OECD publishing

DATA IN AN EVOLVING TECHNOLOGICAL LANDSCAPE

THE CASE OF CONNECTED AND AUTOMATED VEHICLES

OECD DIGITAL ECONOMY PAPERS
December 2022 **No. 346**

OECD
BETTER POLICIES FOR BETTER LIVES

Foreword

Digital technologies underpin the creation, generation, collection, transfer and use of data, and digital technological development and deployment shape data governance policy debates. This report explores data governance in an evolving technological landscape, and offers recommendations to ensure policies remain resilient to technological change over time.

This report was drafted by Angela Attrey under the guidance of Gallia Daor and the supervision of Audrey Plonk, Head of the OECD Digital Economy Policy Division. The feedback and insights of Luis Aranda, Philippe Crist, Alexia Gonzalez Fanfalone, Andras Hlács, Karine Perset, Christian Reimsbach-Kounatze, Maximilian Reisch and Verena Weber is gratefully acknowledged. Mark Foss and Angela Gosmann provided editorial support. This publication is a contribution to IOR 1.3.1.2.3 of the 2021-2022 Programme of Work and Budget of the Committee on Digital Economy Policy.

This report was approved and declassified by the Committee on Digital Economy Policy on 27 September 2022 and prepared for publication by the OECD Secretariat.

This publication is a contribution to Phase III of the OECD Going Digital project, which aims to provide policy makers with the tools they need to design and implement better data policies to promote growth and well-being. For more information, visit www.oecd.org/going-digital.

#GoingDigital

Note to Delegations:

This document is also available on O.N.E under the reference code:

DSTI/CDEP/GD(2022)3/FINAL

Table of contents

Foreword 2

Executive summary 4

1 Introduction 6

2 Digital technologies underpin data and are relevant to data governance policy 8
2.1 An introduction to digital technologies that underpin the use of data 8
2.2 Digital technologies shape the data governance landscape 11
2.3 Governing digital technologies: The example of Artificial Intelligence 14

3 Connected and automated vehicles: A case study of data governance issues 15
3.1 An introduction to connected and automated vehicles 15
3.2 The emerging data landscape of connected and automated vehicles 17
3.3 Emerging approaches to data governance for connected and automated vehicles 20

4 Policy considerations for data governance in a fast-evolving digital technology landscape 24
Policy considerations 24

References 27

FIGURES

Figure 2.1. M2M/embedded mobile cellular subscriptions, December 2021 11
Figure 2.2. Mobile data usage per mobile broadband subscription per month, 2021 12

Executive summary

Insights

Data depend on digital technologies

Data are often compared to other resources, like oil or air. Like air, data are non-rivalrous, while like oil, data are a strategic asset that can be leveraged for economic advantage. However, comparisons to natural resources belie a fundamental characteristic of digital data: they are not naturally occurring. Rather, data in digital formats depend on digital technologies for their creation, generation, collection, transfer and use.

Digital technologies have shaped the policy issues at the heart of data governance discussions

Digital technological development and deployment shape data governance policy debates. The increasing uptake of connected devices, and their growing technical capacity, mean that data are being generated and collected in more and more economic and social contexts. This, in turn, is fuelling growing concerns about privacy and data protection. Digital technologies like data analytics and artificial intelligence (AI) are also often essential for putting data to productive use, and can underlay data's potential competitive advantage. Advances in broadband networks and cloud computing further shape debates on how data should flow, including across borders, and where data are processed and stored.

The example of connected and automated vehicles illustrates how technological development can raise different concerns for data governance

Connected and automated vehicles combine many digital technologies in their operation. They are a part of the Internet of Things, feature many forms of AI-enabled automation, and use short and long-range communication technologies to connect and share data. Connected and automated vehicles are also rapidly advancing as firms adopt different digital technologies in vehicle design. This, in turn, determines how that data will be collected, transferred and analysed to enable technical function. Already, a growing and complex range of actors is involved in collecting, processing and storing data from connected and automated vehicles, much of which is likely to be personal. Meanwhile, policy approaches are still nascent. These factors make connected and automated vehicles an illustrative case study for how technological development can raise different concerns for data governance.

Data governance efforts should be technologically neutral, agile and principles-based

As digital technologies evolve, so too will concerns for data governance. As the case study on connected and automated vehicles demonstrates, digital technologies are evolving rapidly and in ways that are often hard to predict. Data governance policies should therefore remain technologically neutral and sufficiently flexible to remain responsive to technological change. At the same time, they should provide sufficient certainty to enable technological development.

Policy considerations

Keep data governance policies agile and principles-based

In view of the evolving nature of data collection and processing, policies that seek to govern data should aim to remain agile. They should aim not to constrain evolving business models, innovation and competition, while balancing the rights and interests of other parties. The OECD Privacy Guidelines provide an example for data governance policies that are resilient to regulatory obsolescence over time.

Keep data governance policies technologically neutral

Digital technologies determining how data are used are evolving. Therefore, data governance policies should seek to remain technologically neutral. They should govern whether and how data are shared rather than the technology through which the data are exchanged.

Adapt regulations as needed to digital technology developments

Ensuring that regulation remain technology-neutral implies regular review and adaptation to consider digital technology developments. Outcome-based or risk-based regulation could outline outcomes to be achieved but not specify process or a particular means of compliance. Another approach includes the use of regulatory sandboxes, which help enable testing of digitally enabled innovations with fewer regulatory constraints.

Consider horizontal and sectoral data governance policies together

As technological applications are highly heterogeneous, horizontal policies may require complementary sector-specific regulation after identification of gaps in regulation and legislation. General data governance standards can be followed with sector-specific regulations and requirements for data-intensive sectors or areas after gaps are identified. Such an effort, in turn, implies monitoring horizontal data governance policies in various economic and social applications.

Target uncertainties in data governance regulations that can hinder technological development

Data governance policies should seek to maximise the benefits of data use while protecting rights and interests and addressing any related risks and challenges. Horizontal regulations can help set the rules of the road. However, policies should seek to avoid creating uncertainties. They should also clarify any gaps where necessary, particularly where such uncertainties may impact technological development.

1 Introduction

Data have emerged as a new strategic resource for firms, governments and societies, and underlie an increasing number of economic and social activities. However, unlike other strategic, useful economic assets, data do not occur naturally. Instead, they depend on digital technologies for generation, collection, transfer, storage and use. This section introduces the report by first describing its understanding of the terms “data” and “data governance”. It then outlines subsequent sections that will examine, among other issues, the relationship between data and digital technologies in the context of connected and automated vehicles.

The emergence of data governance as a policy issue is directly linked to technological advancement and deployment. As sensor-laden devices become more ubiquitous, more interactions and transactions create flows of data from more economic and social phenomena. Digital technologies like artificial intelligence (AI) enable these data to be processed to generate insights and predictions and yield value for users. How data are stored, processed and managed also depend on advances in digital technologies – from cloud computing to broadband networks. Moreover, these digital technologies are not static: they continue to advance and evolve, further re-shaping how data landscape into the future.

Data and digital technologies are therefore fundamentally linked. Policies related to data and data governance are emerging across policy domains, from health to trade and finance. Often, however, they do not fully account for the relationship between data and the digital technologies. In view of this relationship, efforts to govern data should consider digital technologies and their development.

Box 1.1. What are data? What is data governance?

For the purposes of this report, “data” refer to recorded information in structured or unstructured formats, including text, images, sound and video. Data can be in any format, including analogue formats like paper, or emerging quantum forms like qubits. However, the rise of digital technologies has enabled the growth and policy relevance of digital data – information stored by a computer in binary format. Throughout this report, the word “data” means digital data, unless otherwise noted.

In the context of the OECD Project on Data Governance for Growth and Well-being, ”data governance” refers to diverse arrangements, including technical, policy, regulatory and institutional provisions, that affect data and their creation, collection, storage, use, protection, access, sharing and deletion, including across policy domains and organisational and national borders. Efforts to govern data take many forms. They often seek to maximise the benefits from data, while addressing related risks and challenges, including to rights and interests.

Source: OECD (2022[1]), *Going Digital to Advance Data Governance for Growth and Well-being*, OECD Publishing, Paris,

This report explores the relationship between data and digital technologies, particularly in the context of connected and automated vehicles. It also provides considerations for data governance policies in an evolving technological landscape. Section 2 of this report illustrates the interlinkage of data governance and digital technologies and examines the digital technologies that facilitate the generation, collection, transfer, use and storage of data, and how they have affected the data governance policy issues. Section 3 illustrates the interlinkage of data governance and digital technologies through a case study of connected and automated vehicles, and explores how data governance issues and approaches are emerging in the automotive context. Section 4 concludes with outlining considerations for data governance policies in an evolving technological landscape.

2 Digital technologies underpin data and are relevant to data governance policy

Why consider digital technologies in the context of data governance? This section aims to answer that question by first introducing digital technologies that generate, transfer, collect, store and use data. It then outlines how these digital technologies have shaped the policy issues underlying data governance discussions. Finally, because these digital technologies that underpin data are evolving, it ends with a brief discussion of approaches to the governance of one such technology: artificial intelligence.

2.1 An introduction to digital technologies that underpin the use of data

An ecosystem of digital technologies underpins the productive use of data across economies and societies. OECD research has identified the main digital technologies that underpin digital transformation (OECD, 2019[2]; OECD, 2015[3]) and examined emerging technologies like quantum computing (OECD, 2020[4]). Accordingly, this section focuses on how select digital technologies relate to data and their generation, transfer, collection, storage and use. Some of the discussion shows how digital technologies tend to be used in combination, e.g. the Internet of Things and cloud computing, requiring a comprehensive understanding of how they function.

The Internet of Things

Objects equipped with sensors increasingly proliferate throughout different parts of economic and social life across the OECD. Such devices include familiar objects/gadgets, like the smartphone or laptops. They also encompass other connected devices like smart meters for energy management or industrial robots. Connected devices are broadly referred to as the "Internet of Things" (IoT). The OECD defines the IoT in broad terms, including "all devices and objects whose state can be altered via the Internet, with or without the active involvement of individuals" (OECD, 2018[5]).

The IoT enables the collection, storage and sharing of data about the contexts where the connected devices are located (OECD, forthcoming[6]). The proliferation of the IoT has enabled greater and more varied data collection. A smart meter, for example, collects data on the consumption of electric energy. It can then help consumers better understand their consumption behaviour, and enable providers to better monitor energy demand. Importantly, some critical IoT applications, like remote surgery or connected cars, require a certain quality and speed of data transfers. They thus rely on fast, reliable and low-latency connectivity. These requirements can dictate how data from the IoT are processed, i.e. in a data centre, on the "edge" of the network or in the device itself (OECD, forthcoming[6]). On the other hand, the IoT can span more basic devices, including those that do not rely on high-capacity networks and do not send data often.

Data analytics and artificial intelligence

Data generate much of their value from processing to extract useful insights. Advances in analytical digital technologies enable sophisticated processing that extracts greater value from data. In particular, AI has emerged as a general purpose technology. Its statistical and probabilistic algorithms enable powerful insights and predictions, across a variety of applications, from health care to public services (OECD, 2019[7]).

An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy (OECD, 2019[8]). They use machine and/or human-based data and inputs to (i) perceive real and/or virtual environments; (ii) abstract these perceptions into models through analysis in an automated manner (e.g. with machine learning) or manually; and (iii) use model inference to formulate options for outcomes. AI systems are designed to operate with varying levels of autonomy (OECD, 2019[7]).

Modern and increasingly statistical and probabilistic AI systems use data as an input. Core characteristics of data that can be used to classify AI models include provenance, collection and origin, domain, quality and technical character, including their structure and code (OECD, 2022[9]). As an enabler of AI systems, data can shape the outputs of AI systems and their evolution. Biased data can result in systems delivering biased outcomes (OECD, UNESCO, IDB, 2022[10]).

Broadband networks

High-quality broadband networks with adequate speeds and network response times (low latency) are a prerequisite for data to be transferred at high speed. As more applications become data-driven, from online education to telemedicine, networks are evolving to enable the connection of billions of devices and sensors to the Internet (OECD, 2019[11]). Data consumption increases as next generation networks are deployed across the OECD and support "data-intensive" applications like videoconferencing, streaming and gaming (OECD, 2022[12]).

Broadband networks are also becoming increasingly "virtualised", with software decoupling network functions from hardware appliances (OECD, 2022[12]). This makes it technically possible to shift computing

resources to the "edge" of the network to reduce latency, which enables the provision of edge computing (see below). This is one example of how digital technologies are increasingly converging in their applications.

Cloud and edge computing

Cloud computing can be defined as a "service model for computing services based on a set of computing resources that can be accessed in a flexible, elastic, on-demand way with low management effort" (OECD, 2014[13]). Users of cloud computing services can access computing resources, including storage, software, capacity and networking, over the Internet. This enables customers to transform substantial fixed costs for information and communication technologies (ICTs) into lower marginal costs. At the same time, they can more easily match their purchase of computing resources with their evolving needs. Cloud computing services therefore increase the affordability and availability of computing resources and facilitate the use of other digital technologies like AI, as well as the wider digital transformation (OECD, 2019[2]).

Cloud computing enables the processing and storage of data to be geographically separated from its nexus of production, supporting the provision of cross-border data-driven services. While cloud computing services have a number of different service models, the current dominant model is the public cloud. In this model, data and computing services are provided for a fee by cloud service providers. To date, the provision of public cloud services has involved the transfer of data across wireless networks to public cloud data infrastructure, notably data centres, where data are centralised and analysed. In recent years, much of global data processing and storage has taken place in centralised data centres operated by cloud service providers (CISCO, 2018[14]).

However, as broadband networks are becoming increasingly virtualised (see above), it becomes technically possible to shift computing resources to the "edge" of the network to reduce latency. This could enable new cloud service models, such as mobile network operators hosting edge servers running cloud-based applications (OECD, 2022[12]). Multi-access edge computing is a computing architecture that enables cloud computing capabilities and ICT service environments at the edge of the network (European Technical Standards Institute, 2020[15]). Edge computing enables a shift of data processing and traffic away from centralised clouds. By moving towards the edges of the network, closer to the client or edge device, edge computing reduces latency and increases the performance of high bandwidth applications (OECD, 2022[12]). Some authors also note that edge computing could support resilience, security and privacy protection (IBM, 2021[16]).

Blockchain and distributed ledger technologies

Distributed ledger technologies (DLTs) are a combination of technologies that together create a digital, shared and self-updating ledger of verified transactions or information among parties in a network. DLTs use various types of multi-party consensus mechanisms to validate and record transactions. It has multiple governance systems ranging from "centralised" models through to "decentralised" models that may not be controlled by a central authority(ies) (OECD, 2022[17]).

Thanks to their decentralised nature and independent, immutable verification, DLTs could transform data management. DLTs enable a group of networked databases to hold the same set of data and cannot be changed without consensus. For example, the use of DLTs could be used for personal identification management without needing to divulge personal information ("zero-knowledge proofs"). Under such a model, DLTs would store and transmit encrypted digital attestations rather than the underlying personal data. Some jurisdictions across the OECD are exploring this and other applications of blockchain, although technical and logistical barriers have stymied their widespread uptake (OECD, 2022[18]).

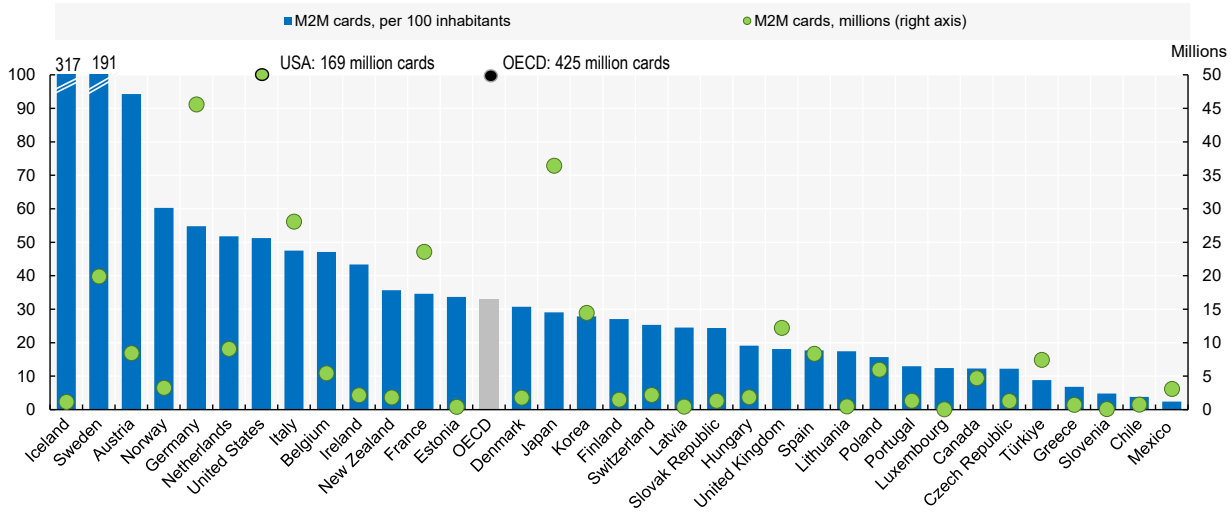
2.2 Digital technologies shape the data governance landscape

The development and deployment of digital technologies in the previous section have transformed the way data are generated, collected, transferred, used and stored. As will be discussed in this sub-section, they have equally affected issues at the heart of policy debates on data governance.

Data generation and collection

Data are generated wherever digital technologies are deployed – namely, in every part of economic and social life across the OECD and beyond. Naturally, the diffusion of digital technologies has led to more data generation and collection. This trend is driven by the penetration of devices like smartphones and the rise of the IoT. CISCO estimates there will be 29.3 billion networked devices, approximately 3.6 devices per capita, by 2023. Approximately half of these connections are expected to be Machine-to-Machine (M2M) connections in 2023, up from 33% in 2018 (CISCO, 2021[19]). Each of these devices is an endpoint that can generate, collect and sometimes store a wide variety of data. M2M-embedded mobile data subscriptions, which are necessary to enable communications for a subset of the IoT, grew by more than 16% over 2021 for countries for which data are available. This indicates strong growth in the IoT overall (see Figure 2.1).

Figure 2.1. M2M/embedded mobile cellular subscriptions, December 2021



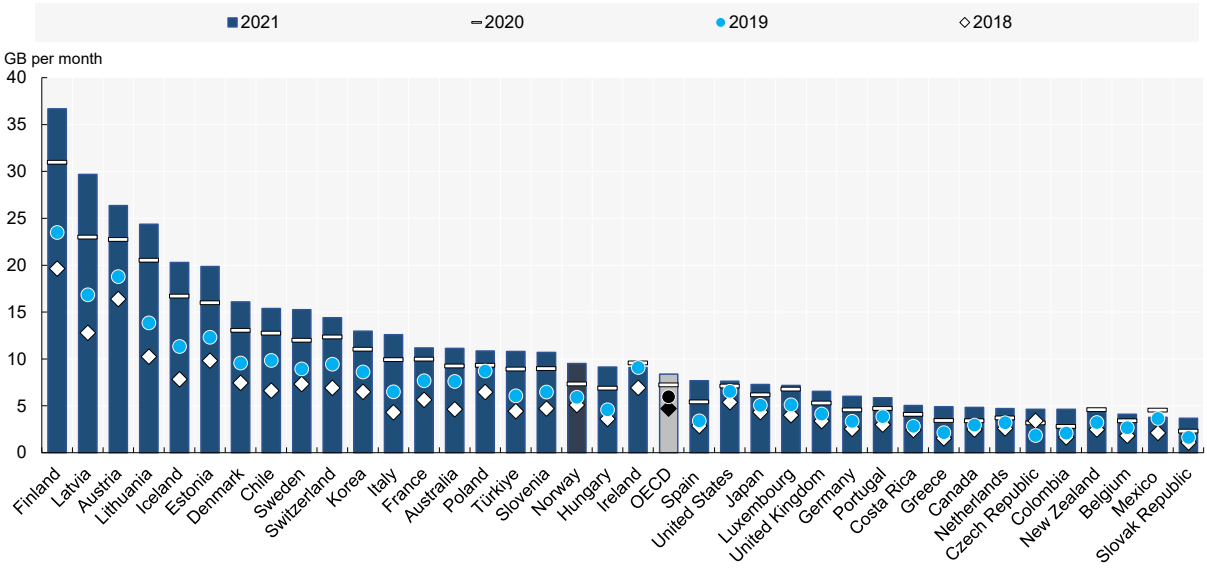
Note: The significant growth of M2M in Iceland is due to the provision by Vodafone Iceland of M2M subscriptions for the benefit of international pharmaceutical companies to manage the transport of COVID 19 vaccines. Most cards from Sweden are provided by a Swedish operator but used in other countries.
Source: OECD (2022[20]) *Broadband Portal* (database), http://www.oecd.org/digital/broadband/broadband-statistics/ (accessed 21 July 2022).

Data transfer and transmission

Next generation wireless networks are necessary to cope with the increasing demand (OECD, 2019[11]) for low-latency upload and download of large streams of data, as well as to accommodate a higher density of connected devices. As networks continue to be deployed across the OECD, the scale and speed of data transfers is likely to increase massively. The deployment will also likely facilitate a broader shift towards more distributed high-bandwidth activities like video streaming (CISCO, 2021[19]). Already, across the OECD, mobile data usage per subscription rose by 15% in 2021. The increase was less than in 2020, but still represented a rise of 79% over the three years until the end of 2021 (OECD, 2022[21]). The amount of

data consumed averages 8.4 GB per OECD subscription per month but varies greatly by country (see Figure 2.2).

Figure 2.2. Mobile data usage per mobile broadband subscription per month, 2021



Note: The multiplier 1024 is used to convert TB into GB; the total amount of GB is divided by the yearly average number of Mobile broadband subscriptions. Mexico and Switzerland: Data are preliminary. United States: Data for 2021 are temporary estimates. OECD average includes estimates.
Source: OECD (2022[20]) *Broadband Portal* (database), http://www.oecd.org/digital/broadband/broadband-statistics/ (accessed 21 July 2022).

Data types and re-identification of personal data

As more connected devices go online and as digital technologies become an increasing part of economic and social activities, organisations can obtain data from numerous sources. Data have thus taken on more "variety". This ranges from graphics, text, audio, video or sensor data (OECD, 2015[22]) to entirely machine-generated data like synthetic data or meta-data. Such variety may require distinct processing capabilities and specialised data analytics to deliver value, including through AI systems (OECD, 2019[7]).

The combination of the increasing variety and volume of data and advances in AI systems has rendered otherwise anonymised and pseudonymised data increasingly easily to (re-)identify as personal data (Rocher, Hendrickx and de Montjoye, 2019[23]). This is particularly true where datasets are, or are combined with, geo-local data, because personal mobility patterns are highly individual and heterogeneous (European Data Protection Board, 2021[24]). The adoption of next generation wireless networks, including hybrid WiFi and global navigation satellite system (GNSS) location services, could increase the precision of geo-local data and amplify potential privacy concerns regarding re-identification (ITF, 2022[25]). This implies that more data collected may fall within the purview of privacy and/or data protection regulation. The recent review of the OECD Privacy Guidelines recognised these issues. It noted the increasing risk of reidentification of personal data as the main concern for 68% of privacy enforcement authorities surveyed (OECD, 2021[26]).

Data processing and degree of distribution

As data volume increases, so too does the importance and complexity of the required data processing and storage architecture. Cloud computing separates the storage and processing of data from where they were

generated or collected. Cloud computing enables the distributed online delivery of data services. However, the data and infrastructure underlying cloud applications, like servers, are usually housed in a physical facility known as a data centre. In this model, data are transferred from where they are generated across wireless networks to a data centre, where they are centralised and analysed. Approximately 80% of the processing and analysis of data around the world takes place in data centres (CISCO, 2018[14]).

These technological developments affecting data processing feed into ongoing data governance debates, notably in the context of cross-border data flows. Data are non-rival, and cloud computing enables remote data processing. Therefore, cloud computing can lead to situations where the service provider and/or data centre, as well as the client, may be subject to conflicting legal obligations related to data protection. Relatedly, policies conditioning the flow of data across borders often focus on the geographic location of data processing and storage (Casalini, López González and Nemoto, 2021[27]). To meet regulatory requirements, and also to improve redundancy and reduce latency, cloud service providers locate data centres in a variety of jurisdictions (AWS, 2022[28]).

However, as digital technologies evolve, so too will the structure of data processing and computing architectures. Possibly, they will move from centralised cloud processing to local edge computing. A main reason for this shift is the practical need for lower latency. As smart devices are increasingly deployed in more critical and bandwidth-intensive applications, they are likely to require the transmission of ultra-high-bandwidth data with near-zero network response times. Many argue that such rapid response times for transmission of high volumes of data are impossible if data are sent to centralised data centres at a distance, even as next-generation wireless networks advance in capacity (AECC, 2021[29]; IBM, 2021[16]).

Advances in edge computing can enable data to be processed closer to, or within, IoT devices and reduce the need to send data to the cloud (IBM, 2021[16]). A distributed network of edge sites could enable local connections that reduce latency and the relevant costs of network traffic. Indeed, some estimate that up to 75% or more of data processing and analytics will run at the edge of the network or within devices by 2027 (European Commission, 2021[30]). This evolution in cloud computing would have the effect of making data processing more local. In this way, it might reduce the need for cross-border transfers of data, and with it, some related policy issues.

Data storage and retention

Digital data are stored in many forms, including magnetic storage, optical disks, flash storage and semi-conductor chips. Bits of information (namely, ones or zeros) can be stored in any material capable of displaying two distinct states. Therefore, digital data could be theoretically stored in other, non-electronic, formats, including in deoxyribonucleic acid (DNA) (Panda et al., 2018[31]). As the capacities of IoT become more advanced, more devices are likely to have some memory capacity that enables them to store data on board rather than relying on constant connectivity (CISCO, 2021[19]).

Much more data are generated than are actually stored. For example, vehicle routing information for connected cars is routinely cached (temporarily stored) and overwritten with new data. The costs of data have been declining for some time (OECD, 2022[32]). However, because data generation is increasing so quickly, this discrepancy between generation and storage is likely to continue . By one estimate, 2% of the additional data generated from the shift of activities on line during the COVID-19 pandemic in 2020 was saved and retained into 2021 (IDC, 2021[33]).

Where and how data are stored, and whether they flow or are required to flow across borders, are relevant to policy. Measures that seek to explicitly store at least one copy of data within domestic territories are on the rise across the world. Recent OECD work identified 92 “data localisation” measures across 39 countries, with at least half emerging over the last five years (López González, Casalini and Porras, 2022[34]). However, as demand for speedy data transmission increases, more data are expected to be cached at the edge of the network, reducing the need for cross-border data flows.

2.3 Governing digital technologies: The example of Artificial Intelligence

Digital technological innovation has been rapid, driving growth and well-being across the OECD, and shaped the policy issues underlying data governance discussions, as discussed in section 2.2. However, as technologies advance rapidly, they can raise new concerns for economies and societies. Indeed, the technologies themselves become the focus of governance.

The rapid advance in digital technologies can challenge traditional policy making, which is often purposefully process-driven and deliberative (OECD, 2019[35]). For example, advances in AI have exploded in recent years and hold much promise. However, the technology can lack transparency that challenges traditional accountability mechanisms and could propagate biases (OECD, 2019[7]). These challenges call for policy action that minimises risks but is also agile enough to foster continued research, innovation and technology diffusion.

To address this technology governance challenge, governments across the OECD are adopting principles-based approaches to governing AI. The OECD AI Principles, for example, were adopted in 2019 and subsequently formed the basis of the Group of Twenty's AI principles (OECD, 2019[8]). To date, OECD countries and eight non-OECD countries, including five low- and middle-income countries, adhere to the OECD AI Principles. These values-based principles aim to foster confidence in the adoption of trustworthy AI. They are also designed to be adaptable to technological developments. The OECD AI Principles are an example of upstream governance that can be later complemented by downstream elements such as regulation and technical standards if necessary (OECD, 2021[36]). Given rapidly evolving technology, understanding how principles are implemented is essential to shaping future downstream policy action. To this end, the OECD supports and tracks the implementation of the OECD AI Principles through its AI Policy Observatory. This covers hundreds of policy initiatives in more than 60 countries and territories, and the European Union.

3 Connected and automated vehicles: A case study of data governance issues

This section turns to connected and automated vehicles, which hold great potential to transform economies and societies. The chapter examines the emerging data landscape of connected and automated vehicles, examining how they collect vast and varied volumes of data that hold value in many applications. After exploring the data ecosystem, it identifies emerging approaches to data governance for these vehicles. To that end, it covers topics such as data collection and management at the vehicle level, location of data processing, access to in-vehicle data, models for data sharing and access, and finally, reporting and mandatory collection of data.

3.1 An introduction to connected and automated vehicles

Connected vehicles can connect to networks, like the Internet, to send and receive data with other networked devices, both inside and outside the vehicle. They communicate with other systems, including other vehicles, roadside infrastructure and third-party service providers.

Automated vehicles are vehicles that operate on a spectrum of declining input from the driver. Many aspects of vehicles manufactured today already involve some automation, including adaptive cruise control and active lane-keeping assistance. However, a fully automated, or autonomous or self-driving, vehicle

would include the "full-time performance by an automated driving system of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver" (SAE, 2014[37]).

Automated vehicles periodically connect to the Internet, e.g. to receive software updates. In some models of automated vehicle deployment, achieving increasing levels of automation is expected to require increasing levels of connectivity, as vehicles use short-range communication technologies to communicate with road-side infrastructure and other vehicles (OECD, 2018[5]).

Connected and automated vehicles highlight how different digital technologies increasingly converge in their application. Connected vehicles commonly use both cellular and non-cellular communication technologies to connect and share data with other vehicles, infrastructures, road services, satellites and other third parties. Given they are intelligent devices using sensors to collect data, connected and automated vehicles are part of the "Internet of Things" (OECD, 2018[5]). Connected vehicles on the market already display some forms of automation enabled by artificial intelligence (AI), including driver assistance systems (ITF, 2015[38]). As technologies continue to develop and vehicles become more automated, the use of data processing technologies like big data analytics and AI systems within vehicles are also expected to increase.

Box 3.1. Models of connected and automated vehicle deployment

Connected and automated vehicles hold great potential to transform economies and societies. For example, connected vehicles that can interact with other vehicles could "platoon". In other words, they would adjust their speed and trajectories to move together in a group, thereby reducing carbon dioxide emissions (Winder, 2016[39]). Connected vehicles could share road safety information and reduce traffic accidents. Fully autonomous vehicles have not been commercially deployed yet. However, many estimates cite the role of human error in road fatalities and note that road safety is likely to improve with vehicle automation (ITF, 2015[38]).

Connected and automated vehicles could also change the delivery of mobility itself. Mobility as a service (MaaS) is a model for supplying a wide range of passenger transport services through a single digital customer interface (ITF, 2021[40]). Connected and automated vehicles could fit into this vision through the widespread deployment of automated taxi fleets. This would enable more seamless delivery of transport services, potentially reducing the need for personal vehicle ownership and parking spaces near homes and workplaces. Some deployments of automated personal vehicles take the form of taxi fleets (Wayland, 2022[41]). Other forms of vehicular automation, like fully automated train systems (Alstom, 2022[42]) and autonomous tractors (OECD, 2022[43]), are evolving in other contexts.

The discussion in this paper refers specifically to the use of connected and automated vehicles for personal, non-professional use, and the governance of data collected by the vehicle and/or exchanged between the vehicle and other machines and actors, unless otherwise specified. Many of the considerations in this paper are relevant to data governance questions in professional, MaaS and shared and public transport contexts. However, those topics may also raise additional issues or considerations that are out of scope for this paper. This paper uses the terms "cars" and "vehicles" interchangeably.

Cars are increasingly connected: the connected car is expected to be the fastest growing application type for the IoT by 2023 (CISCO, 2021[19]). By 2030, about 95% of new vehicles sold globally will be connected (Bertoncello et al., 2021[44]). Manufacturers are also actively experimenting with and testing models of automated driving with different mixes of digital technologies. As more aspects of vehicles become data-driven, traditional automotive companies are increasingly competing, and sometimes partnering with, information communications technology (ICT) firms (OECD, forthcoming[45]). The data generation and

collection capabilities of a vehicle are largely determined during vehicle design, which therefore differs across manufacturers (ACEA, 2022[46]). Vehicles are evolving in their level of connectivity and automation and their technical function. As they do, the ways data are generated, processed and transferred within and by vehicles are also varied and evolving.

Nevertheless, the volume of data generated by connected vehicles is already immense and is expected to increase as vehicles become more automated. Much of these data are personal in nature, meaning that they can be identified and associated with an individual. Indeed, a growing number of actors, from regulators to third-party service providers, has acknowledged the potential uses and value of these data (NTC, 2020[47]; BEUC/FIA, 2021[48]; CLEPA, 2022[49]). The following section turns to the emerging and evolving data landscape of connected and automated vehicles.

3.2 The emerging data landscape of connected and automated vehicles

Connected vehicles with a degree of automation are already commonplace across the OECD, and competition to develop increasingly automated vehicles is underway. This sub-section explores how connected and automated vehicles generate and collect data, and different models for data processing as these vehicles advance.

Connected and automated vehicles collect and generate vast and varied volumes of data

Connected vehicles can collect and transmit data differently than previous generations of vehicles. As vehicles become increasingly automated, data generation, collection, storage, transfer and use are expected to become increasingly essential to their function (OECD, 2018[5]). The types of data collected or generated by such vehicles will include the following:

- Locational data – data about the precise geographic location of a vehicle, including direction and speed. Such data will interact with both static data (such as high-definition 3D maps) and semi-static data (on temporary events like weather, accidents and traffic jams).
- Sensor data – data about how the car perceives the external environment, including infrastructure, traffic signals and other road users. Such data are typically derived from radar or light detection and ranging (LIDAR) sensors or cameras.
- Diagnostic data – data on how the vehicle is performing with respect to fuel consumption, energy emissions, engine operation, battery status and performance, among other indicators.
- Driving behaviour data – data on driver behaviour, such as seatbelt use, speeding and frequent stopping.
- Identity and biometric data – identity data, like names and other identifying information. This may also include biometric information, like fingerprints.

Estimates of the volumes of these data vary. However, many authors note that connected cars can already produce up to 25 gigabytes of data per hour. They suggest the amount of data generated or collected is likely to increase as vehicles are increasingly automated (NTC, 2020[47]; Bertoncello et al., 2021[44]). These data help ensure the operation of the vehicle and optimise vehicle function. However, the Global Privacy Assembly notes the numerous sensors in connected vehicles imply a “very high risk of excessive data collection compared to what is necessary to achieve the purpose” (GPA, 2018[50]). Vehicles are expected to have a relatively long ownership and usage lifecycle, in the order of 10-20 years. This implies a longer scope of data collection over time, although not all these data are necessarily retained or stored (ACEA, 2022[46]).

Much data collected by connected and automated cars are likely to be personal in nature. Some data will be directly identifiable, like volunteered personal information or inferred personal information from driving habits. Other data might be indirectly identifiable, such as details of journeys made or vehicle usage data. Moreover, much data from connected and automated vehicles are likely to be geolocal in nature. This will include many indirectly identifiable data points related to individuals and their behaviour.

Even anonymised, data from connected and automated vehicles are likely to be sufficiently heterogeneous, varied and precise to enable re-identification of individuals based on their behaviour (Rocher, Hendrickx and de Montjoye, 2019[23]). This may become increasingly relevant as AI systems advance in their capabilities. Moreover, location data will become increasingly precise with the evolution of next generation wireless networks and their eventual combination with global navigation satellite system-based localisation and positioning systems (Ghosh et al., 2021[51]; Mukkavilli and Zhang, 16 December 2021[52]).

A complex data ecosystem for connected and automated vehicles

Connected and automated vehicles increasingly collect a variety of data. Approaches to processing vehicular data are also evolving. Under current data protection policies and industry practices, data from vehicles are typically collected and processed after consent is granted by the vehicle user (ACEA, 2015[53]; European Data Protection Board, 2021[24]; Autos Innovate, 2022[54]). As vehicles grow in their technical capacity, automated cars are likely to be increasingly equipped with compute resources to support their data-intensive, high-performance applications. This will enable some data to be processed within the vehicle. Some models of car automation keep mission-critical data processing on board and off line. This ensures that vehicles can still function without connection to cellular networks (Condliffe, 2017[55]). However, many models of automated driving expect vehicles to exchange data with each other, and with infrastructures, to add spatial range to data collected by the vehicles' own sensors.

Where models of connected and automated driving involve increasing connectivity with other vehicles and infrastructures, the data processing landscape can become more complex. Most connected vehicle data leave the car via cellular networks and are processed in data centres or cloud platforms owned by the equipment manufacturer (Otonomo, 2019[56]). However, as data needs evolve, and vehicles potentially become more connected with each other and infrastructures, many expect that distributed computing architectures will take on more data processing (AECC, 2021[29]).

Edge computing architectures would have the advantage of reducing latency and enabling computing and data processing to take place closer to the vehicle. Many connected driving applications have limited temporal and geographic area of interest, where information is only needed close to the source and time where it was generated, i.e. a traffic jam or an accident (Kousaridas et al., 2020[57]). Therefore, more distributed computing may reduce the need to transmit, store or analyse data after they become irrelevant (IBM, 2021[16]).

Similarly, an edge digital solution can prioritise which data need to be processed by the vehicle's on-board computing system, and which should be relayed to data centres for further analysis (IBM, 2021[16]). As data flows from connected vehicles become more "local", it could reduce the need for cross-border transfers of data. In this model, network operators would host edge servers running cloud-based applications closer to the vehicle. This implies that network operators, and cloud service providers, may also be involved in the processing of data from vehicles. Where data are shared with road infrastructures or other vehicles, road authorities, telecommunication operators and other vehicles may also play a role in data processing (AECC, 2021[29]).

Moreover, as driving becomes a more data-enabled activity, new players can offer services adjacent to the driving process. Services like autopilot software or traffic management services, for example, would involve processing data generated by vehicles (European Data Protection Board, 2021[24]). Similarly, as vehicles become more automated and drivers have less direct input into the driving process, there may be

more demand for in-vehicle applications that require real-time data download, like media consumption. Furthermore, as new service providers emerge, new actors with data-driven offers, such as motor vehicle insurance companies, may also play a role in accessing and processing data generated by vehicles.

Data from connected and automated vehicles hold value in many applications

As outlined above, vehicles are increasingly gaining the ability to collect a greater variety and volume of data. In most jurisdictions, these data are primarily controlled by the original equipment manufacturer following consent from users as part of the terms of service. Some authors describe original equipment manufacturers as in a "gatekeeper position" with control over data generated by connected and automated vehicles (Kerber, 2018[58]; Martens and Muller-Langer, 2018[59]; Iacob, Campmas and Simonelli, 2021[60]). Data from connected and automated vehicles have value to a wide and complex array of actors.

Data collected by connected and automated vehicles can be valuable to other vehicles and infrastructures to support their basic function. Some models of automated driving rely on real-time flows of data to achieve real-time two-way communication. This is true both among connected vehicles and between them and connected infrastructures, like parking spaces and motorways (OECD, 2018[5]). In these models, effective automated driving relies both on short-range communications technologies in cars, as well as mechanisms to enable data access and sharing between a diverse set of stakeholders. These could include vehicle manufacturers, regulators, communication service providers and third-party service providers.

Similarly, data generated by vehicles could improve public policy development and delivery. Potential applications for connected car data include improved traffic management and safety, more dynamic management of traffic control and public spaces, and better and faster responses to incidents in real time. Data from connected and automated vehicles could also help authorities plan the deployment and improvement of new infrastructures and public services, like roads or bike paths. (ITF, 2022[61]).

Data generated by connected vehicles may also emerge as an input into production for firms. Automotive ecosystems are already characterised by a wider variety of actors. These include suppliers and downstream sectors with strong linkages, like the wholesale and retail trade and repair of motor vehicles sector (OECD, forthcoming[45]). Data from connected vehicles could have value for vehicle manufacturers, enabling new business models and revenue streams associated with improving customer experience and operational efficiency. Similarly, data from connected vehicles could provide the basis for the provision of a range of other data-driven business models to car users. These might include predictive repair and maintenance services or dynamic adjustment of insurance services based on driving behaviour (Bertoncello et al., 2021[44]).

The sharing of data generated and collected by vehicles demands that data are interoperable. In the context of vehicle-generated data, both semantic and technical interoperability are needed for sharing to take place. Even where third parties may have access to data from connected and automated vehicles, the use of proprietary standards can increase the costs of the effective use of these data. This can preclude the participation of less technologically adept firms (Iacob, Campmas and Simonelli, 2021[60]).

Connected and automated vehicles across borders

Connected and automated vehicles are mobile by definition with the ability to move across borders and jurisdictions. For fully automated vehicles that rely on cellular connectivity to function across borders, high quality of service for handovers between mobile network operators is needed. This can be a challenge if each network authority has a different network architecture (Kousaridas et al., 2020[57]; AECC, 2021[29]).

Many potential models of automated driving involve constant communication between vehicles and road authorities to quickly and efficiently ensure that automated vehicles adapt to traffic events. However, moving between jurisdictions can mean adapting to data transmission and reception with the digital systems of road authorities. These systems may be closed and proprietary (Hetzer et al., 2021[62]). For

automated vehicles to be effective across contiguous borders, mechanisms may be needed to enable the flow of driving-relevant transport data between jurisdictions.

Restrictions on the flow of data across borders can also impact connected and automated vehicles. As noted above, privacy enforcement authorities believe a large share of data generated by connected and automated vehicles are likely to be personal. Moreover, an increasing share of OECD countries place conditions on the flow of such data across borders. Industry players also note that jurisdictions differ on which aspects of vehicle data are likely to be "personal" in nature, and which data practices are likely to require driver consent (Otonomo, 2019[56]). This can complicate decisions for data processing and sharing across borders. For example, conditions on the flow of data across borders could preclude an automotive company from collecting and analysing accident and fault data across the globe (GSMA, 2019[63]).

Policies relevant to connected and automated vehicles often differ across national and regional borders. For example, in the United States, states differ in their approaches to data collection and storage from event data recorders (NCSL, 2022[64]). Some states have passed comprehensive regulation that enables autonomous vehicle testing (NCSL, 2022[65]). Some models of automated driving require scanning of external terrain and capturing images of surrounding vehicles and pedestrians. However, regulations relating to filming public spaces differ among European Union member states (European Data Protection Board, 2021[24]; Iacob, Campmas and Simonelli, 2021[60]). Similarly, divergent enforcement practices can lead to a fragmented regulatory landscape with uncertain protections as vehicles cross borders. Some authors highlight the issue as a source of uncertainty for individuals and businesses alike (Hickman, 21 October 2019[66]; Iacob, Campmas and Simonelli, 2021[60]) that could preclude the provision of data-driven services in vehicles across borders (Iacob, Campmas and Simonelli, 2021[60]).

3.3 Emerging approaches to data governance for connected and automated vehicles

Data from connected and automated vehicles are the subject of policy debate across the OECD. Discussions on how the data should be collected, processed, shared and reported are ongoing across OECD jurisdictions. This sub-section discusses emerging policy, technological and organisational approaches to addressing the above issues.

Approaches to data collection and management at the vehicle level

As outlined above, data collected by connected and automated vehicles are likely to be personal in nature. Much technical data like vehicle movement and condition are likely to concern the driver or the passengers. In view of the privacy risks associated with the data generated and collected by vehicles, both the public and private sectors have developed principles for data collection and management at the vehicle level. These complement data protection regulations that usually do not directly address automotive data.

Among original equipment manufacturers, alliances of industry actors have promulgated codes of conduct and privacy principles. The European Automobile Manufacturers Association has developed the Principles of Data Protection (ACEA, 2015[53]). In the United States, the Alliance for Automotive Innovation has developed Consumer Privacy Protection Principles. The Alliance represents the manufacturers of 98% of personal vehicles sold in the United States (Autos Innovate, 2022[54]). Other players in the automotive ecosystem have also developed privacy principles. Automotive data aggregator Otonomo, for example, released a "Privacy Playbook" in 2019 (Otonomo, 2019[56]). Some data protection and privacy enforcement authorities have also released principles for data management and collection at the vehicle level for connected and automated vehicles (European Data Protection Board, 2021[24]).

These principles commonly feature a focus on seeking consent for data collection and management at the vehicle level, while minimising overall data collection. Regulatory authorities also tend to include a specific

focus on the management and collection of different types of data in view of their potential risks to data protection and privacy. For example, the European Data Protection Board outlines recommendations for the collection and management of geolocal data, biometric data and data related to fines and traffic-related offences (European Data Protection Board, 2021[24]).

Data processing near or within vehicles

Questions of where and how to process data from connected and automated vehicles are informed by both technical requirements, including the need to ensure vehicle operation, as well as regulatory and other requirements and obligations of the data controller, including with respect to the management of personal data.

In some models of connected and automated driving, data are expected to be shared and processed with a widening ecosystem of actors. These range from equipment manufacturers, regulatory authorities and cloud service providers to network operators and third parties providing in-vehicle services. Noting this expansion, the European Data Protection Board has encouraged local processing of data, notably within the vehicle. It also recommends avoiding external cloud computing of data generated in connected vehicles (European Data Protection Board, 2021[24]). The Board notes that such an approach can increase user control over personal data, as well as "mitigate the potential risks of cloud processing" (European Data Protection Board, 2021[24]).

More local processing of data, namely closer to the vehicle, is increasingly viewed as a solution to both technical and privacy concerns. Edge computing enables reduced response times for the processing of large volumes of data generated by vehicles. In addition, data processing on the network edge can minimise the need to store sensitive data or share it with the cloud, thereby potentially enhancing privacy and security (IBM, 2021[16]).

Access to in-vehicle data from connected and automated vehicles

As noted above, data from connected and automated vehicles are of value to a wide range of stakeholders. To date, such data have been under the effective control of the original equipment manufacturers after consent from the vehicle owner and/or driver. This consent is typically required to use the vehicle and related services.

Manufacturers often argue that tight controls on this data are necessary for privacy and security. They contend such data are often personally identifiable and subject to appropriate laws. Other actors, including those in the motor vehicle aftermarkets and services areas, argue that access to such data could enable a more competitive ecosystem of services. (Kerber, 2018[58]; Andraško et al., 2021[67]; Iacob, Campmas and Simonelli, 2021[60]) In addition, the owner and purchaser of a vehicle may not be its sole user. This can further complicate questions of how collected data should be appropriately managed and shared (Iacob, Campmas and Simonelli, 2021[60]). Efforts to enhance access to in-vehicle data in a mechanism compliant with the European General Data Protection Regulation are ongoing at the European level. Meanwhile, some jurisdictions like Austria, France and Germany have already set rules on access to vehicle data by third parties, including regulatory authorities and insurance companies.

Notably, such discussions intersect with emerging discussions related to access to data generated by the use of connected devices by consumers and businesses. Measures enabling the portability of some kinds of personal data are increasingly commonplace across the OECD (OECD, 2021[68]; OECD, 2021[69]). However, the standards and requirements for access to non-personal industrial data derived from users are less clear. In a connected vehicle context, such data could include information about fuel usage patterns. Such data are not necessarily personal but are nonetheless derived from the individual usage of the vehicle. Efforts to facilitate access to and sharing of such data are nascent across the OECD (Caplan, Kim and Arthur, 2022[70]; European Commission, 2022[71]).

The many technical approaches to accessing data generated by vehicles are often classified into two main groups: solutions "inside" or "outside" the vehicle. Solutions inside the vehicle include on-board application platforms. These would enable applications to be hosted and deployed within the vehicle, using the vehicle's resources and data, and displayed on the vehicle's interface. Another "inside" solution could enable external devices to connect to the vehicle to use vehicle data, while the applications processing the data are run outside the vehicle's system (McCarthy et al., 2017[72]). Solutions "inside" the vehicle would enable providers of data-based services to interact seamlessly with the driver, and enable real-time access to vehicle-generated data (Iacob, Campmas and Simonelli, 2021[60]). Technical solutions "outside" the vehicle for accessing vehicle-generated data usually involve the transfer of car data to an external server (e.g. controlled by equipment manufacturers, or a third party). Meanwhile, applications processing the data operate outside the vehicle system (Iacob, Campmas and Simonelli, 2021[60]). "Outside" solutions would limit real-time access to data for third parties. However, some argue "outside" solutions may have the advantage of reducing safety, security and liability risks by minimising access to the vehicle system (Iacob, Campmas and Simonelli, 2021[60]).

Models of data sharing and access for connected and automated vehicles

Data from connected and automated vehicles have value for a wide variety of actors. These range from players directly involved in the automotive industry ecosystem to third-party players like entertainment service providers and aftermarket service providers of repair and maintenance. In the context of automated vehicle trials, data sharing among innovators could help speed the rate of experimentation.

However, unlike other valuable inputs into production, several economic characteristics, including non-rivalry, economies of scale and information asymmetries, preclude the emergence of markets for data (OECD, 2022[32]). In the absence of widespread data trading, and since data might be used repeatedly without depletion, new frameworks are required that enable greater data sharing among actors. This is particularly relevant for the automotive industry, which features a variety of upstream and downstream industrial activity (OECD, forthcoming[45]) and where incentives may not exist for data sharing.

Various actors have proposed new solutions to enable wider data sharing in the automotive industry. Industry participants have developed Catena-X, an integrated open data ecosystem that enables data sharing among approved participants. It allows the original data holder to retain granular control over data access rights, which can be continuously adjusted. Catena-X is built on the International Data Spaces standard for trusted data sharing and data sovereignty (International Data Spaces Association, 2021[73]). The network features over 100 members throughout the automotive value chain, including service providers and ICT providers, as well as small and medium-sized enterprises. It aims to build an open and collaborative data ecosystem providing standards, applications and services (Catena-X, 2022[74]). Among other things, Catena-X will provide a common interface to reduce reliance on the closed data platforms from established cloud providers. This, in turn, obliges suppliers to adopt multiple different interfaces (Bundeskartellamt, 2022[75]). Use cases include better monitoring and tracing of vehicles after delivery, providing new opportunities to better recycle vehicle components (Catena-X, 2022[76]; Leeuw, 13 January 2022[77]).

Similar efforts to build a decentralised data ecosystem for mobility data are underway at the European level through the Gaia-X initiative (Leeuw, 13 January 2022[77]). Also built on the International Data Spaces standard, the Mobility Data Space would provide a secure ecosystem where data providers can control the conditions under which third parties can use the data. This would enable secure and comprehensive access to real-time traffic data and sensitive mobility data between public transport providers, car sharing providers, public authorities and other third parties. These players may otherwise be reluctant to share data due to a lack of infrastructure and established data formats and interfaces (Pretzsch et al., 2022[78]).

Third-party service providers also act as data aggregators of connected and automated vehicle data. In general, their service offering involves processing raw vehicle data from vehicle manufacturers, applying

relevant privacy norms and then enabling sharing with other third parties. The largest such data aggregators host up to 50 million vehicles, and several billion data points, on their platforms per day (Otonomo, 2022[79]). Notably, the use of external data aggregators implies the transmission of data to an external server, which intersects with debates on how data from vehicles should be accessed (see above).

The use of decentralised ledger technologies (DLTs) could also enable neutral and secure data sharing for connected and automated vehicles. Although no solutions have reached maturity, industry participants and the European Commission are experimenting with using DLTs to enable decentralised identity services. This could enable better and more trusted sharing of data such as on emissions from vehicles. The pilot indicates that sharing vehicle information, including fuel consumption or emissions, to a blockchain-based monitoring system would be technically feasible (MOBI, 2022[80]). Previous work from the International Transport Forum also identifies other use cases for DLTs for data sharing in a mobility context. These include improving identity management and traceability throughout supply chains, and enabling neutral data sharing among potentially competing service providers (ITF, 2021[81]).

Data reporting and mandatory data collection

Data from connected and automated vehicles hold great potential for regulators and public authorities to improve development of public policies – from transportation to planning. The use of such data can create public value by enhancing network operations, investment, maintenance, planning and road safety in jurisdictions across the OECD.

Initiatives across the OECD encourage or oblige the reporting of some kinds of data generated or collected by vehicles to relevant public authorities. For example, in Europe, the Data for Road Safety initiative enables data sharing between vehicle manufacturers and national road authorities and service providers. This initiative recognises that vehicles are already equipped with technologies that can detect dangerous road conditions and warn drivers. It also points out that greater sharing of these data could benefit road operators and the wider public (Data For Road Safety, 2021[82]). The International Transport Forum recently released good governance principles for frameworks for reporting mobility data to public authorities (ITF, 2022[25]).

In some jurisdictions, authorities mandate the collection of some kinds of data. For example, event data recorders are compulsory in vehicles in some jurisdictions (European Commission, 2021[83]). These devices record technical vehicle and occupant information for a brief period before, during and after a crash. Developing similar data storage systems for automated vehicles was the subject of international negotiation in 2020 (UNECE, 2020[84]).

4 Policy considerations for data governance in a fast-evolving digital technology landscape

This section outlines policy considerations for data governance in a rapidly evolving technological landscape. It looks at the need for policies to stay flexible so they can balance the rights and interests of different parties. It also reflects on the need for policies to remain technologically neutral, which implies regular review and adaptation. Given that emerging applications for digital technologies are highly heterogeneous, the chapter suggests that horizontal policies may need sector-specific rules to identify gaps in regulation and legislation. This, in turn, also implies monitoring horizontal data governance policies in various economic and social applications. Finally, the section discusses the need for policies to target uncertainties in regulation without undermining technological development.

Policy considerations

Keep data governance policies agile and principles-based

Data governance is fundamentally a cross-sectoral policy issue as digital technologies continue to be deployed widely. Section 3 of this report highlights that data governance is both a framework condition and

a horizontal issue to the automotive industry. Other industries are also likely to transform towards a more software-driven and digitally enabled service offering. As industries innovate and digitalise, their generation, use, collection, transfer and storage of data are also likely to evolve. This raises the strong likelihood that new considerations for data governance will also emerge.

In view of the evolving nature of data collection and processing, policies that seek to govern data should aim to remain agile. They should aim not to constrain evolving business models, innovation and competition, while balancing the rights and interests of other parties. The OECD Privacy Guidelines provide an example for data governance policies that are resilient to regulatory obsolescence over time. Almost ten years after the revision of the Privacy Guidelines, privacy enforcement authorities continue to affirm the continued relevance and impact of these guidelines due to their principles-based and technological neutral nature (OECD, 2021[26]).

Keep data governance policies technologically neutral

As digital technologies evolve, the ways in which data are generated and collected will evolve as well. For example, in the context of the automotive sector, vehicle design largely determines which data are generated, processed or collected. This, in turn, is evolving as vehicles increase in their level of connectivity, capacity and automation.

Data governance policies should therefore seek to remain technologically neutral. They should govern whether and how data are shared rather than the technology through which the data are exchanged. In the context of connected and automated vehicles, for example, specific requirements to partition certain kinds of services within vehicles, or regulate the means by which in-vehicle data are accessed, will directly impact vehicle design and manufacturing choices.

Ensuring that regulation remains technology-neutral implies regular review and adaptation to consider digital technology developments. For example, as vehicles and automated driving systems advance, regulations may need to adapt to remove references to “steering wheel” and “pedals” (US National Highway Traffic Safety Administration, 2022[85]). Outcome-based or risk-based regulation could outline outcomes to be achieved but not specify process or a particular means of compliance. This could help ensure regulations remain up to date with technological development (OECD, 2021[86]). Another approach includes the use of regulatory sandboxes, which help enable testing of digitally enabled innovations with fewer regulatory constraints. Some jurisdictions explicitly use such sandboxes to inform the process of technologically neutral regulatory development (Attrey, Lesher and Lomax, 2020[87]).

Consider horizontal and sectoral data governance policies together

Applications for digital technologies continue to proliferate across economies and societies. However, these applications are highly heterogeneous. All applications of digital technologies should not be regulated in the same way. Similarly, the relevant data associated with these digital technologies should not be governed in the same way. For example, previous OECD work has highlighted the diversity of the Internet of Things (OECD, 2018[5]), which varies in terms of both connection technology (i.e. short-range or wide area) and applications. This heterogeneity is well understood in an automotive context, where vehicles comply with strict type approval legislation. This reflects their importance from a legislative perspective, and the potential for serious consequences on economies and societies in the event of a failure (Iacob, Campmas and Simonelli, 2021[60]).

As technological applications are highly heterogeneous, horizontal policies may require complementary sector-specific regulation after identification of gaps in regulation and legislation. General data governance standards can be followed with sector-specific regulations and requirements for data-intensive sectors or areas after gaps are identified. Such an effort, in turn, implies monitoring horizontal data governance policies in various economic and social applications.

Target uncertainties in data governance regulations that can hinder technological development

Data governance policies should seek to maximise the benefits of data use while protecting rights and interests and addressing any related risks and challenges. Horizontal regulations can help set the rules of the road. However, policies should seek to avoid creating uncertainties. They should also clarify any gaps where necessary, particularly where such uncertainties may impact technological development.

Already, the purchase of connected cars often includes buying a stream of services, which can make questions of liability and consumer product safety more complex (OECD, 2018[88]). However, as vehicles increase their level of automation, questions related to the collection of data and the liability framework for automated vehicles persist. Notably, equipment manufacturers argue that uncertainties about liability may be magnified by efforts to increase access to data generated by vehicles (Iacob, Campmas and Simonelli, 2021[60]). Such uncertainties may impact incentives to invest in, experiment with or use increasingly automated vehicles. Further guidance from policy makers to clarify relevant responsibilities and risks could help address such uncertainties.

References

ACEA (2022), "Proposal for a Data Act", *Position Paper*, European Automobile Manufacturers' Association, Brussels, https://www.acea.auto/publication/position-paper-proposal-for-a-data-act/. [46]

ACEA (2015), *ACEA Principles of Data Protection in Relation to Connected Vehicles and Services*, European Automobile Manufacturers' Association, Brussels, https://www.acea.auto/uploads/publications/ACEA_Principles_of_Data_Protection.pdf. [53]

AECC (2021), "Connected Cars: On the edge of a breakthrough", *White Paper*, Automotive Edge Computing Consortium, Wakefield MA, and Mobile World Live, https://aecc.org/wp-content/uploads/2021/05/MWL_-_AECC_whitepaper_-_Design_v2.0.pdf. [29]

Alstom (2022), *Autonomous mobility: The future of rail is automated*, https://www.alstom.com/autonomous-mobility-future-rail-automated (accessed on 15 June 2022). [42]

Andraško, J. et al. (2021), "Sustainable data governance for cooperative, connected and automated mobility in the European Union", *Sustainability*, Vol. 13/9, https://doi.org/10.3390/su131910610. [67]

Attrey, A., M. Lesher and C. Lomax (2020), "The role of sandboxes in promoting flexibility and innovation in the digital age", *Going Digital Toolkit Notes*, No. 2, OECD Publishing, Paris, https://doi.org/10.1787/cdf5ed45-en. [87]

Autos Innovate (2022), *Consumer Privacy Protection Principles: Privacy Principles for Vehicle Technologies and Services*, Established 12 November 2014, Reviewed 18 May, March 2022, Alliance for Automotive Innovation, Inc., Washington, DC, https://www.autosinnovate.org/innovation/Automotive%20Privacy/Consumer_Privacy_Principlesfor_VehicleTechnologies_Services-03-21-19.pdf. [54]

AWS (2022), *General Data Protection Regulation (GDPR) Center*, https://aws.amazon.com/compliance/gdpr-center/ (accessed on 15 June 2022). [28]

Bertoncello, L. et al. (2021), "Unlocking the full life-cycle value from connected-car data", Our Insights, 11 February, McKinsey & Company, Atlanta, https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/unlocking-the-full-life-cycle-value-from-connected-car-data. [44]

BEUC/FIA (2021), "Subject: Urgent need for a legislative proposal on access to in-vehicle data and functions", *Letter to Magrethe Vestager*, The European Consumer Organisation and Fédération internationale de l'automobile, European Bureau, Brussels, [48]

https://www.beuc.eu/sites/default/files/publications/beuc-x-2021-062_beuc_and_fia_joint_letter_on_urgent_need_for_a_legislative_proposal_on_access_to_in-vehicle_data_and_functions.pdf.

Bundeskartellamt (2022), "First component for Gaia-X: Bundeskartellamt gives green light for establishing data network for automotive industry (Catena-X)", 24 May, Press Release, Bundeskartellamt, Bonn, https://www.bundeskartellamt.de. [75]

Caplan, M., J. Kim and L. Arthur (2022), "Data sharing is caring: Consumer access to IoT data under the Australian Consumer Data Right and the European Data Act", 27 June, Lexology, https://www.lexology.com/library/detail.aspx?g=73b13797-49e4-4810-859a-dcfa212c52b7. [70]

Casalini, F., J. López González and T. Nemoto (2021), "Mapping commonalities in regulatory approaches to cross-border data transfers", *OECD Trade Policy Papers*, No. 248, OECD Publishing, Paris, https://doi.org/10.1787/ca9f974e-en. [27]

Catena-X (2022), "Our Strengths at a Glance", webpage, https://catena-x.net/en/mehrwerte (accessed on 15 June 2022). [76]

Catena-X (2022), "The Catena-X Network grows to 104 members", Catena-X, 1 June, News,, https://catena-x.net/en/aktuelles-termine. [74]

CISCO (2021), *Annual Internet Report 2018-23 White Paper*, CISCO, San Jose, CA, https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html. [19]

CISCO (2018), *Cisco Global Cloud Index: Forecast and Methodology, 2016-2021*, CISO, San Jose, CA, https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.pdf. [14]

CLEPA (2022), "Access to in-vehicle data and resources: On a regulatory framework ensuring a level playing field in automotive digitalisation", *Position Paper*, European Association of Automotive Suppliers, Brussels, https://www.fiev.fr/media/2022/07/CLEPA-Position-Paper-Access-to-in-vehicle-data.pdf. [49]

Condliffe, J. (2017), "Why some autonomous cars are going to avoid the Internet", 10 January, MIT Technology Review, https://www.technologyreview.com/2017/01/10/154642/why-some-autonomous-cars-are-going-to-avoid-the-internet/. [55]

Data For Road Safety (2021), "Safety Related Traffic Information Ecosystem: Data for Road Safety", webpage, https://www.dataforroadsafety.eu/ (accessed on 20 June 2022). [82]

European Commission (2022), "Data Act: Commission proposes measures for a fair and innovative data economy", Press Release, 23 February, European Commission, Brussels, https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113. [71]

European Commission (2021), "Next-generation Internet of Things and edge computing", *Event Report from the Fireside Chat of 9 March 2021*, 31 March, European Commission, Brussels, https://digital-strategy.ec.europa.eu/en/library/next-generation-internet-things-and-edge-computing. [30]

European Commission (2021), "Vehicle safety – technical requirements and test procedures for EU type-approval of event data recorders", Draft Act, European Commission, Brussels, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12989-Vehicle-safety- [83]

technical-requirements-&-test-procedures-for-EU-type-approval-of-event-data-recorders-EDRs-_en.

European Data Protection Board (2021), *Guidelines 01/2020 on Processing Personal Data in the Context of Connected Vehicles and Mobility-related Applications*, European Commission, Brussels, https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-012020-processing-personal-data-context_en. [24]

European Technical Standards Institute (2020), "Multi-Access Edge Computing", webpage, https://www.etsi.org/technologies/multi-access-edge-computing (accessed on 19 June 2022). [15]

Ghosh, A. et al. (2021), "The evolution of 5G new radio positioning technologies", *White Paper*, 22 February, Bell Labs and Nokia, Murray Hill, NJ, https://www.bell-labs.com/institute/white-papers/evolution-5g-new-radio-positioning-technologies/#gref. [51]

GPA (2018), "Connected vehicles", *Working Paper*, 63rd meeting, 9-10 April, Budapest, International Working Group on Data Protection in Telecommunications, https://globalprivacyassembly.org/the-icdppc-working-group-on-data-protection-in-telecommunications-adopts-a-working-paper-on-connected-vehicles/. [50]

GSMA (2019), "Protecting privacy and data in the Internet of Things", *Policy Paper*, GSMA, https://www.gsma.com/iot/wp-content/uploads/2019/02/Protecting-Privacy-big-data-report-gsma.pdf. [63]

Hetzer, D. et al. (2021), "5G connected and automated driving: Use cases, technologies and trials in cross-border environments", *Wireless Com Network*, Vol. 97, https://doi.org/10.1186/s13638-021-01976-6. [62]

Hickman, T. (21 October 2019), "Dashcams and autonomous vehicles: Dodging legal landmines in the EU", Our Thinking blog, https://www.whitecase.com/insight-our-thinking/dashcams-and-autonomous-vehicles-dodging-legal-landmines-eu. [66]

Iacob, N., A. Campmas and F. Simonelli (2021), "Big Data and B2B platforms: The next big opportunity for Europe", *Final Report*, CEPS, Brussels, https://www.ceps.eu/ceps-publications/big-data-and-b2b-platforms-the-next-big-opportunity-for-europe/. [60]

IBM (2021), "Edge computing for automotive", *Act on Insights Closer to Where your Data is Created with Edge Computing*, IBM, Armonk, NY, https://www.ibm.com/resources/guides/edge-computing-industry-use-cases/#section-4. [16]

IDC (2021), "Data creation and replication will grow at a faster rate than installed storage capacity, according to the IDC Global DataSphere and StorageSphere Forecasts", Press Release, 24 March, International Data Corporation, Needham, Mass., https://www.idc.com/getdoc.jsp?containerId=prUS47560321. [33]

International Data Spaces Association (2021), "GAIA-X and IDS", *Position Paper, Version 1.0*, International Data Spaces Association, Berlin, https://internationaldataspaces.org//wp-content/uploads/dlm_uploads/IDSA-Position-Paper-GAIA-X-and-IDS.pdf. [73]

ITF (2022), "Reporting Mobility Data: Good Governance Principles and Practices", *International Transport Forum Policy Papers*, No. 101, OECD Publishing, Paris, https://doi.org/10.1787/b988f411-en. [61]

ITF (2022), "Reporting Mobility Data: Good Governance Principles and Practices", *International* [25]

Transport Forum Policy Papers, No. 101, OECD Publishing, Paris, https://doi.org/10.1787/b988f411-en.

ITF (2021), "Forging Links: Unblocking Transport with Blockchain?", *International Transport Forum Policy Papers*, No. 93, OECD Publishing, Paris, https://doi.org/10.1787/738454bb-en. [81]

ITF (2021), *Integrating Public Transport into Mobility as a Service: Summary and Conclusions*, ITF Roundtable Reports, No. 184, OECD Publishing, Paris, https://doi.org/10.1787/94052f32-en. [40]

ITF (2015), "Automated and Autonomous Driving: Regulation under Uncertainty", *International Transport Forum Policy Papers*, No. 7, OECD Publishing, Paris, https://doi.org/10.1787/5jlwvzdfk640-en. [38]

Kerber, W. (2018), "Data governance in connected cars: The problem of access to in-vehicle data", *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, Vol. 14 November, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3285240. [58]

Kousaridas, A. et al. (2020), "5G cross-border operation for connected and automated mobility: Challenges and solutions", *Future Internet*, Vol. 12/1, https://doi.org/10.3390/fi12010005. [57]

Leeuw, V. (13 January 2022), "Gaia-X: A software framework on top of cloud infrastructure for and by multiple ecosystems", Gaia-X blog, https://gaia-x.eu/news/latest-news/gaia-x-a-software-framework-on-top-of-cloud-infrastructure-for-and-by-multiple-ecosystems/. [77]

López González, J., F. Casalini and J. Porras (2022), "A Preliminary Mapping of Data Localisation Measures", *OECD Trade Policy Papers*, No. 262, OECD Publishing, Paris, https://doi.org/10.1787/c5ca3fed-en. [34]

Martens, B. and F. Muller-Langer (2018), "Access to digital car data and competition in aftermarket maintenance service", *Journal of Competition Law and Economics*, Vol. 16/1, pp. 116-141, https://doi.org/10.1093/joclec/nhaa005. [59]

McCarthy, M. et al. (2017), "Access to in-vehicle data and resources", report commissioned by the European Union, TRL, May, https://www.figiefa.eu/wp-content/uploads/TRL-access-to-in-vehicle-data-and-resources.pdf. [72]

MOBI (2022), *European Commission Pilot with Citopia & ITN on CO2 Emissions Monitoring*, MOBI, Wellington, New Zealand, https://dlt.mobi/eupilot/. [80]

Mukkavilli, K. and X. Zhang (16 December 2021), "5G: Bringing precise positioning to the connected intelligent edge", OnQ blog, https://www.qualcomm.com/news/onq/2021/12/5g-bringing-precise-positioning-connected-intelligent-edge. [52]

NCSL (2022), *Autonomous Vehicles State Bill Tracking Database*, (database), https://www.ncsl.org/research/transportation/autonomous-vehicles-legislative-database.aspx (accessed on 19 June 2022). [65]

NCSL (2022), *Privacy of Data from Event Data Recorders: State Statutes*, National Conference of State Legislatures, https://www.ncsl.org/research/telecommunications-and-information-technology/privacy-of-data-from-event-data-recorders.aspx. [64]

NTC (2020), "Government access to in-vehicle data", *Discussion Paper*, National Transport Commission, Melbourne, https://www.ntc.gov.au/transport-reform/ntc-projects/government- [47]

access-vehicle-generated-data.

OECD (2022), "Blockchain at the frontier: Impacts and issues in cross-border co-operation and global governance", *OECD Business and Finance Policy Papers*, No. 04, OECD Publishing, Paris, https://doi.org/10.1787/80e1f9bb-en. [18]

OECD (2022), "Broadband networks of the future", *OECD Digital Economy Papers*, No. 327, OECD Publishing, Paris, https://doi.org/10.1787/755e2d0c-en. [12]

OECD (2022), *Broadband Portal*, (database), https://www.oecd.org/digital/broadband/broadband-statistics/ (accessed on 21 July 2022). [20]

OECD (2022), "Fibre overtakes cable as the primary fixed broadband technology in OECD countries", *OECD Broadband Statistics Update*, 21 July, OECD, Paris, https://www.oecd.org/sti/broadband/broadband-statistics-update.htm. [21]

OECD (2022), *Going Digital to Advance Data Governance for Growth and Well-being*, OECD Publishing, Paris, https://doi.org/10.1787/e3d783b0-en. [1]

OECD (2022), "Measuring the value of data and data flows", *OECD Digital Economy Papers*, No. 345, OECD Publishing, Paris, https://doi.org/10.1787/923230a6-en. [32]

OECD (2022), "OECD Framework for the Classification of AI systems", *OECD Digital Economy Papers*, No. 323, OECD Publishing, Paris, https://doi.org/10.1787/cb6d9eca-en. [9]

OECD (2022), "Recommendation of the Council on Blockchain and Other Distributed Ledger Technologies", *OECD Legal Instruments*, OECD, Paris, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0470. [17]

OECD (2022), *The OECD Tractor Codes*, OECD, Paris, https://www.oecd.org/agriculture/tractors/. [43]

OECD (2021), "Data portability, interoperability and digital platform competition", *OECD Competition Committee Discussion Paper*, http://oe.cd/dpic. [68]

OECD (2021), "Mapping data portability initiatives, opportunities and challenges", *OECD Digital Economy Papers*, No. 321, OECD Publishing, Paris, https://doi.org/10.1787/a6edfab2-en. [69]

OECD (2021), *OECD Regulatory Policy Outlook 2021*, OECD Publishing, Paris, https://doi.org/10.1787/38b0fdb1-en. [86]

OECD (2021), *OECD Science, Technology and Innovation Outlook 2021: Times of Crisis and Opportunity*, OECD Publishing, Paris, https://doi.org/10.1787/75f79015-en. [36]

OECD (2021), *Report on the Implementation of the Recommendation of the Council Concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, OECD, Paris, https://one.oecd.org/document/C(2021)42/en/pdf. [26]

OECD (2020), *OECD Digital Economy Outlook 2020*, OECD Publishing, Paris, https://doi.org/10.1787/bb167041-en. [4]

OECD (2019), *Artificial Intelligence in Society*, OECD Publishing, Paris, https://doi.org/10.1787/eedfee77-en. [7]

OECD (2019), *Going Digital: Shaping Policies, Improving Lives*, OECD Publishing, Paris, [2]

https://doi.org/10.1787/9789264312012-en.

OECD (2019), "Recommendation of the Council on Artificial Intelligence", *OECD Legal Instruments*, OECD, Paris, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. [8]

OECD (2019), "The road to 5G networks: Experience to date and future developments", *OECD Digital Economy Papers*, No. 284, OECD Publishing, Paris, https://doi.org/10.1787/2f880843-en. [11]

OECD (2019), "Vectors of digital transformation", *OECD Digital Economy Papers*, No. 273, OECD Publishing, Paris, https://doi.org/10.1787/5ade2bba-en. [35]

OECD (2018), "Consumer product safety in the Internet of Things", *OECD Digital Economy Papers*, No. 267, OECD Publishing, Paris, https://doi.org/10.1787/7c45fa66-en. [88]

OECD (2018), "IoT measurement and applications", *OECD Digital Economy Papers*, No. 271, OECD Publishing, Paris, https://doi.org/10.1787/35209dbf-en. [5]

OECD (2015), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, https://doi.org/10.1787/9789264229358-en. [22]

OECD (2015), *OECD Digital Economy Outlook 2015*, OECD Publishing, Paris, https://doi.org/10.1787/9789264232440-en. [3]

OECD (2014), "Cloud Computing: The Concept, Impacts and the Role of Government Policy", *OECD Digital Economy Papers*, No. 240, OECD Publishing, Paris, https://doi.org/10.1787/5jxzf4lcc7f5-en. [13]

OECD (forthcoming), *Measuring the Internet of Things*, OECD Publishing, Paris. [6]

OECD (forthcoming), "The automotive Sector and its industrial eco-system", *OECD Science, Technology and Industry Policy Papers*, OECD Publishing, Paris. [45]

OECD, UNESCO, IDB (2022), *The Effects of AI on the Working Lives of Women*, UNESCO/OECD/Inter-American Development Bank, https://wp.oecd.ai/app/uploads/2022/03/The-Effects-of-AI-on-the-Working-Lives-of-Women.pdf. [10]

Otonomo (2022), *Powering the Mobility Economy*, Otonomo, Herzeliya, Israel, https://otonomo.io/. [79]

Otonomo (2019), "A privacy playbook for connected car data", *Otonomo White Paper*, Otonomo, Herzliya, Israel, https://fpf.org/wp-content/uploads/2020/01/OtonomoPrivacyPaper.pdf. [56]

Panda, D. et al. (2018), "DNA as a digital information storage device: Hope or hype?", *Biotech*, Vol. 8/5, p. 239, https://doi.org/10.1007/s13205-018-1246-7. [31]

Pretzsch, S. et al. (2022), *Mobility Data Space: Secure Data Space for the Sovereign and Cross-Platform Utilisation of Mobility Data*, Fraunhofer Institute for Transportation, Dresden, Germany, https://www.mobility-data-space.de/content/dam/ivi/mobility-data-space/documents/Mobility_Data_Space_2022_EN.pdf. [78]

Rocher, L., J. Hendrickx and Y. de Montjoye (2019), "Estimating the success of re-identifications in incomplete datasets using generative models", *Nature Communications*, Vol. 10/3069, https://doi.org/10.1038/s41467-019-10933-3. [23]

SAE (2014), "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles", webpage, https://www.sae.org/misc/pdfs/automated_driving.pdf (accessed on 15 June 2022). [37]

UNECE (2020), "UN regulation on automated lane keeping systems is milestone for safe introduction of automated vehicles in traffic ", News, 24 June, United Nations Economic Commission for Europe, Geneva, https://unece.org/transport/press/un-regulation-automated-lane-keeping-systems-milestone-safe-introduction-automated. [84]

US National Highway Traffic Safety Administration (2022), *Occupant Protection for Vehicles with Automated Driving Systems*, Department of Transportation, Washington, D.C., https://www.nhtsa.gov/sites/nhtsa.gov/files/2022-03/Final-Rule-Occupant-Protection-Amendment-Automated-Vehicles.pdf. [85]

Wayland, M. (2022), "'A ghost is driving the car' — my peaceful and productive experience in a Waymo self-driving van", 30 March, CNBC, https://www.cnbc.com/2022/03/30/waymo-self-driving-experience-mostly-peaceful-and-productive.html. [41]

Winder, A. (2016), *Study of the Scope of Intelligent Transport Systems for Reducing CO2 Emissions and Increasing Safety of Heavy Goods Vehicles, Buses and Coaches*, 9 September, ERTICO-ITS Europe, Brussels, https://erticonetwork.com/wp-content/uploads/2016/09/ITS4CV-Report-final-2016-09-09.pdf. [39]