

# AI and data governance from principles to practice: auto insurance

Workshop report  
20 September 2018

# Summary

In 2017, the Royal Society and British Academy's joint report *Data management and use: governance in the 21st century* called for a connected approach to governance of data use<sup>1</sup>. This means the need for stewardship of the data governance landscape; shared principles to underpin that governance; and learning lessons and sharing best practice across sectors. This workshop explored how principles for governance of data use might apply in practice for the automobile insurance sector and set out to identify examples of good practice.

The automobile insurance sector sits at the intersection of the automobile sector and the insurance sector, which both stand to be transformed by advanced data capture and analysis. Additionally, these are consumer-facing sectors which involve granular, individual-level data. As a result, automobile insurance faces a broad set of challenges and opportunities, and how the sector responds could in turn influence how the automobile sector and the insurance sector more widely adapt to data's potential.

Workshop attendees considered how each of the principles proposed in *Data management and use* played out in the automobile insurance sector. By doing so, they explored the ways in which advanced data capture and analysis in the automobile sector could contribute to improving driving and road safety. Examples included:

- mechanisms for standardising the collection and analysis of data across vehicles and initiatives for data sharing about the risks and causes of vehicle accidents;
- mechanisms for anticipating and clarifying possible concerns around cybersecurity in connected and autonomous vehicles;
- the potential for connected and autonomous vehicles to improve access to transport for social groups that might currently be disadvantaged or disenfranchised in this regard – as well as potentially create a new category of vulnerable groups or individuals; and
- how industry and government could collaborate to create the necessary infrastructure for connected and autonomous vehicles.

The workshop also explored the ways in which data-driven disruption and innovation in insurance and automobile insurance could lead to better priced and fairer products for consumers. Examples included:

- the development of new services and business models;
- mechanisms for improving transparency and fairness in automobile insurance;
- the balance between consumer privacy about lifestyle choices and insurer transparency about which data is used in pricing or personalising services; and
- the role of data controller accountability in insurance pricing and data privacy impact assessments.

The growth in the availability of data, and changing expectations around car ownership and access to personal transport, are likely to disrupt both the automobile sector and automobile insurance – creating change in business models and raising questions about the need for new forms of governance. These changes all depend on the increasing importance of understanding and standardising the value, quality and provenance of data.

*AI and data governance from principles to practice: auto insurance* was a joint workshop between the Royal Society, The Alan Turing Institute, the Leverhulme Centre for the Future of Intelligence and the Royal Academy of Engineering. This report is a summary of discussions at the workshop and does not present the views or positions of any of the partner organisations.

# Background

## Data governance in the 21st century: from principles to practice

In 2017, the Royal Society and British Academy's joint report *Data management and use: governance in the 21st century* argued that new uses of data create a series of pervasive tensions and disconnects which illustrate the kinds of dilemmas that society will need to navigate<sup>2</sup>. Because of this, grounding data governance efforts in underlying principles will provide a source of clarity and of trust across application areas. The report argued for an overarching principle that systems of data governance should promote human flourishing. This framing includes concepts such as wellbeing and the need for individuals and communities to prosper. Four high-level principles complement the need to promote human flourishing, as a framework for well-founded debate about the tensions inherent in data governance. These principles are:

- enhance existing democratic governance;
- protect individual and collective rights and interests;
- ensure transparent, accountable and inclusive decision-making in data trade-offs; and
- seek out good practice and learn from success and failure.

A detailed outline of the existing tensions and disconnects in data management and use, and our principles for data governance, is provided in Annex A. This workshop was part of a series of activities to explore how these principles can be practically applied in different sectors, underpinning models of good practice in data governance for human flourishing. The discussion at the workshop related to each of these principles in turn, exploring what they mean for governance of data use in automobile insurance.

# Context

## Data disruption in automobile insurance

The automobile insurance sector sits at the intersection of the automobile sector and the insurance sector, which both stand to be transformed by advanced data capture and analysis. These are consumer-facing sectors which involve granular, individual-level data, and as a result automobile insurance faces a broad set of challenges and opportunities. How the sector responds could in turn influence how the automobile sector and the insurance sector more widely adapt to data's potential.

At the heart of this transformation is telematics – the storing, sending and receiving of data between or about remote objects. In the automobile sector and automobile insurance sector, these remote objects are typically sensors or devices internal to the vehicle, but potentially also external to it (such as satellites or road cameras). Telematics data is therefore a rich category of data encompassing, for example, driver-generated data, geolocation data and satellite images, among other things. Telematics data can now be combined with other rich datasets and data sources to build sophisticated models of driver behaviour and the lifecycles of vehicles. Consequently, a number of data governance questions arise. For example: what data about an individual is it acceptable to use when determining an insurance premium? At what point does differential pricing of insurance premiums become a form of discrimination? Who has rights over the data about a driver that is collected when driving and what privacy safeguards should exist? What data should be collected in the event of an accident, and could there be any duties to share data?

### AI and data disruption in the automobile sector

Connected and autonomous vehicles incorporate a range of different technologies, potentially improving road safety and efficiency, and potentially also widening access to transport for people with mobility difficulties. Increased connectivity allows vehicles to communicate with their surrounding environment, providing information to the driver about road, traffic and weather conditions. Increased automation uses information from on-board sensors and systems to analyse a vehicle's position and environment and enable the vehicle to control its own functions without the input of a human driver.

However, there might need to be differences in the frameworks that apply to connected vehicle data and the data used or generated by automated driving systems. Connected vehicle data is already regulated in some jurisdictions, such as Event Data Recorders (EDRs) as mandated in the US, and so can typically be stored and standardised in some ways. In contrast, autonomous vehicle data differs between developers, and is heavily dependent on the suite of sensors and software system used. Additionally, even a common class of sensor might have different manufacturers and specifications in the market, making it more challenging to apply common standards or treatment across autonomous vehicles.

The market for connected and autonomous vehicles can be stratified as level 0 – 5, depending on the vehicle's level of connectivity and autonomy, as set out in Box 1<sup>3</sup>.

The capability gap between levels 1 – 2 autonomy and levels 4 – 5 autonomy is significant. Consequently, the transition period from 'dumb' vehicles to fully connected and autonomous vehicles will face significant challenges as a 'mixed economy' of human drivers and partially or fully automated vehicles share transport infrastructure. In 2017, the Transport Systems Catapult (TSC) in collaboration with the Centre for Connected and Autonomous Vehicles (CCAV) published a report estimating that by 2035 the connected and autonomous vehicles sector would be worth £28bn to the UK economy<sup>4</sup>. The report also estimated that by 2035 approximately 31% of new vehicles would be considered levels 4 – 5 connectivity and automation, but the figure could be as low as 5% if 'remaining challenges for autonomy are not resolved quickly and many consumers remain suspicious or untrusting of the technology'. On this basis, the length of the transition period could run into decades; but the impact on UK road networks of this 'mixed economy' or 'mixed fleet' is not yet understood<sup>5</sup>.

## Levels of vehicle connectivity and autonomy

**Level 0** vehicle connectivity and autonomy is when a human driver performs all aspects of dynamic driving tasks and capability is not in place for either tracking or monitoring vehicle usage or performance or for digital services. These vehicles have no connectivity or automated driving functionality.

**Levels 1 – 2** vehicle connectivity and autonomy is a market segment that is beginning to gather momentum and to attract the focus of automobile manufacturers. This includes features such as advanced driver assisted systems (ADAS) and adaptive cruise control. A strategic element of this market segment is for retrofitting, where data-enabling devices or apps are added to older or less advanced vehicles in order to keep abreast with industry expectations set by the new generation of vehicles that are manufactured with data-enabling devices embedded. For example, on-board diagnostics (OBD) are a category of computer port that will allow an engineer to identify and analyse potential faults in the vehicle and that might be included in the design of new vehicles but that can also be retrofitted to older vehicles.

**Level 3** vehicle connectivity and autonomy is when the application of data capture and analysis enables a vehicle to effectively be in control of core functions such as accelerating, braking and steering. Crucially, the human driver remains responsible for the vehicle at all times, must always be in a position to resume control and may be requested to intervene.

**Levels 4 – 5** vehicle connectivity and autonomy is a market segment that is in its infancy and that is the focus of specialist companies. The work is trailblazing in establishing methodologies, but there is not yet sufficient data for the work to be translated into real-world application and used at scale. Regulation for these levels of connectivity and autonomy is still developing: it's anticipated that an automated driving system will need to demonstrate safety across a comprehensive range of driving scenarios. Simulation will be heavily used and the volume of data that would need to be collected for this is in the several billions of miles. For context, Google Waymo is estimated to have collected around 500 million miles of data at sufficient granularity to start to support modelling for levels 4 – 5 autonomy.

### Data disruption in insurance and automobile insurance

Data-driven transformation in the automobile sector is combining with data-driven transformation in the insurance sector to present a broad set of challenges and opportunities for automobile insurance. The insurance sector is already being disrupted by advanced data capture and analysis<sup>6</sup>: the speed with which a claim can be resolved is vital, since it can deliver savings of up to £1,600 per claim. This speed could be accelerated either through quicker access to, and analysis of, relevant data; or by automated decision-making in processing claims.

The insurance business is predicated on analysing the risk of certain events and the likely impact of them. This gives rise to a broader unintended consequence of the new availability of significant datasets and sophisticated data analysis: the insurance market might be deconstructed because of the precision with which the risk of an event and its likely impact can be modelled and predicted. It also raises significant concerns for consumers.

For example, is a fair price an accurate price which reflects actuarial risk, or is a fair price an equitable price for the consumer? In 2012, the European Court ruled that it was unacceptable for there to be a difference in the automobile insurance premiums offered to men and to women, even though men are typically less safe drivers than women, because this would be a form of gender discrimination<sup>7</sup>; however, in 2017 *The Guardian* argued that offering lower automobile insurance premiums to women to reflect their reduced risk would have been fairer and that the economic cost of the higher premiums have since widened existing gender inequalities between men and women<sup>8</sup>.

# Enhancing existing democratic governance

## Business models and regulation

The Royal Society and British Academy joint report *Data management and use* argued that data management and use should support democratic processes, help enact democratic decisions and be subject to democratic oversight. An important mechanism for this is regulatory frameworks and their role in innovation, and in managing disruptive new business models and markets, to promote human flourishing. The workshop considered ways in which data-enabled digital disruption in the automobile, insurance, and automobile insurance sectors built on existing regulation and created new governance considerations.

### Regulation and liability

Regulation can support innovation by providing a legal framework to navigate new or ambiguous concepts and building citizen or consumer confidence in the innovation by clarifying and protecting their rights. However, over-regulation can restrict innovation in a way that can inhibit the development of new markets until the social and economic costs of the restrictions compel the regulation to be revisited. At present, advanced data capture and analytics in the automobile sector are regulated by the 1999 Database Directive<sup>9</sup>, and the 2018 General Data Protection Regulation (GDPR)<sup>10</sup>.

### Data sharing after vehicle accidents

How can advanced data capture and analytics generate better insight into the causes of vehicle accidents, from telematics data, vehicle performance data and driver behaviour data? What responsibilities and liabilities might this create?

- If data about the causes of accidents was centralised for accident analysis, effective analysis might need to be accompanied by a stipulation of an acceptable level of road fatalities. In Britain there is no officially-determined level for the acceptable number of fatality-causing road accidents per year, and currently there are approximately 2,000 road fatalities annually<sup>11</sup>. If this level is reduced to zero, it is likely to constrain the functioning of the transport system overall.
- Pooling or sharing data for accident analysis might remove some of the competitive advantage to automobile manufacturers or insurers who have collected that data. In light of this, there could be a role for trusted independent bodies, such as the Royal Society, to work with relevant stakeholders to develop an appropriate regulatory environment.

- If an insurer can ascertain, from data from a connected or autonomous vehicle, that a driver is driving dangerously or breaking laws, might there be a duty of care for insurers to drivers or to the wider public, compelling them to act on what that data tells them?

Seed funding of £480,000 has been provided from the Department for Transport to the RAC Foundation for, potentially, the establishment of a highways accident investigation branch<sup>12</sup> to parallel those that are already in place for rail, marine and aviation<sup>13</sup>. This might also be a suitable independent forum should there be a regulatory requirement that, in the case of accidents, insurers and automobile manufacturers provide the relevant data so that there is a growing data bank on the causes of accidents.

### Standardising data collection and analysis

As vehicles develop higher levels of automation and connectivity, the insurance framework might have to change. This raises the following challenges:

- During a transition period from human drivers to autonomous cars, when there is a 'mixed economy' of drivers and vehicles, in the event of a collision between a vehicle driven by a human and an autonomous vehicle, there could be significant differences in the availability of data for event analysis. The autonomous vehicle is likely to be a very rich source of data, while the non-autonomous vehicle might provide almost no data.
- Consequently, the kinds of insurance available to the user or manufacturer of the autonomous vehicle would be very different from the insurance available for the driver and manufacturer of the non-autonomous vehicle.
- The point of liability for a collision could also be unclear, leading to delays or difficulties in resolving an insurance claim. In light of this there might be a need for standardising the collection of accident data, including identifying the kinds of data that must be recorded in the event of an accident.
- There might be a need to standardise the equipment used in this data collection, so that data quality is assured. There might also be a need to standardise the interpretation of the data, so that the approach is seen to be fair and consistent.

### Cyber-security and software upgrades

New security risks are created by connected vehicles. What responsibilities does this confer onto vehicle manufacturers and owners?

- If a connected or autonomous vehicle has a reputation for being more susceptible to cybersecurity attacks, what would be the implications of this for insurance premiums for individual vehicles and for the robustness of a connected vehicle network overall?
- If owners of connected and autonomous vehicles acquired software patches or downloaded updates from an independent garage instead of from the vehicle dealership, could this impact insurance premiums or the robustness of the vehicle network overall?

### Business and market disruption

Advanced data capture and analytics is having a significant effect on the markets and business models for the automobile industry and for automobile insurers. Some automobile manufacturers might be hesitant to share the data they have access to with insurers, as they explore opportunities to monetise it in new business models; and tech companies with rich datasets and sophisticated data analytics tools are well-positioned to disrupt the market further by creating new insurance-orientated products and services. What might this mean for the market and for consumers?

### Price stretching

How is the richness of available data transforming insurance pricing? Insurers have traditionally sought more information from customers about themselves and their vehicles; that has enabled pricing to be differentiated between customers and also for pricing to be stretched away from an average. How might this be amplified in the context of rich data?

- Information that was originally declared by the customer is increasingly being augmented by a wide array from other data sources, which may be both public and private, or granular data from the driver's use of the vehicle. This data richness can affect the risk assessment of the individual and the vehicle and allow price stretching to increase.
- As cars become autonomous and connected, their capabilities and programming become part of the insurance pricing calculation – potentially reducing the scope for price differentiation or stretching. But insurance companies cannot assume that the data they consider relevant will be captured or made available as an add-on or by-product of an autonomous vehicle's design.

- At levels 1 – 2 vehicle connectivity and autonomy, an app that functioned as an automated backseat driver could give a human advice for becoming a better or safer driver. By levels 4 – 5 autonomy, these technologies are expected to reduce accident frequency or severity – and potentially also prevent some kinds of accidents from occurring at all.

### New services, business models and incentives

These developments, in combination with the increased capability for insurers to assess and predict risk, is expected to reduce the volatility of risk over time. There are a number of consequences of this change:

- Increasing accuracy in predicting risk is likely to result in less market differentiation around price for insurance customers, and narrower variance in risk in the customer base for insurers.
- Consequently, some of the tools that became aspects of competitive advantage for insurers in the past might change and customer data might have less value in the historical aspects of underwriting and pricing, but more value in deriving insight for new products and services. This could replace the lost income from shrinking risk pools of non-crashing vehicles with new income streams.
- Insurers have had to innovate to access richer customer data because customers have proved reluctant to share data from their dashcams to insurance companies. In response, insurers have developed new models offering customers free or waived deductibles in return for data-sharing, even if the customer might be at fault in an accident.
- Some insurance companies are now experimenting with new, green entities that they have set up separately from the current business to scope new products, services and partnerships.

### Vehicle ownership

In the future, the notion of ownership of vehicles might change or attitudes towards vehicle ownership become quite stratified<sup>14</sup>. Presently, older age groups tend to have a stronger sense of personal ownership of their vehicles, whereas younger people are less likely to learn to drive and are more comfortable using short-term vehicle rental services such as Zipcar. Over time, the structure of the industry might also change radically, with mobility being classified as a service rather than vehicles being classified as a product. The anticipated shift from vehicle ownership to mobility-as-a-service (MaaS) will be a significant disrupter for the automobile insurance sector.

# Protecting individual and collective rights and interests

## Consumer rights and social good

Another key principle for data governance from the Royal Society and British Academy report *Data management and use* was that it should protect individual and collective rights and interests. The workshop considered ways in which data-enabled digital disruption in the automobile, insurance, and automobile insurance sectors presented new challenges and opportunities for individual and collective rights, interests and benefits. Automobile insurance may be different from some other kinds of insurance, such as health insurance, where there might be more resistance to personalising insurance premiums to the exact characteristics of an individual.

### Fairness

The importance of transparency about the data that has been used, for example in the calculation of an insurance premium quote, can have a major impact on perceptions of fairness. Transparency about the data sources could also help civil society organisations to examine price discrimination across markets and engage with consumers on their understanding of what might be fair or unfair, and to engage with government on the balance between allowing greater competition versus narrowing access for vulnerable consumers.

### Which data is relevant and representative?

Consumer trust may depend on whether the data that is used to calculate insurance premiums is perceived to be relevant and representative. For example, the use of telematics data to identify and reward good driving is likely to be seen as relevant and fair by most consumers. In contrast, the use of social media data to help improve risk profiling is likely to be perceived by consumers as irrelevant and therefore more likely to be perceived as unfair. Datasets that have been historically biased, asymmetric, or non-representative of a broader population, or that do not have existing relevant data points for new consumer profiles, might not help calculate risk fairly.

### Differentiation and discrimination

Insurers must treat consumers fairly and equitably. This raises a number of considerations in the context of richer data:

- Individuals with particular characteristics might generally be at higher risk of accidents and therefore be appropriate candidates for differential pricing; but these characteristics might also pick out social groups protected from discrimination by equality legislation.
- Granular data might create new vulnerable social groups whose participation in the market becomes impaired as a result.
- As vehicles become more connected and more autonomous, there is potential for benefits to be experienced by individuals currently disadvantaged in traditional automobile insurance markets.
- The combination of the availability of more granular driver data and vehicles' enhanced safety systems, could allow traditionally higher-risk customers to secure more competitive insurance premiums.
- When levels 4 – 5 autonomy are achieved, the benefits could extend to groups previously disenfranchised in transport systems, such as people with impaired vision or other disabilities, who would be able to have parity with drivers without disabilities.

### New ethical challenges in the context of connected and autonomous vehicles

The 'trolley problem' is an ethical dilemma where an agent is put in control of a trolley that is about to run over a group of individuals and must decide whether to divert the trolley to a parallel track where it will run over only one individual. People respond to the challenge presented by the trolley problem differently, and there is no consensus on the correct answer. In the context of autonomous vehicles, a vehicle of level 4 – 5 autonomy might be required to conform with a pre-determined and consistent set of priorities for its decision-making (potentially varying according to legal jurisdictions and their associated insurance practices). One possible implication of this is that for fully-connected and autonomous vehicles, more standardised responses to the trolley problem might start to emerge.

## Transport data for social good

Recent research by the British Standards Institution (BSI) and the TSC examined the major barriers and challenges to the deployment of connected and autonomous vehicles<sup>15</sup>. One area of focus was promoting the sharing of safety-critical data, so that there is greater public acceptance of the new technology and more confidence in it because it is understood to be low-risk. Workshop participants explored the significant potential for advanced data capture and analysis in the automobile and automobile insurance sectors to contribute to improving driving and road safety.

## Driver safety and vehicle safety

There are good reasons for sharing non-safety-critical data at lower levels of vehicle autonomy and connectedness. For example, access to granular data about vehicles and drivers may contribute to improving road safety for identifying when a vehicle is beginning to develop a fault, without waiting for the annual MOT. This granular data could also distinguish between drivers who are proactive about addressing faults in their vehicles and help reduce their insurance premiums.

- Although there is an understanding that insurance companies are comfortable trying to nudge customers towards reducing their risks, it is unclear whether they have further responsibilities about other kinds of harmful behaviour, such as potential criminal activity. There might be unintended consequences in how customers react to try to inhibit data collection, such as jamming GPS signals.
- However, smaller businesses such as independent workshops and garages are at risk of being frozen out of the market if they are unable to access proprietary data about vehicle performance. If automobile manufacturers charge for access to this data, and independent garages pass this charge on to local customers, then customers might be deterred from having their vehicles repaired.

# Ensuring transparent, accountable and inclusive decision-making in data governance trade-offs

## Trust and data integrity

The workshop considered ways in which data-enabled digital disruption in the automobile, insurance and automobile insurance sectors presented new challenges and opportunities for decision-making in data governance trade-offs. Particular issues are the balance between privacy and transparency, the role of accountability and data privacy impact assessments, and the increasing importance of understanding and standardising or quantifying the value and provenance of data.

### Privacy and transparency

#### Privacy and personalisation

Consumers are demonstrating more nuance in their expectations of privacy – being more concerned with the impact of specific instances of data collection and analysis on them rather than having a blanket attitude of privacy as a default. Consumer expectations around privacy are also likely to be weighed against benefits, such as personalisation of service. This has a number of repercussions:

- Personalisation, with appropriate levels of granularity, could help drivers make more informed choices to improve their experience of driving or to reduce their insurance premiums. For example, a driver could input their intended route or destination into a device and receive a series of information points on costs or risks, such as congestion charging and thus have the option to modify their route. This should be done while preserving privacy – there are emerging technologies that allow fine differentiation at the user level but that do not allow the service provider to access the data at the same level of granularity. For example, a vehicle might measure usage by the minute, but the vehicle manufacturer or insurer might only be able to access charging information by the month.
- The volume of data generated could make it more difficult for consumers to make informed choices when exercising their data rights under the GDPR. Additionally, the range of devices, sensors, organisations and modelling techniques involved in

data capture and analysis makes it very difficult for any individual consumer to maintain perfect privacy.

- Privacy can become expensive for consumers if not sharing data results in higher insurance premiums for them. Due consideration must be given to the ethical implications of this. There might need to be a category of insurance available for consumers with high privacy requirements who do not wish to provide or share data, with a role for a regulator to cap prices to maintain equitable premiums for these consumers. The challenge to insurers will be differentiating between customers with high privacy requirements as a personal lifestyle choice, versus those customers opting for high privacy to conceal their risky profiles.
- The Royal Society's work on Privacy Enhancing Technologies has relevance here<sup>16</sup>: this is a category of emerging technologies and approaches that have the potential to preserve data privacy while allowing sophisticated analysis on the data. Privacy can be built in through design and data management principles such as 'highly personal data never leaves the car.'

### Transparency

The use of sophisticated algorithms, made possible by advanced data capture and analytics, raises issues of transparency about the data use and decisions made:

- There is a risk of 'invisible processing': individuals might not know that data analysis is being carried out, or how they might be affected. However, GDPR has transparency provisions that require that people be informed if automated decision making and profiling is taking place.
- This legislation has provisions that prohibit pure automated decision-making on something that might have a significant effect on individuals. It also has some wider-ranging exceptions that include safeguards, such as the right to obtain human intervention, the right to request that an automated decision is reconsidered by humans and the right to contest that decision.

- Unfortunately, because many algorithms work in a way that is not transparent, it can be difficult for a human to provide an explanation for the outcome of an automated decision.
- In terms of both hardware and software, it is becoming increasingly difficult to know what data-enabled devices are in a car and what kinds of data-capture and data analysis is taking place.
- This is compounded by the circulation of vehicles in the second-hand market, the importance of the retrofitting market in augmenting vehicle capabilities and owner variation when it comes to activating or updating hardware and software. Consequently, it can be difficult to determine a vehicle's capability at a single moment or in a given situation.

### Accountability, impact assessments and demonstrating compliance

The principle of accountability in the GDPR means that data controllers need to be able demonstrate their compliance with the legislation. Demonstrable compliance can be an asset in a business's relationship with its customers, with the regulator and with civil society, as well as with other businesses.

- One measure or tool for compliance that is provided for within the legislation is industry-led, enforceable codes of conduct; the insurance sector could be an area where there is potential to introduce this.
- GDPR legislation also provides for data protection impact assessments, which assess the impact on the privacy and rights of data subjects of a data controller processing their data; this impact assessment must be done before the data processing occurs.
- Prior to the introduction of the GDPR, the Information Commissioner's Office (ICO) had published some guidance on privacy impact assessments in the context of advanced data capture and analysis, drawing on input from industry expertise including the insurance sector. It was argued that the guidance remains largely valid today.

### Data value and data provenance

The sharing of data is key to new business models in both the automobile sector and the insurance and automobile insurance sectors; it is also key to understanding road transport safety. However, it depends crucially on the ability to assess the value, and the quality, of data.

### Data quality and data infrastructure

The importance of data quality standards, and a framework for defining what kind of data needs to be collected, was highlighted for the following reasons:

- There is significant variation in how data is recorded across the claims databases of different insurers. Consequently, the historical tagging of insurance claims data is probably not fit for purpose for contributing to analyses of how many past accidents a new technology could have prevented or mitigated against. This makes it harder to build an evidence base for the introduction or regulation of new technologies in the sector.
- There is significant data fragmentation even within an individual insurer's database, because of legacy IT systems. This might be inhibiting insurers' capacity for leveraging their own data assets.
- The data being produced through retrofitting hardware is often higher quality than the data being produced by embedded data-capture systems in new vehicles.
- There is not yet consensus on the balance to be found around open standards for data: some service providers might see the provision of data as their chief differentiator in the market and so open data standards would represent a loss of their market niche or competitive advantage.

### Data markets and data value

Because of the potential consumer and societal benefits, it would be beneficial for there to be an understanding of the value exchange of data-sharing and mechanisms developed for putting both a societal and a financial value on the data and its exchange<sup>17</sup>. Insurance companies are already buying up significant data on individuals: it may appear to an individual that only a small amount of data is being sought from them, for example, only their name and address, but insurance companies can compare and augment that with data they buy elsewhere. Current regulation, such as GDPR, can underpin some of that exchange with rules around data portability; data trusts for mutual benefit and mutual sharing of value might also have an important role to play.

### Data trusts and data rights

Data markets depend on understanding data rights. The relevance of a concept such as ‘property rights’ as assigned to data raises a number of questions. Most things that have property rights are not reproducible, like property; but data is easily reproducible, making data a challenge for property rights frameworks. Ownership rights for data are another model that is being explored, akin to copyright rights<sup>18</sup>. But currently there is ambiguity about whether data about a driver is owned by the individual who generated it or by the company that produced the device that collected the data.

Industry-level data trusts might be a mechanism of exchanging data, in terms of either the legal framework, mechanisms to build trust, or a culture in an industry that is not already inclined to information exchange. Data trusts might also play an important role in determining what kind of regulation would be possible and effective, as well as agile enough to respond to the emerging technology in this field, without stifling innovation.

# Annex A

## Summary tables of key arguments from *Data management and use: governance in the 21st century*<sup>18</sup>

TABLE 1

Tensions and disconnects in data management and use.

TENSIONS	Using data to improve offerings...	...without limiting available information or choices.
	Promoting benefits fairly across society...	...while ensuring acceptable risk for individuals, communities and organisations.
	Promote innovation...	...while addressing societal needs and reflecting public interest.
DISCONNECT	Existing data governance concepts, such as 'privacy', 'ownership', and 'consent', are under strain.	Cause: the traditional data lifecycle (collection, processing, application) is no longer linear because of 'open networks of data' with interconnected and interdependent data lifecycles. Data collection and data use are harder to separate; non-sensitive data can hold sensitive insights; and data provenance can be unclear because of weak audit trails of meta-data and data trading and selling.

TABLE 2

Principles for data governance.

PRINCIPLE	ARGUMENT
<p><b>Overarching principle:</b> Promote human flourishing.</p>	<p>Human flourishing is multi-dimensional, dynamic and context-specific. It includes concepts such as wellbeing and the need for individuals and communities to prosper. At moments of contention, the principle should serve to reflect the fundamental tenet that society does not serve data but that data should be used to serve human communities.</p>
<p><b>Complementary principle:</b> Protect individual and collective rights and interests.</p>	<p>Data governance should offer meaningful and effective protection against both tangible and intangible harms, such as discriminatory treatment or exclusion from opportunities respectively. It should protect both individual rights, goods and benefits, such as health; and collective rights, goods and benefits such as the environment.</p>
<p><b>Complementary principle:</b> Trade-offs between data management and use are transparent, accountable, and inclusive.</p>	<p>Effective data governance must identify competing considerations and balance them. To achieve this, decision-making must be multi-stakeholder and if necessary iterative. It should also be recognised that relevant expertise might come from both traditional and non-traditional perspectives, backgrounds and approaches. Finally, transparency alone is not sufficient: it must be accompanied by accountability.</p>
<p><b>Complementary principle:</b> Enhance existing democratic governance.</p>	<p>Data governance and data use should support democratic processes, help enact democratic decisions and be subject to democratic oversight. There should be consistency and proportionality in governance frameworks and mechanisms, and appropriate balance between competing interests. Finally, enforcement powers and resources should be appropriate for achieving regulatory aims.</p>
<p><b>Complementary principle:</b> Seek out good practice, and learn from success and failure.</p>	<p>Effective data governance should display a commitment to promoting good practice and embedding continuous learning as a way of improving practices and standards.</p>

# Annex B

This report is a summary of the workshop *AI and data governance from principles to practice: auto insurance* held at the Royal Society, 6 – 9 Carlton House Terrace, London, on 20 September 2018, in partnership with The Alan Turing Institute, the Leverhulme Centre for the Future of Intelligence and the Royal Academy of Engineering.

## Workshop speakers

### Workshop Chair

#### **Dr Adrian Weller**

Programme Director for Artificial Intelligence,  
The Alan Turing Institute

### Workshop panellists (alphabetical by surname)

#### **Maria Axente**

PwC UK

#### **Professor Andrew Blake FREng FRS**

FiveAI

#### **Roger Bickerstaff**

Bird & Bird

#### **Glen Clarke**

Allianz

#### **James Edgar**

Which?

#### **Bruno Fernandez-Ruiz**

getnexar

#### **Nick Fleming**

British Standards Institution

#### **Toby Fountain**

Department for Transport

#### **Ed Klinger Leon**

Flock

#### **Dr Natasha McCarthy**

Head of Policy (Data), The Royal Society

#### **Charles Smith**

Collision Management Systems

#### **Antony Walker**

Deputy CEO, techUK

#### **Carl Wiper**

Information Commissioner's Office

#### **Steve Yianni FREng**

The Royal Academy of Engineering

## Acknowledgements

The Alan Turing Institute

The Leverhulme Centre for the Future of Intelligence

The Royal Academy of Engineering

## Staff at the Royal Society

### **Louise Pakseresht**

Senior Policy Adviser (Data)

### **Ellen Phillipson**

Project Coordinator (Education and Data)

### **Dr Mahlet Zimeta**

Senior Policy Adviser (Data)

# References

1. *Data management and use: governance in the 21st century*, June 2017, the Royal Society and the British Academy (see <https://royalsociety.org/topics-policy/projects/data-governance/>, accessed 8 August 2019)
2. *Data management and use: governance in the 21st century*
3. This table is adapted from *Connected and autonomous vehicles: the UK economic opportunity*, March 2015, the Society of Motor Manufacturers and Traders (see <https://www.smm.co.uk/2015/03/connected-and-autonomous-vehicles-the-uk-economic-opportunity/>, accessed 8 August 2019)
4. *Market forecast for connected and autonomous vehicles*, July 2017, the Transport Systems Catapult (see <https://ts.catapult.org.uk/news-events-gallery/news/connected-autonomous-vehicle-industry-worth-28bn-uk-2035/>, accessed 8 August 2019)
5. *Connected and autonomous vehicles: the future? The second report of session 2016-17*, House of Lords Science and Technology Select Committee (see <https://www.parliament.uk/autonomous-vehicles>, accessed 8 August 2019)
6. See, for example, *The future of insurance is happening without insurance firms*, 20 July 2019, *The Economist* (available at <https://www.economist.com/finance-and-economics/2019/07/20/the-future-of-insurance-is-happening-without-insurance-firms>, accessed 8 August 2019)
7. *EU rules on gender-neutral pricing in insurance industry enter into force*, 20 December 2012, the European Commission (see [http://europa.eu/rapid/press-release\\_IP-12-1430\\_en.htm](http://europa.eu/rapid/press-release_IP-12-1430_en.htm), accessed 8 August 2019)
8. *How an EU gender equality ruling widened inequality*, 14 January 2017, *The Guardian* (see <https://www.theguardian.com/money/blog/2017/jan/14/eu-gender-ruling-car-insurance-inequality-worse>, accessed 8 August 2019)
9. Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases (see <https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31996L0009>, accessed 8 August 2019)
10. *EU data protection rules*, the European Commission (see [https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules\\_en](https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en), accessed 8 August 2019)
11. See *Road accidents and safety statistics*, Department for Transport (available at <https://www.gov.uk/government/collections/road-accidents-and-safety-statistics#history>, accessed 8 August 2019)
12. *Towards an accident investigation branch for roads*, December 2017, the RAC Foundation (see <https://www.racfoundation.org/research/safety/towards-an-accident-investigation-branch-for-roads>, accessed 8 August 2019); and *Crash investigation stepped up a gear*, June 2018, the RAC Foundation, (see <https://www.racfoundation.org/media-centre/crash-investigation-stepped-up-a-gear>, accessed 8 August 2019)
13. Rail Accident Investigation Branch (RAIB) (see <https://www.gov.uk/government/organisations/rail-accident-investigation-branch>, accessed 8 August 2019); Marine Accident Investigation Branch (MAIB) (see <https://www.gov.uk/government/organisations/marine-accident-investigation-branch>, accessed 8 August 2019); Air Accidents Investigation Branch (AAIB) (see <https://www.gov.uk/government/organisations/air-accidents-investigation-branch>, accessed 8 August 2019)
14. *Motoring of the future: the eighth report of session 2014-15*, House of Commons Transport Committee (see <https://www.parliament.uk/business/committees/committees-a-z/commons-select/transport-committee/inquiries/parliament-2010/motoring-of-the-future/>, accessed 8 August 2019)
15. *Connected and autonomous vehicles: A UK standards strategy*, March 2017, the British Standards Institution (see <https://www.bsigroup.com/en-GB/Innovation/cav/>, accessed 8 August 2019)
16. *Protecting privacy in practice: the current use, development and limits of Privacy Enhancing Technologies in data analysis*, April 2019, the Royal Society (see <https://royalsociety.org/topics-policy/projects/privacy-enhancing-technologies/>, accessed 8 August 2019)
17. See, for example, *Towards trusted data sharing: guidance and use cases*, December 2018, the Royal Academy of Engineering (available at <http://reports.raeng.org.uk/datasharing/cover/>, accessed 8 August 2019)
18. See, for example, *Data ownership, rights, and controls: seminar report, 2018*, The British Academy, The Royal Society and techUK (available at: <http://www.thebritishacademy.ac.uk/publications/data-ownership-rights-controls-seminar-report/>, accessed on 21 August 2019)
19. *Data management and use: governance in the 21st century*



The Royal Society is a self-governing Fellowship of many of the world's most distinguished scientists drawn from all areas of science, engineering, and medicine. The Society's fundamental purpose, as it has been since its foundation in 1660, is to recognise, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity.

The Society's strategic priorities emphasise its commitment to the highest quality science, to curiosity-driven research, and to the development and use of science for the benefit of society.

These priorities are:

- Promoting excellence in science
- Supporting international collaboration
- Demonstrating the importance of science to everyone

**For further information**

The Royal Society  
6 – 9 Carlton House Terrace  
London SW1Y 5AG

**T** +44 20 7451 2500

**W** [royalsociety.org](http://royalsociety.org)

Registered Charity No 207043

Issued: August 2019 DES6485