bike, sharing that data with the user, and sometimes selling it to third parties.

Some bikeshare services are exploring ways of sharing data that help urban planners make better planning and investment decisions. For instance, Urbo aims to work with local authorities whenever they enter a new city. Services increasingly share usage data with local authorities, and it is now often a condition for dockless services seeking local authority consent.

> " *"We always open a line of dialogue with the government and local council… it's very much a collaborative relationship"*
> – Eanna Lalor, Urbo

Some bikeshare services are working to increase accessibility for excluded groups. For example, Germany's Nextbike service can be accessed via an RFID card bought in shops, to give access to people with no smartphone or bank card. Hourbike offers electric bikes in some locations, to give access to those less able to cycle.

Open standards and industry bodies are helping these services share more data and in ethical and equitable ways. The General Bikeshare Feed Specification, started in the US, is a new operator-led standard for sharing real time data on bike availability from different services. It increases accessibility to bikeshare services by helping journey planners include them in their applications, and improves urban planners' decision-making by giving them aggregated mobility patterns.

Industry groups are also emerging to help these new services engage with key stakeholders in the public and private sectors, including large employers and community groups. For example, Bikeplus, the representative body for bikeshare services in the UK, analyses and shares data and works to ensure social and environmental benefits are maximised and evenly spread[73].

Journey planners

Journey planners are growing in popularity around the world, transforming how we access and analyse transport ecosystems. Applications like Google Maps, Moovit and Citymapper, typically accessed via smartphones, combine transport data from many modes of transport with maps to help people choose and monitor trips from and to precise locations. This makes them 'end-to-end' and 'multimodal' services.

---

[71] http://bikes.oobrien.com

[72] See http://www.myurbo.com, https://www.o.bike and https://www.ofo.com

[73] https://www.carplusbikeplus.org.uk/projects/bikeplus

Users can give these applications permission to collect journey data, anonymously or linked with other data about them. As well as helping to plan journeys, it can help users track other things, such as distances travelled, calories burned or money saved. It can also help improve the accuracy of the application's data for all users, and help the application itself develop other services, such as data analytics for public and private mobility providers, or new transport operations.

Moovit crowdsources data from users to plot public transport stations and routes in cities with little or no open transport data. This helps people in those cities access and engage with public transport more easily, and those cities can license this information to improve their data infrastructure[74]. Citymapper encourages cities to partner with it to develop open data, and also uses the journey data it collects to identify and fill gaps in transport services (see Smart Buses in the UK case study).

The benefit of open data infrastructure to these companies has influenced how they communicate. They strive to explain the different ways and value of sharing data to users, building communities and making the 'greater good' part of their service and ethos. This can increase people's engagement as they are motivated to help improve a system for the benefit of others as well as themselves. Companies with this type of engagement and business models that back it up foster and earn trust.

Journey planners have real time network views of how people move through transport ecosystems. These overarching views could be of great value to public services that often provided the open transport data which helped these private services start. For instance, Moovit's public transit index shares anonymised and aggregated data on how people experience public transport in cities around the world, from wait times, to number of changes in a single journey and the average length of a commute[75].

Further exploration of such sharing is encouraged, particularly where it could help public transport fill service gaps that private companies have identified but can't or won't fill themselves. If MaaS develops to position journey planners as centralised gateways into our transport ecosystems, it will also be important their displays continue to enable innovation by supporting awareness of and enabling access to new mobility services as they emerge, using open APIs.

Case studies for the future

Digital Matatus

Collecting journey data for open data infrastructure in informal transport systems

Nairobi's Digital Matatus was the first group to develop open data infrastructure for the informal transport systems common in cities across Africa, Asia and Latin America[76]. The University of Nairobi's C4D Lab, MIT's Civic Data Design Lab and Boston design agency Groupshot collaborated to map the city's network of matatus, privately-owned minibuses used by most Nairobi residents[77].

To do this, university students collected journey data using mobile phones and GPS technology. They identified routes and stops, as well as fares, all of which can vary journey to journey. To standardise and open the data, they

---

[74] https://www.solutions.moovit.com
[75] https://www.moovitapp.com/insights
[76] http://www.digitalmatatus.com
[77] 'The Digital Matatu Project', https://doi.org/10.1016/j.jtrangeo.2015.10.005

asked Google to help them modify the GTFS open standard, built for formal transport systems with fixed routes and stops, to fit informal transport systems with a lot of variable data.

The group used the data to publish route maps in both digital and paper form. Because the data was standardised using GTFS, Google and OpenStreetMap could also include the routes in their maps. Making the routes visible and accessible helped passengers save time by finding routes to meet their travel needs easily. It also changed travel experiences, for instance helping women feel safer at night[78].

Making the data open and easy to use allowed others to create further benefits. Matatu drivers could see underserved areas and try new routes. Third-party developers used the data to create new applications such as ma3route, which crowdsources location and other data to provide updates on traffic and driving conditions[79]. New mobility services like this also expanded the city's transport ecosystem to include its technology community. The data can also support transport planning to help combat the estimated $1 billion Nairobi losses a year to wasted gridlock hours[80].

Similar initiatives are also running in Manila, Accra, Dhaka and Mexico City[81]. Linking resources and skills across cities, Digital Transport 4 Africa is a collaborative initiative using open data to improve urban transport throughout Africa[82]. Could learnings from these initiatives also be applied to cities with more formal public transport where frequent journeys via new mobility services may be evolving into informal transport systems?

Smart ticketing

Ticketing is a core function of any transport system and has changed significantly in recent years. Many public transport operators have moved from paper to magnetic tickets, then barcodes, and now mobile ticketing, smart cards and contactless payments. These new technologies have simplified travel for people and allowed transport operators to collect journey data and use it to make better planning and investment decisions, as well as lower the cost of their ticketing solutions.

Smart cards using RFID and NFC technology let passengers in cities such as Hong Kong, Sydney, Stockholm and London access services across multiple modes of transport with a single card[83]. In London, people can use contactless bank cards. Other cities, such as Oslo and Helsinki, focus on mobile ticketing and smartphones.

While many rural communities do not yet have access to smart ticketing, some countries have moved to a national ticketing platform that can be used in urban and rural area and for all modes of public transport. This solution is seen, for example, in

[78] Wired (2015), 'How Nairobi Got its Ad-hoc Bus System on Google Maps', https://www.wired.com/2015/08/nairobi-got-ad-hoc-bus-system-google-maps

[79] https://www.ma3route.com

[80] Reuters (2015), 'Nairobi's billion dollar traffic blues', https://www.reuters.com/video/2015/03/23/nairobis-billion-dollar-traffic-blues?videoId=363596944

[81] See https://sakay.ph, http://data.afd.fr/accramobility, Albert Ching (2012), 'A User-Flocksourced Bus Experiment in Dhaka: New Data Collection Technique with Smartphones', http://web.mit.edu/czegras/www/Flocksource_JUT.pdf and Quartz (2016), 'Mexico City is attempting to map its more than 1,000 unwieldy bus routes with a crowdsourcing app', https://qz.com/598895/mexico-city-is-attempting-to-map-its-more-than-1000-unwieldy-bus-routes-with-a-crowdsourcing-app/

[82] http://digitaltransport4africa.org

[83] https://en.wikipedia.org/wiki/List_of_smart_cards

The Netherlands via OV-Chipkaart and in Denmark via Rejsekort. These systems are spreading the benefits of smart ticketing more equitably and could provide rich journey datasets to help policy makers make better planning and investment decisions at a national level. Some systems, such as OV-Chipkaart, let users choose between an anonymous version of the card, which does not store personal data, and a personal version which does, and offers benefits as a result[84].

Smart cards dedicated to specific groups of people can also help public services ensure their travel needs are met. In the UK, West Yorkshire used journey data collected on Senior Pass smart cards to see how people over 60 use buses there[85].

Smart ticketing also means smartphone makers and banks are now part of our transport infrastructure. Smartphones have helped new mobility services to emerge and enabled the collection and sharing of real time data. Smartphone makers can already collect large amounts of location data from users, storing or using it in different ways. If bank cards become a common form of ticketing, they can also build large journey datasets and link it with other data. It is paramount that organisations with such large amounts of personal data use it ethically and equitably at all times, and in ways users can understand.

---

[84] https://www.ov-chipkaart.nl/purchase-an-ov-chipkaart/anonymous-ov-chipkaart.htm
[85] Data Mill North (2017), 'English National Concessionary Travel Scheme boarding data for West Yorkshire', https://datamillnorth.org/dataset/english-national-concessionary-travel-scheme-boarding-data-for-west-yorkshire

A framework for the future

Data-enabled transport must combine innovation with ethics in open ways that build trust

The future value of transport to our societies and economies is much more than financial. Quantifying the potential of data-enabled transport can yield some striking results: as noted earlier, intelligent mobility may be worth £900bn by 2025; not sharing transport data may cost an economy like the UK £15bn. But it is how transport can change our lives and how we feel, its qualitative value, and businesses' reliance on access to data about us to make this possible, that makes the need for an approach to data in transport that combines openness and innovation with ethics and trust so urgent and great.

As shown earlier, personal data like journey data can help transport organisations be more efficient by understanding and meeting travel needs in more targeted ways. This can change our quality of life – better transport can help us live longer, sleep more, even divorce less[86]. It can also affect how we feel – when transport gives us a greater sense of control it can increase our feelings of independence or happiness[87].

Businesses can build sustainable models delivering these services by being more open, as well as fostering and earning the trust of people who share data about them. As services like journey planners and bikesharing show, businesses can in turn make the data infrastructure they build more open to support the public transport infrastructure and interoperability they need to continue to operate and innovate. This can help make transport better for everyone and gain people's trust while also raising awareness of a business's services with potential customers.

This sharing in all directions and the innovation and trust it can foster is core to how transport works. Public and private transport services operate interdependently as a person's single journey will often involve several modes of transport. New mobility services innovate using data made openly accessible by public services which have a special mandate to serve the public good. New businesses like ridesharing and bikesharing find innovative ways to share resources, and can help public services and accredited researchers better understand people's travel needs. In areas with less public transport, private transport services can even help create foundations for data infrastructure which leads to new physical infrastructure (see the Digital Matatus case study). In transport, a more open approach can create value for all.

Contrast this with the HERE survey noted earlier which found that while nearly 70% of us will share location data if we feel in control, only 20% currently feel that way, and 80% of us feel nervous or vulnerable. A loss of trust is bad for any business, but for consumer-facing groups using personal data, the risk can be far more than reputational and the cost material. For example, research suggests 5% of Facebook users deleted their account in just the first few weeks of the Cambridge Analytica crisis[88]. As new services become more commoditised, and as data about people becomes more portable, these risks and costs for businesses using personal data

---

[86] Lifehacker (2012), Long Commutes Are Sucking the Life Out of You: Shortening Yours by 20 Minutes Could Save Your Health', https://lifehacker.com/5908879/long-commutes-may-be-killing-us-shorten-yours-by-20-minutes-to-save-your-health

[87] See https://ts.catapult.org.uk/readyforinnovation and The Happiness Index (2017), 'How your morning commute affects your happiness and wellbeing', https://the-happiness-index.com/does-your-morning-commute-affect-your-happiness-and-wellbeing/

[88] Campaign (2018), 'One in 20 Brits delete Facebook accounts after the Cambridge Analytica scandal', https://www.campaignlive.co.uk/article/one-20-brits-delete-facebook-accounts-cambridge-analytica-scandal/1460836

are likely to go up, not down.

Some may assume that the compelling nature of their service makes a loss of trust affordable in the short term, others that the cost of a loss of trust is low anyway because customers get hooked to convenience. Even if both of these assumptions are true when tested, shareholders with long-term investment targets and ESG (environmental, social and governance) mandates are unlikely to tolerate them. And if customers and shareholders do nothing, regulators will intervene on their behalf.

Similarly, as personal data becomes part of business models, commercial imperatives and social responsibilities are harder to separate. An 'us versus them' approach can erode trust when exposed[89]. And for a universal need such as transport, how personal data is used and shared affects employees and investors just as it does customers. A large-scale loss of trust can create losses for all, through increased risk-aversion, social instability and even violence[90].

For these reasons, we urge transport businesses who use personal data to become more open and build trust now. Doing so will both reduce business risks and increase growth opportunities.

Recommendations

1. Use existing open transport data to get a competitive edge

This additional insight at low cost can support internal decision-making, improve your operations and help your organisation engage with and lead sector change.

Actions could include:

- using third-party sources of data in your internal processes
- combining your data with other data from public services, other businesses or community groups
- engaging with existing external groups enabling open data for transport.

Businesses hoping to gain a competitive edge in this way can explore open transport data provided by governments, public transport operators and open data groups[91], or on platforms like GitHub. Examples of existing sector initiatives at the national level include the UK's Transport Systems Catapult and Belgium's Intelligent Transport Systems. Broader alliances include Digital Transport 4 Africa and Open Transport Net[92].

2. Turn your transport data into infrastructure that can benefit all

This can let other organisations integrate your transport services with theirs, help people access your services more easily, and inform decision-making for public transport infrastructure. It can also help your teams collaborate more internally.

---

[89] TechCrunch (2018), 'Facebook retracted Zuckerberg's messages from recipients', https://techcrunch.com/2018/04/05/zuckerberg-deleted-messages

[90] Pew Research Centre (2017), 'The Fate of Online Trust in the Next Decade', http://www.pewinternet.org/2017/08/10/the-fate-of-online-trust-in-the-next-decade

[91] https://github.com/theodi/shared/wiki/Finding-Open-Data

[92] See https://ts.catapult.org.uk, http://its.be, http://www.digitaltransport4africa.org and http://www.opentnet.eu

For instance:

- Adopt existing open standards and create open APIs that help your services integrate into journey planners and public planning processes
- Find other groups with a common or adjacent need that can be addressed using data and partner with them to build new data infrastructure
- Identify data you manage anyway that could be shared or made open safely and at low cost to benefit customers or public services

See how TransportAPI and Navitia.io use open APIs in transport[93], how Thomson Reuters has built open infrastructure for and with its customers[94], and an idea for how location data could be shared with local authorities to improve air quality[95].

CRITICAL QUESTIONS
What does 'as open as possible' mean for journey data?

Making data as open as possible can create benefits for all, and as open as possible means different things when dealing with different data. For personal data, a balance must be struck between the benefits of providing access to data – and individuals' rights to privacy and being protected from harm. It means this data can only be shared in very safe ways and, if the person it is about has not consented, not at all. Only in very rare cases – such as politicians' expenses – should identifiable personal data be open for everyone to access and use.

Developing ways to share journey data safely and securely and in ways that foster trust with people will be hard, but we encourage transport organisations to explore doing it in three ways:

1. As open data – with careful and appropriate aggregation and anonymisation, and accompanied by privacy impact assessments

2. In secure environments – in particular with accredited researchers, who should openly publish their results

3. With people the data is about – through open APIs, in ways that enable innovation

There are several examples of ways this can be done in transport and other sectors. For instance, Transport for London makes network statistics such as Oyster card journey data open to support innovation[96], while Uber Movement makes anonymised and aggregated journey data open to help urban planning[97]. Open banking in the UK is using open APIs to give customers more control over data about them[98], and Projects by IF have explored other prototypes for open APIs using personal data[99].

---

[93] See https://www.transportapi.com and https://www.navitia.io

[94] Open Data Institute (2016), 'Open enterprise: How three big businesses create value with open innovation',
https://theodi.org/article/open-enterprise-how-three-big-businesses-create-value-with-oPen-innovation

[95] Projects By If (2018), Improving a city's air quality using bulk location data from mobile phones,
https://openapis.projectsbyif.com/scenario-3-improving-a-citys-air-quality-using-bulk-location-data-from-mobile-phones

[96] https://tfl.gov.uk/info-for/open-data-users/our-open-data

[97] https://movement.uber.com

[98] https://www.openbanking.org.uk

[99] https://newdigitalrights.projectsbyif.com

3. Use impetus from the GDPR to drive data literacy and ethics

This can differentiate your business from others and build customer loyalty. It can also help your teams make more informed decisions, and help your organisation stay ahead of further potential regulation on the use of personal data.

Consider:

- going beyond technical and regulatory requirements to develop governance guidelines and processes for the collection and use of personal data which give people more control over data about them
- providing training and resources to support different levels of data literacy across your technology, strategy, operations, marketing and other teams
- generating debate and discussion, internally and externally, about the use of personal data and exploring different ways to use and store it

The Data Ethics Canvas developed by the ODI can help identify, debate and address issues around how data is used by your organisation[100]. Co-op is using the GDPR to drive a broader debate on the use of personal data[101]. Caution Your Blast and Birmingham City Council have piloted a journey planner that stores personal data differently[102].

> **CRITICAL QUESTIONS**
> **What's next in the regulation of personal data?**
>
> The GDPR is one of the most comprehensive pieces of regulation yet to deal with personal data and will likely lead to more regulation in future, both from the EU and individual countries. Just as technology can evolve quickly and policy discussions around the use of personal data can move fast, so organisations have a better chance of informing and being prepared for further regulatory change if they engage in public debate.
>
> Potential focuses of future regulation may be group privacy and actions[103] (while the GDPR allows group actions, national laws like the UK's Data Protection Bill do not), inspection measures and conditions around when (if at all) people can sell data about them. Authorities like the UK's ICO are also likely to make new recommendations about how personal data can be used in light of issues around Facebook and Cambridge Analytica. In future, businesses may also be able to use data protection sandboxes to develop new services in a way regulators can assess as they emerge[104].

4. Engage openly with people on how data is collected and used

This can let your organisation show customers how you consider and care for their privacy needs, for instance explaining how personal data is safeguarded as well as when and why it is shared. This can build trust and address privacy concerns so

---

[100] https://theodi.org/article/data-ethics-canvas

[101] Co-op Digital (2017), 'How the ODI is helping Co-op Digital put data at the centre of the organisation', https://digitalblog.coop.co.uk/2017/12/12/how-the-odi-is-helping-co-op-digital-put-data-at-the-centre-of-the-organisation/

[102] https://www.cautionyourblast.com/smart-routing

[103] Linnet Taylor (2017), 'Group Privacy: The Next Generation of Privacy Problems', https://linnettaylor.wordpress.com/2017/01/10/group-privacy-a-new-book-on-the-next-generation-of-privacy-problems/

[104] The Register (2018), 'UK regulator moots data protection sandbox for organisations to play in', https://www.theregister.co.uk/2018/03/05/ico_promises_data_protection_sandbox_for_organisations_to_play_in

people continue to share data about them, and strengthen customer relationships more broadly.

Ways to engage more openly include:

- designing a user experience for your service so people know when and how sharing data about them creates benefits so they can make more informed decisions about whether or not to use the service
- enabling more granular control over types of personal data and sharing
- engaging with customers through surveys, focus groups and other feedback forums to understand what data-sharing different groups are comfortable with, for example urban commuters in cities with multiple mobility services, or remote rural communities with few transport services
- embedding a user experience for any automated decision-making using personal data, in order to explain why a decision was taken[105]. This both complies with the GDPR and builds trust with users (self-driving vehicles are an obvious focus here)
- clearly telling users which, if any, third parties their personal data is shared with, what is shared, and how it is used. For example, a mobility service may need to share personal data with a bank to take payment
- if both collecting personal data and selling it to third parties, making every effort to explain the trade-offs and risks to both the individuals the data is about and the third parties buying a product based on it. Over time this may even suggest a new kind of caveat emptor that is closer to caveat venditor, 'seller beware', for business models that offer a free service in exchange for personal data, and use that personal data to sell another service to third parties. This type of shift in responsibility becomes more important the greater the asymmetry of power between buyer and seller in both types of transaction.
- publishing more detailed assessments on the use of personal data in an accessible way, whether they are mandatory privacy impact[106] or legitimate interests[107] assessments, or voluntary ones such as the data ethics canvas mentioned earlier.

See how journey planners like Moovit and Citymapper engage with users on how they use data. Consider the ODI's openness principles for organisations handling personal data[108]. Explore Projects By IF's open source catalogue of design patterns for consent and sharing data[109].

5. Use open innovation to be more efficient and equitable

Opening up your data for research and development can help address your organisation's specific challenges more efficiently, as well as systemic challenges in a wider context. It can also widen your networks and access to new ideas. The engagement and accessibility around open innovation, and its potential to support transport systems more broadly, can also support your implementation of other recommendations made here, by increasing your ability to capture the benefits of open data and building more trust with people.

Organisations can engage in open innovation by:

---

[105] https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling
[106] https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments
[107] https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/legitimate-interests
[108] Open Data Institue (2016), 'Openness principles for organisations handling personal data',
https://theodi.org/article/openness-principles-for-organisations-handling-personal-data
[109] https://catalogue.projectsbyif.com

- working with open data startups
- giving academics access to their data
- exploring new ways to collaborate with public services around data
- getting involved with sector innovation programmes using open data
- using data portability to help customers share data to support innovation

For examples in transport, see how data accelerator [Data Pitch](#) is working with [Deutsche Bahn](#)[110] and explore TfL's [innovation portal](#)[111]. Explore how GTFS began with collaboration between public and private organisations in the US (see Open Standards in Transport), and expanded by partnering with academics in Kenya (see the Digital Matatus case study). In other sectors, see how [Arup](#) works with open data startups[112], how [OpenActive.io](#)'s sector innovation programme is developing new ways to help people get active[113], and how the [Open Up Challenge](#) is using data portability in UK banking to support innovation[114].

> **CRITICAL QUESTIONS**
> **How can re-identification risks be mitigated?**
>
> Our research showed that a number of transport organisations want to share anonymised and aggregated journey datasets to support our transport systems and enable innovation, but they are very concerned about [re-identification risks](#)[115]. The technical complexity of these risks was highlighted again recently when anonymised and aggregated open data shared by Strava was subsequently shown to raise concerns about [national security](#)[116].
>
> A lot more work needs to be done around de-identification processes and re-identification risks, and there are initiatives organisations can join now to help do this. For example, the [UK Anonymisation Network](#) has set out an anonymisation decision-making framework[117] which can be adopted alongside techniques from the ICO's [Anonymisation Code of Practice](#)[118]. In the US, the Future of Privacy Forum has developed an [assessment](#) model for preserving privacy in open data and suggested strategies to mitigate re-identification risks[119].
>
> More must be done to reduce these risks to preserve trust in services using journey data and to enable open data that can make transport better for all.

6. A collaborative approach to transport data could benefit the sector

Sharing insights and resources across the sector can maximise the benefits of transport data for our economies and societies, reduce costs and risks for individual companies, and build public trust.
Transport businesses can pilot working together around shared technical problems such as re-identification risks, around more complex and ultimately philosophical

---

[110] https://datapitch.eu

[111] https://tfl.gov.uk/info-for/business-and-commercial/innovation-portal

[112] Open Data Institute (2016), 'Open enterprise: How three big businesses create value with open innovation', https://theodi.org/article/open-enterprise-how-three-big-businesses-create-value-with-Open-innovation

[113] https://www.openactive.io

[114] https://www.openupchallenge.io

[115] 'De-Identification of Personal Information', http://dx.doi.org/10.6028/NIST.IR.8053

[116] The Guardian (2018), 'Fitness tracking app Strava gives away location of secret US army bases', https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases

[117] http://ukanon.net

[118] https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf

[119] Future of Privacy Forum (2018), 'FPF Publishes Model Open Data Benefit-Risk Analysis', https://fpf.org/2018/01/30/fpf-publishes-model-open-data-benefit-risk-analysis

issues like informed consent, or campaigns raising public awareness of and trust in the collection and use of data.

These organisations can also collaborate locally, nationally or even globally, for instance joining initiatives mentioned earlier like the UK's Transport Systems Catapult. Organisations can also collaborate within a single mode of transport or type of mobility service, like Bikeplus in the UK, or as a multimodal group such as the MaaS Alliance[120]. These different groups can then share knowledge or coordinate around broader shared problems relevant to them all as they emerge.

Such collaborative groups can also more effectively engage other stakeholders, like:

- public bodies responsible for transport policy, infrastructure and regulation
- civic society, consumer rights and other groups representing passengers
- newer stakeholders in transport like smartphone makers and banks
- open data and personal data experts

Sector-wide groups can also drive change programmes by creating a forum to:

- develop, advocate and measure transport-wide data strategies
- give guidance on best practices in transport data governance
- research the role of and impact of journey data and other data in transport
- standardise open data protocols for transport and provide tools for their use
- facilitate data training via face-to-face workshops, eLearning and webinars
- run competitions and accelerators for data-enabled innovation in transport

---

[120] https://maas-alliance.eu/the-alliance

Appendix: Methodology

For this research, we commissioned the TravelSpirit Foundation to produce an overview of the relevant literature related to the use of journey data within the transport ecosystem. In addition to submitting regular reports of their findings, TravelSpirit planned and conducted a workshop for ODI staff in order to help us identify the modes of transportation worth focusing on and the types of themes worth exploring.

We then conducted a landscape review which involved extensive research of the current use of journey data by organisations involved in the transport ecosystem. We identified a list of startups and large firms using journey data to improve services for customers, which we analysed and categorised according to the way that journey data was deployed within each organisation. A number of businesses were then selected for targeted, in-depth interviews.

We conducted 11 interviews with participants from a range of backgrounds, striving for a balance between people from startups, SMEs, public sector organisations, research institutions and academia. Our interviewees were mainly from companies based in the UK (7), though we also interviewed one person each from Ireland, Singapore, Israel and Finland.