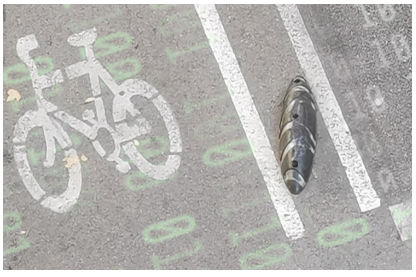
International Transport Forum | CPB Corporate Partnership Board

Reporting Mobility Data

Good Governance Principles and Practices



Corporate Partnership Board Report

Reporting Mobility Data

Good Governance Principles and Practices

Corporate Partnership Board Report

The International Transport Forum

The International Transport Forum is an intergovernmental organisation with 63 member countries. It acts as a think tank for transport policy and organises the Annual Summit of transport ministers. ITF is the only global body that covers all transport modes. The ITF is politically autonomous and administratively integrated with the OECD.

The ITF works for transport policies that improve peoples' lives. Our mission is to foster a deeper understanding of the role of transport in economic growth, environmental sustainability and social inclusion and to raise the public profile of transport policy.

The ITF organises global dialogue for better transport. We act as a platform for discussion and pre-negotiation of policy issues across all transport modes. We analyse trends, share knowledge and promote exchange among transport decision-makers and civil society. The ITF's Annual Summit is the world's largest gathering of transport ministers and the leading global platform for dialogue on transport policy.

The Members of the Forum are: Albania, Armenia, Argentina, Australia, Austria, Azerbaijan, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Canada, Chile, China (People's Republic of), Colombia, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Hungary, Iceland, India, Ireland, Israel, Italy, Japan, Kazakhstan, Korea, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Mexico, Republic of Moldova, Mongolia, Montenegro, Morocco, the Netherlands, New Zealand, North Macedonia, Norway, Poland, Portugal, Romania, Russian Federation, Serbia, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Tunisia, Turkey, Ukraine, the United Arab Emirates, the United Kingdom, the United States, and Uzbekistan.

About the Corporate Partnership Board

The Corporate Partnership Board (CPB) is the International Transport Forum's platform for engaging with the private sector and enriching global transport policy discussion with a business perspective. The members of the ITF Corporate Partnership Board are: AB InBev, Airbus, Allianz Partners, Alstom, Aramco, Argo AI, Arrival, AutoCrypt, Bosch, CEIIA, Cruise, ExxonMobil, Iberdrola, Kakaomobility, Michelin, Mott Macdonald, NXP, PTV Group, RATP Group, Rolls Royce, Shell, Siemens, SPEA Engineering, TIER Mobility, Total Energies, Toyota, Trucknet, Uber, Valeo, Voi, Volvo Cars and Volvo Group.

Disclaimer

Funding for this work has been provided by the ITF Corporate Partnership Board. This report is published under the responsibility of the Secretary-General of the ITF. It has not been subject to the scrutiny of ITF or OECD member countries, and does not necessarily reflect their official views or those of the members of the Corporate Partnership Board.

Cite this work as: ITF (2021), "Reporting Mobility Data: Good Governance Principles and Practices", *International Transport Forum Policy Papers*, No. 101, OECD Publishing, Paris.

Acknowledgements

This report was written by Philippe Crist with substantive contributions from Camille Combe, both of the International Transport Forum (ITF). The project was managed by Philippe Crist.

The work for this report was carried out in the context of a project initiated and funded by the International Transport Forum's Corporate Partnership Board (CPB). CPB projects are designed to enrich policy discussion with a business perspective. They are launched in areas where CPB member companies identify an emerging issue in transport policy or an innovation challenge to the transport system. Led by the ITF, the project development is carried out collaboratively in working groups consisting of CPB member companies, external experts and ITF staff.

A virtual workshop information sharing and discussion with members of the ITF Corporate Partnership Board, including Rem Dekker (Waymo), Jacques Ferriere (RATP), Benoit Marichal (RATP), Paulo Humanes (PTV), Uttara Sivaram (Uber), Laurent Tridemy (Michelin) and Mathieu Voisin (RATP). Also contributing to the workshop discussions were Diego Canales (Populus), Sebastian Castellanos (NUMO), Jascha Franklin-Hodge (Open Mobility Foundation), Michael Schnuerle (Open Mobility Foundation), Kevin Webb (SharedStreets), Rebecca Williams (Belfer Center – Harvard University). Philippe Crist, Sharon Masterson, Asuka Ito and Maria Santos Alfageme all participated for ITF.

Several external experts were also consulted in the preparation of this report. These included Jean Coldefy (ATEC-ITS), Jonathan Couppe (Mairie de Paris), Antoine Courmont (Sciences Po), Laetitia Dablanc (University Gustave Eiffel), Thomas Geier (EMTA), Mélanie Gidel (Mairie de Paris), Anabelle Huet (UITP), Gayang Ho (UITP), Vincent Neumayer (Wiener Linien), Gerald Stöckl (Upstream Mobility).

The authors would like to thank Jari Kauppila, Sharon Masterson and Mary Crass (all ITF) for their review of a draft version of the report Gemma Nellies (independent) for editorial support and Hilary Gaboriau (ITF) for co-ordinating publication.

Table of contents

Figures

Tables

Boxes

Executive summary

What we did

Transport systems and the people using them generate an ever-increasing amount of data. These data represent a largely untapped potential source for transport system performance improvement but also pose significant and often poorly understood risks. Mobility data-governance frameworks are comprised of two pillars – data sharing and data reporting. Data sharing refers to data shared among market actors and other stakeholders, which enables the delivery of mobility and other services and which supports the functioning of transport markets. Data reporting refers to data provided by stakeholders and market actors to public authorities that enables the latter to monitor, guide or intervene to enact public policy. This report explores the issues that public authorities must address when establishing data-reporting mandates and policies.

What we found

Good governance often requires access to sensitive or personal data – this is especially evident in the field of mobility. These data are increasingly found in the private sector and thus must be collected by public authorities. The balance of benefits and harms emerging from government access to data underscores the need for appropriate and effective data-governance frameworks that recognise, respect and enshrine individual privacy rights. At the core of these frameworks is the need to reconcile what is technically possible, what is desirable and what is legally permitted.

Reporting of personal data should adhere to principles that ensure the highest level of privacy protection. Personal data can be linked directly or indirectly to natural persons and therefore poses the most significant risk to potential impacts on privacy rights. The definition of "personal data" should be expansive, given that the risks of re-identification continue to increase over time. Whether data are considered personal does not prevent the collection or reporting of such data but should trigger additional care in its processing and handling.

The sharing or release of data deemed to be commercially confidential or sensitive poses another class of risks. The commercial sensitivity of informational or operational data depends on the context and nature of collected and reported data. Not all technical or operational data releases will harm competition, although transactional information and other information regarding investments or service development are almost always commercially sensitive.

Public authorities can reduce both real and perceived data-reporting risks by adopting proportionate and transparent data-collection and handling protocols that protect personal and commercially sensitive data.

Data collection by authorities may be compulsory, for example, by making data reporting conditional to accessing a service or a set of rights, voluntarily or on commercial terms. Public authorities select data acquisition pathways depending on the types of public mandates they hold and the policy objectives they aim to achieve. Governments and firms must consider the relative costs of different data acquisition pathways, as well as the proportionality and balance of these costs in relation to the expected benefits from such data collection. Public authorities must also consider their technical and human resource capacity to materially collect, process and manage data when selecting a data acquisition pathway. When mandates for public authorities are clear and well-articulated – for example, the need to ensure high levels

of road safety or to manage public spaces for the public good – then direct or conditional compulsion of data reporting is warranted, provided that the burdens imposed on reporting parties are in line with expected benefits.

Data reported to public authorities enables them to plan, manage transport operations or enforce regulations. For this reason, public authorities often require data reporting from market actors or make it a condition of licensure. The level of detail or aggregation of data required by authorities is linked to the specific task at hand.

Planning data are not principally concerned with where individuals travel but with how communities function and, therefore, do not need to be highly disaggregated or available in real-time. Operational data allows authorities to control traffic, manage public spaces and respond to incidents in real-time. Operational data requires greater granularity than planning data but can still be collected and reported at an appropriate level of aggregation. Data supporting enforcement actions relate directly to unique vehicles, individuals and their specific behaviours. As such, this is the most privacy-sensitive of the three data-reporting streams.

Not everything of consequence for urban mobility produces digital data, which may lead to observational biases. Just because a data stream is digital and can easily be collected does not mean it provides an accurate or useful assessment of the "ground truth".

Data reporting necessarily entails imposing certain burdens on reporting parties. Some form of data harmonisation can improve regulatory efficiency. In certain instances, data syntaxes encode personal or commercially sensitive data and thus, reporting requirements that mandate the use of these syntaxes must be accompanied by appropriate risk avoidance or mitigation policies.

Current practices in data governance are generally unfit for delivering desired policy outcomes – including the protection of individuals' fundamental rights, welfare-improving competition, equity and sustainability. These data-governance frameworks were developed outside the sphere of public governance and not designed to deliver on such outcomes. Unlike public spaces, digital spaces are largely unregulated, or rather, they are only partially regulated, imperfectly, and with no clear consensus on how and for whom they should be regulated. New governance frameworks are needed in order to govern digital spaces effectively. The European Union has undertaken the most extensive and ambitious effort to update regulation for modern data-rich societies – The General Data Protection Regulation 2016 (GDPR) – alongside a number of other developing data-governance initiatives.

New mechanisms and building blocks are required to ensure that the data ecosystem preserves privacy and protects commercial interests by default. These mechanisms will build on new concepts, such as data trusts, personal data spaces and data architectures expressly built to deliver on public outcomes – the "public stack" – but will require tangible and actionable tools. Such tools are only just emerging.

What we recommend

Embed individual privacy rights at the heart of data-reporting policies

Data-governance frameworks must explicitly seek to avoid eroding privacy rights while at the same time maximising benefits. Reporting data for planning and operational purposes should avoid personal data by default. Data reporting related to enforcement should be given the highest protection and treated with adapted and privacy-preserving reporting pathways and processing protocols – separate, distinguishable and more secure than other data-reporting pathways. In this respect, a fundamental

precautionary principle is that enforcement actions should only collect, process and retain data on contravening incidents.

Adopt coherent data-governance frameworks

Developing and adopting a coherent data-governance framework helps defuse tensions around personal or commercially sensitive data reporting and build trust concerning how and why authorities require data from the private sector. Where data-governance frameworks do not exist at the supra-national or national level (e.g. unlike in Europe with the GDPR), local and regional authorities should seek to clearly specify and adopt best-practice data-governance approaches – including undertaking assessments of personal data protection. As part of this approach, authorities should clearly identify data-stewardship and data-custodianship roles and responsibilities and establish an inventory of data-reporting mandates.

Establish, document and communicate the basis for public authority data-reporting mandates

Public authority data-reporting mandates should be linked to explicit and lawful objectives. The specific purposes for which a public authority is collecting data – especially when compelled or required for licensure or service operation – should be lawful, clearly stated, publicly documented and reference the legal basis for data collection.

Align data-reporting mandates to targeted outcomes

Authorities must evaluate whether the data they require reporting is necessary for carrying out the stated purposes of its collection – especially regarding personal or commercially sensitive data. Public authorities should evaluate alternatives to mandatory or conditional reporting of personal data. This assessment should be open and provide opportunities for stakeholders to identify and provide evidence on the ability and efficacy of non-personal data alternatives to achieve stated objectives. Authorities should also ascertain the accuracy and representativeness of the data reported to them when establishing and carrying out their reporting mandates.

Create and adhere to clear personal data processing, retention and destruction policies

Clear data policies are necessary to build confidence that personal data will be handled and disposed of according to its sensitivity. This includes ensuring that data subjects were notified of public authority processing when their consent was obtained, assessing if the outcome of data processing is suited for the purposes for which it was collected and if any biases have emerged from the processing of personal data. Public authorities should limit the sharing of personal and sensitive data to the minimum extent necessary to achieve the purpose of its collection. They should also ensure that sufficiently strong data access controls are in place. Finally, personal data should be retained and stored securely according to its sensitivity and only for as long as necessary. Public authorities should apply specific and documented protocols for irreversibly de-identifying or destroying personal data.

Explore ways to ensure that data reporting preserves privacy and protects commercial interests by default

Efforts to govern rapidly evolving data ecosystems with decades-old regulatory frameworks designed for analogue services are bound to result in suboptimal, inefficient and possibly privacy-damaging outcomes. In particular, data-sharing and data-reporting frameworks built on outdated regulations inherit deficiencies stemming from the poor fit between those regulatory frameworks and the needs of digital regulation. Public authorities should explore the development and adoption of new and adapted data-governance building blocks that enable the emergence of in-built privacy protection within the data ecosystem. This would enable data protection for data sharing and reporting by design, not by retro-fit.

Governance, data and data governance

Governance is an essential function of society – when framed by democratic and inclusive practices, it enables the development of policies that improve people's lives. Governance is the framework of rules, relationships, systems and processes that enable collective goals to be met (Waag, 2021). Governance frameworks have emerged in all sectors of the economy and all facets of society. They are prominent in the transport sector, addressing the material conditions under which mobility operates and in which transport improves connectivity and accessibility. Digital spaces are a more recent manifestation of human activity, and comprehensive data-governance frameworks are lacking – or are only just emerging. Just as consensus has formed around other spheres of governance, so too will it have to coalesce in relation to digital public spaces (Lehrer, 2021).

Good governance, in mobility and elsewhere, requires access to sensitive or personal data. The nature and scope of data available to carry out effective governance are rapidly evolving with the digital revolution and there is a generalised shift of potentially relevant data from the public to the private sector. These changes require adapted data collection and reporting frameworks that minimise burdens on and risks to firms and individuals while maximising public benefits. At the heart of the evolving governance challenge in mobility and elsewhere is how to balance the needs and capacity for public authorities to gather or otherwise access data in the public interest versus maintaining individual privacy and fostering space for commercial innovation. This report explores principles and rules that can help achieve these goals and frame data-reporting policies which maximise public value.

Governance and data are inextricably linked. Governments play a role in monitoring, guiding, incentivising and compelling action by firms and individuals in order to achieve public policy goals and improve social welfare. Data enables governments to govern, and good data enables governments to govern efficiently. Public authority decision making and action requires information about society, markets and, in certain circumstances, even individuals. The need to access relevant data and use them to govern effectively extends to many domains of public action, including health, education, taxation, market oversight and control, public safety and transport.

Public data governance must reconcile what is technically possible, what is desirable and what is legally permitted. Public authorities often lack a coherent vision of what data are being collected from citizens and how that data will be used. On the one hand, governments collect a wide variety of statistical data which help inform policy action. Governments also seek to gather data from emerging data sources, for which established data collection procedures and protocols do not exist. These data give rise to multiple, unco-ordinated requests for data across government departments, which may leave fundamental issues relating to privacy, data ownership and control, processing and data access rights and responsibilities, left partially or fully unaddressed by policy (Custer, 2019).

The relationship between data collection, processing and public governance is durable but evolving. For as long as there has been organised governance, there has been a concomitant need to collect, record and use data for carrying out public action. In Mesopotamia, the world's first form of recorded data-keeping was tied to accounting systems which enabled public authorities to tax and govern. The later spontaneous emergence of writing systems in China and then Mesoamerica also gave rise to data collection supporting

public governance (Schmandt-Besserat, 2015). What has changed over time is the range and volume of data collected, processed and archived, as well as the breadth of activities covered (ITF, 2016). The digitalisation of the economy and of peoples' lives has accelerated these trends to the point where the virtual representation of a firm or an individual can increasingly be transcribed and accessed as digital data (ITF, 2019a). This evolution generates both risks and new opportunities.

Data collected by public authorities has value beyond efficiently and effectively carrying out the act of governance. In particular, widespread access to some forms of data collected by public authorities also generates significant benefits for society, individuals and firms, and supports the development of welfare-improving secondary markets (OECD, 2013). In these instances, there is a case for providing open access to that data if doing so does not comprise privacy or commercial sensitivity risks.

Access to data by public authorities also poses specific and potentially significant risks. Such risks are linked to the personal or sensitive nature of some forms of data collected by public authorities. These risks are also linked to the ability of governments to compel or constrain individuals' and firms' actions on the basis of data collected. In democratic societies, these risks are generally mitigated by open, transparent and participatory governance processes that seek to contain the potential for overreach and abuse by public authorities. In less democratic or authoritarian regimes, these powers may be unchecked. In both instances, misuse, abuse or incautious handling of personal or sensitive data reported to public authorities can lead to significant harm.

Digital data collection in support of public policy purposes accentuates fundamental tensions concerning government oversight and control. If left unaddressed, these tensions and conflicts may inhibit the acceptance of data reporting by the public, erode trust between private and public stakeholders and limit possibilities for innovative, effective and ethical public authority use of data that they do not themselves collect. The Dutch Ministry of the Interior identifies six fundamental tensions that must be addressed in public authority data-governance frameworks (Geist, Klievink and Steunenberg, 2019):

- the conflict between privacy and technology
- the conflict between companies and individuals, and between the individual and the collective public interest
- the conflict between transparency and the extent to which data use in the public domain can be motivated and justified
- the conflict between the original indicated purpose of data collection […] and subsequent data reuse
- the conflicting interests in the partnership between public and private parties
- the question of what role government should play within the "smart society".

The balance of benefits and harms emerging from government access to data underscores the need for appropriate and effective data-governance frameworks that recognise, respect and enshrine individual privacy rights. Privacy is a fundamental human right formally established by the United Nations in 1948 (United Nations, 1948). In common with other forms of data that can be linked to individuals, mobility data may erode this right (see Box 1). Data-governance frameworks must explicitly seek to avoid eroding privacy rights or contributing to other potential harms while maximising benefits. In many instances, data-governance frameworks are no longer fit for purpose or are simply lacking – especially for those sectors characterised by emerging or rapidly evolving data sources. This is the case with many sources of transport and mobility data.

Box 1. Data governance, data reporting and the fundamental human right to privacy

The right to privacy is a recognised fundamental right outlined in Article 12 of the United Nation's Universal Declaration of Human Rights: "No one shall be subjected to arbitrary interference with [their] privacy, family, home or correspondence [….] Everyone has the right to the protection of the law against such interference or attacks"(United Nations, 1948). The office of the United Nations Special Rapporteur on the right to privacy underscores that: "Privacy enables the enjoyment of other rights: the free development and expression of an individual's personality, identity and beliefs, and their ability to participate in political, economic, social and cultural life" and notes that new privacy risks emerge with the potential for digitally-enabled mass surveillance and exploitation of new data sources, including big data, and the application of artificial intelligence in numerous fields of governance and commercial activity (UN-OHCHR, 2021).

The collection of personal data in digital form or via digitally-enabled surveillance of public and private spaces poses significant privacy risks that require adapted and proportionate responses (Zuboff, 2019; OECD, 2021a; Williams, 2021). These risks may manifest themselves inadvertently or indirectly as a result of inherent structural biases in data collection and processing (e.g. in the case of some forms of policing of open public spaces [Seo, 2019; Webb, 2020a] or as an explicit design feature of state or commercial forms of intelligence gathering [Zuboff, 2019; Xu, Leibold and Impiombato, 2021]). The risks related to collecting personal data in these contexts warrant strong privacy preservation by default and by design in data-governance frameworks (Cavoukian, 2011; OECD, 2021a).

For these reasons, a number of governments have established (or are working to establish) legal frameworks that define the right to privacy and provide legal and operational guidance on how to ensure privacy protection. The OECD published such guidance in 1980, amended in 2013, in its recommended Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data (OECD, 2013). The OECD Guidelines outline eight principles relating to the limitation of data collection, data quality, purpose specification, use limitation, security safeguards, openness with respect to personal data-privacy policies, individual participation and accountability (OECD, 2013). The OECD guidelines set out general guidance and establish key definitions (including the nature of the data subject and the data controller) that have been used in many subsequent data-privacy regulations, including, most notably, the EU's General Data Protection Regulation of 2016 (GDPR) (EU, 2016a).

Over 120 countries and jurisdictions have enacted laws relating to personal data protection (DLA Piper, 2021; Thales, 2021). The EU's GDPR is, as of 2021, the most extensive and stringent of these legal frameworks and applies to all data collected within the European Union or relating to data subjects who are EU citizens. While some aspects of the GDPR have proven difficult to manage or enforce effectively (e.g. obtaining meaningful consent from data subjects), its robust enforcement and consequential fines have set the context for many other data-privacy frameworks. These include Brazil's *Lei Geral de Proteção de Dados* (LGPD), the Canadian Personal Information Protection and Electronic Documents Act (PIPEDA) (slated to be replaced by the Canadian Consumer Privacy Protection Act [CPPA]), the People's Republic of China (hereafter "China") forthcoming Personal Information Protection Law (PIPL), India's Personal Data Protection Bill, Japan's amended Act on the Protection of Personal Information, Korea's Personal Information Protection Act (PIPA) and the UK's Data Protection, Privacy, and Electronic Communications (DPPEC) Regulations of 2019, which adapt the UK's privacy law in the post-Brexit context. In some federal states, privacy protection is enacted at the sub-national level, as in the United States, where California has enacted the California Consumer Privacy Act (CCPA) and the California Privacy Rights Act (CPRA), or in a mix of Federal and state legislation, as in Australia (DLA Piper, 2021).

Factors to consider when establishing or adapting data-reporting frameworks

Public governance requires some level of data reporting to function effectively. Data-reporting frameworks must account for a number of factors in order to balance the risks and benefits linked to data collection. These include potential data biases, the distinction between data sharing and data reporting, the type of data acquisition methods considered, data heterogeneity, data-related risks, the usefulness of the data to carry out public policy objectives (including data quality and relevance) and data security. These are discussed below.

Data reporting versus data sharing

Mobility data-governance frameworks are comprised of two primary and complementary pillars – data sharing and data reporting. Mobility data sharing refers to data shared among market actors and other stakeholders that enables the delivery of mobility and other services and supports the functioning of transport markets. Data reporting refers to data provided by stakeholders and market actors to public authorities that enables the latter to monitor, guide and intervene to enact public policy (see Figure 1).

This report will focus on data reporting, not on data sharing. Data reporting comprises some element of direct or conditional compulsion. The remainder of this report will focus on data reported to public authorities under direct reporting mandates or via conditional reporting requirements (e.g. in return for a licence to operate).

A key data-reporting question is, what is the data used for? This is particularly pertinent when data reporting is mandatory or made conditional to licensure or obtaining an operating permit. From a public authority perspective, there are three broad uses for such data: data for planning (DR1 in Figure 1), data for the operational management of traffic and public spaces (DR2) and data for enforcement actions (DR3).

Data in support of planning

These are data that improve the capacity of public authorities to carry out their transport, urban and other planning activities. Fundamentally, “planning data” are not principally concerned with where individuals travel but with how communities function (Webb, 2020b). Understanding the latter is more challenging than the former but only requires aggregate insight into mobility patterns. ITF (2021a) addresses many of the issues that must be considered with respect to the use and usefulness of data – especially “big data” – for modelling and planning purposes.

Planning data may enable the creation of origin-destination matrices, which provide insights into how, when and where travel takes place. This information helps governments deliver policies and infrastructure, for example, by providing insights on where new cycling or public transport infrastructure may be needed, where parking may be under- or over-supplied, and where road expansion or re-allocation may be helpful. Alternatively, planning-supportive data may include information on accessibility and mobility service levels. Data reported by mobility operators and other stakeholders may also be used for planning government services or actions outside of the field of transport, for example, in assessing coverage of health or education facilities or adequacy of green space provisions.

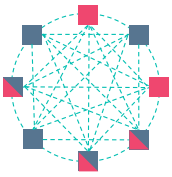
How should public authorities approach the mandatory or conditional reporting of data to be used for planning purposes? First, by collecting only data which is disaggregated or robustly anonymised. Because of its focus on “how communities function”, planning data should not require a high degree of spatial or temporal granularity, nor do they require low latency or real-time access (Webb, 2020b; ITF, 2021a). Disaggregated data revealing personal information are also less useful for planning purposes since they

cannot be shared publicly, limiting discourse and participation in policy making (Webb, 2020b). Exceptions to this rule should be explicitly motivated by a clearly stated and publicly communicated purpose, which cannot be demonstrably met with more aggregate or robustly anonymised data. If personal data are collected, they should be managed in line with legal and other obligations to preserve individual privacy.

Figure 1. Two pillars of mobility data governance: Data sharing and data reporting

Data **sharing**

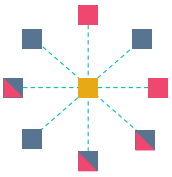
amongst actors and stakeholders to **enable** mobility services and market function



Market actors and other stakeholders share data to support the delivery of mobility services and to enable markets to function.

Data **reporting**

from market actors to public authorities to **monitor** market function



Market actors and other stakeholders report data to public authorities so that authorities can carry out their public mandates.

DS1

Informing
Data allowing stakeholders to plan or co-ordinate services (e.g. schedule, availability, latency, price, etc).

DS2

Operations
Data allowing the fulfillment of travel services (vehicle/ gate access, co-ordination of multi-leg trips, reservation, trip end and final clearing).

DS3

Transactions
Data that enable people to book, pay and gain access to transport services.

DR1

Planning
Anonymised and aggregate (spatially and temporally) data that allow public authorities to understand mobility demand and market operation in order to plan or invest accordingly. Reporting is periodic and does not comprise personal data.

DR2

Operations
Anonymised and aggregate (temporally) data that allow public authorities to intervene in real-time (e.g. via traffic light control or road closure/opening) or otherwise carry out their mandates. Reporting is more frequent (nearing real-time for some specifc data) but data is still de-identified. Data are fully aggregated after use and original data is destroyed.

DR3

Enforcement
Personal or vehicle-specific data reported to support enforcement of code infractions. These data are treated under a separate secure reporting pathway with restricted and vetted access only. All data are destroyed after the completion of the enforcement action in line with data retention policies for enforcement actions. Only aggregate meta-data is retained.

Second, planning data should be collected in response to specific questions. There is no such thing as "universal" or completely objective "raw" data. As a result, the impartiality of questions that would emerge from such data is contestable. Data collection methods are determined by the questions being asked, and

in many cases, the question being asked of mobility operator data is linked to the delivery of that particular service, not to achieving public policy outcomes (Webb, 2020b; ITF, 2021a).

Third, data quality and accuracy must be such that the insights from data used for planning purposes are as relevant and useful as possible (ITF, 2021a). Private-sector data infrastructure serves narrow operational purposes and is generally uncalibrated. In the context of public planning and policy making, it should be applied with significant caution and only after appropriate investment in calibration (Webb, 2020b; ITF, 2021a).

Planning authorities may wish to have access to original data in order to aggregate and anonymise the data themselves. This may be because they expect that aggregated or anonymised data reported to them may not be trustworthy or may have processing errors that limit its accuracy or usefulness. Public authorities must establish trust in what aggregate or otherwise processed data they collect if they do not have access to the original data. One way to do this is to delegate a third party to receive and process the original data and transmit the processed data to public authorities. Another is to specify an audit mechanism that may be used to ensure that mobility operator processed data are accurate. In both cases, audit pathways and functions applicable to either the third-party data processor or the original mobility operator must be defined and enacted in law and/or in policy.

Data in support of operations

These are data that improve the ability of public authorities to carry out their mandated functions. These functions include traffic control, managing public spaces (including parking), ensuring safety, etc. These types of data may include traffic flow information that allows authorities to operate traffic signal and control systems, speed data that allows authorities to monitor and adjust operations to improve safety or parking data that enables authorities to manage public spaces efficiently. They may also allow authorities to identify and respond to incidents in a timely manner.

Ensuring operational functionality requires some spatial and temporal precision, although many outcomes can be ensured with data aggregated into zones and binned into meaningful time bands. Real-time access to these data should generally not be necessary; however, some level of reduced latency may be required for some forms of operational management (e.g. in the order of several minutes rather than seconds).

More fundamentally, traffic operations and parking management data are not designed to be privacy-preserving. That is, they are collected and then processed *ex-post* in order to render them less privacy revealing. *Ex-ante* privacy preservation is technically possible for both traffic and parking management (e.g. edge computing, fog computing, in-stream data processing [Laroui et al., 2021]), although the methods being developed to do this are nascent and not yet generally built into traffic and parking management systems.

Location and trajectory privacy can be ensured via mixed differential privacy, anonymisation, pseudonymisation and cryptographic methods (Chatzigiannakis, Vitaletti and Pyrgelis, 2016; Safi et al., 2017; Zhang, Yan and Kantola, 2017; Ren and Tang, 2020; Atmaca et al., 2021; Mundhe, Verma and Venkatesan, 2021; Qi et al., 2021). As with any privacy-preserving data processing, a trade-off may emerge between data utility and privacy, but this too may be minimised by state-of-the-art data processing methods (for instance, by translating vehicular location data into graph-structured data [Atmaca et al., 2021]). Public authorities should seek to integrate these *ex-ante* privacy-preserving processing methods by default when collecting data for traffic, parking and other operational management functions.

As with planning data, it is necessary to ensure that operator-generated data for operational management tasks are fit for purpose. The first question asked should be if the data present or create biases. If the purpose of the data is to manage traffic flow or parking, do the data accurately represent which traffic

participants or parking incidents are creating the greatest impact? For example, data from app-based ridesourcing or shared micromobility services are, in principle, easy to collect and process. However, they do not necessarily explain the main causes of traffic congestion (ITF, 2019b) or of illegal or disruptive occupation of public space (ITF, 2021b). As noted earlier, just because a data stream is digital and can easily be collected does not mean it provides an accurate or useful assessment of the "ground truth".

Data in support of enforcement actions

Mobility data collected for enforcement purposes typically relates directly to unique vehicles, individuals and their specific behaviours. As such, it is the most privacy-sensitive of all the three data-reporting streams. It is also fraught with higher-order questions linked not to the data itself but to how it will be used. Optimising data collection for enforcement purposes matters little if structural biases present in the criminal justice system lead to inequitable and biased outcomes (Webb, 2020a; Williams, 2021). These structural deficiencies are beyond the scope of this report. However, where they exist, they may erode trust in mobility data-reporting mandates and thus reduce the effectiveness of using data in support of enforcement actions that otherwise would improve traffic safety, better manage public spaces or contribute to other desired public policy goals.

Data-reporting mandates in support of enforcement typically relate to access control (does a vehicle have a right to access or occupy a space? – e.g. a loading/unloading bay, a congestion tolling area, an urban vehicle access restricted zone), traffic enforcement (e.g. speed, respect of traffic rules, vehicular/driver behaviour, etc.) or parking enforcement. Highly detailed data must be collected in order to detect, prosecute or otherwise carry out enforcement actions. These data may include specific vehicle and operator identifiers, precise location or trajectory data and timestamps for actions triggering an enforcement action.

For these reasons, data related to enforcement actions should, by default, be given the highest protection and treated with adapted and privacy-preserving reporting pathways and processing protocols. These should be separate, distinguishable and more secure than other data-reporting pathways and processing protocols. Policies relating to access to these data, processing and retention protocols, and post-enforcement action archiving, aggregation or data destruction should be clearly established and audited for security vulnerabilities and privacy preservation.

Where data are collected in support of law enforcement actions, specific and robust rules generally apply concerning processing, handling, access to and retention of that data (e.g. Article 10 of the GDPR and its transcription into national legislation). These must be respected in the data-reporting cycle. However, not all enforcement actions are law enforcement actions. Parking fine collection and control, for example, has been decriminalised in a number of jurisdictions to reduce the burden on law enforcement and to increase the efficacy of revenue collection and enforcement-related deterrence. Separate and robust data collection and handling protocols should equally be enacted for non-criminal enforcement actions to protect personal privacy and to ensure strict alignment between the collection of personal data and the targeted enforcement action.

Public authorities may feel they require data on all traffic participants to effectively detect and prosecute those that break the law or otherwise trigger an enforcement action. Policing of public space in democratic societies has avoided such ubiquitous surveillance as it raises real concerns with respect to fundamental rights, including privacy, agency and freedom of movement (Green, 2019; Zuboff, 2019; Williams, 2021). New technologies make such surveillance possible, but this does not mean they should be employed to this end.

A fundamental precautionary principle in this respect is that enforcement actions should only collect, process and retain data on contravening incidents. In practice, this means that only data relating to incidents where a law or regulation can be demonstrably shown to have been violated should be retained and transmitted to an enforcement body. In instances where this may not be possible, only those incidents where a defined and transparent threshold for establishing suspicion of a violation should trigger an enforcement action and the retention and processing of data related to that instance. All other data collected in the detection of such incidents should be irreversibly deleted (e.g. automated license plate reading systems used for parking enforcement should only retain data on offending vehicles). Data collected for enforcement purposes should only retain data linked to the enforcement action (e.g. traffic light running cameras should only record still images of offending vehicles and blur out all other traffic participants). Access to enforcement data should be limited only to those whose involvement in the enforcement action is vetted according to established and transparent criteria. The data should be irreversibly deleted or anonymised at the end of the enforcement action and any related appeals processes. One way to ensure robust data access control throughout the lifespan of the enforcement data is through encryption, with only limited and vetted parties having access to the encryption key.

Data reporting on data sharing

Additionally, some forms of data sharing may help support broader public policy outcomes. Data sharing among mobility operators and with mobility service integrators can improve intermodal efficiency and potentially reduce inefficient use of cars. Timetable and real-time system performance and service availability data delivered to travellers can better inform choices and nudge behaviours in support of agreed policy outcomes. These types of data sharing do not involve direct reporting to public authorities but may call for reporting on how well and openly data are being shared in order to gauge the impact of such initiatives or requirements.

The potential bias of no data and "raw" data

Not everything of consequence for urban mobility produces data, let alone digital data. While much of the discourse around digitally-enabled mobility services centres on the large, and often real-time stream of potentially exploitable data, much of what moves in cities does not produce such easily exploitable data streams. Insofar as these modes (walking, cycling, motorised two-wheelers and car driving) form the basis of overall trips in cities, this is a significant blind spot. Data-monitoring methods for these modes are based on observation, not on self-produced and granular digital data. This difference should be accounted for where it may impact the ability of public authorities to monitor overall system performance. Public authorities should seek to avoid asymmetric data-reporting requirements.

Data are never context-less or "raw" – care must be taken to understand what the data actually represents. There is a common misunderstanding that raw data are an objective expression of material reality. This is almost never the case. Raw data are already replete with framing choices and biases that propagate themselves as the data are processed and re-used. These potential biases are numerous and may relate to access and use of technology (e.g. assuming that "drivers" equates "people"), income, gender, physical abilities, etc. Data are generated in very specific and deliberate ways – even sensor-based data. Whitelaw (2010) notes that: "those sensors are designed to measure specific parameters for specific reasons, at certain rates, with certain resolutions. Or more correctly: [data] is gathered by people, for specific reasons, with a certain view of the world in mind, a certain concept of what the problem or the subject is. The people use the sensors, to gather the data, to measure a certain chosen aspect of the world". Data-reporting policies should account for and address the specific framing biases inherent in all data and sensor-based data in particular.

Data acquisition models

In many instances, data are not just observable or openly available but are held by individuals and firms and must be acquired or accessed by the public sector in order to carry out their governance functions. How public authorities do this differs according to data type and public policy outcome targeted. ITF (2016) identifies five data acquisition pathways: public-private data partnerships, public-citizen data partnerships, mandatory data sharing, new data-sharing model and open data. More recently, the European Investment Bank expanded on these and identified seven data acquisition models for public authorities (EIB, 2021) (Box 2). Looking across these and other reviews of public authority data acquisition methods, four over-arching data acquisition models emerge for gathering data that is not generated by public authorities themselves. These models are compulsion, conditionality, co-operation or commercial terms.

Compulsion – the act of requiring or compelling the transfer of data from an individual or a firm to a public authority – is the strongest and most constraining method by which public authorities may acquire data. Compulsion is a common approach to data collection and is generally deployed when the social benefits of acquiring data from individuals and firms outweigh individual privacy or commercial sensitivities linked to that data remaining out of government hands. Public authorities must still ensure that personal or sensitive data are not released or used in ways not aligned with public policy objectives, including protecting individual privacy and commercial competitiveness. Examples of data compulsion include reporting revenue to fiscal authorities and reporting infectious disease cases to health authorities. In transport, examples include reporting of vehicle fuel economy performance, personal identity and drivers' licence data, vehicle registration information, carriage of dangerous goods and passenger identity data for air travel.

Conditionality – the act of requiring the transfer of data in order to access a service, a set of rights or to be granted a licence to operate – is also a common data acquisition pathway for public authorities. Conditionality and compulsion are closely related as some services or licences are so ubiquitous or necessary to function that their acquisition is almost compulsory. The notion of a *quid pro quo*, access to an outcome in exchange for the data that allows the public authority to monitor the conditions in which that outcome is delivered, is especially common in the transport sector. Examples include driver licensing, delivery of public transport contracts and granting mobility service licences to operate (e.g. for taxis, ridesourcing and shared micromobility).

Co-operation – the act of firms or individuals volunteering data to public authorities in order to obtain mutually beneficial outcomes – is a less common data-acquisition pathway for public authorities. Nonetheless, individuals may report some data to governments, for example, relating to individuals' experience of the quality of infrastructure or services through apps such as Paris' "*Dans ma rue*" (On my street) (Ville de Paris, 2021) or to firms' experience of the quality of government services. Firms may also wish to share data with public authorities via voluntary agreements as a way of building trust and avoiding more constraining data-reporting requirements.

Commercial terms – governments may also acquire data by purchasing it from data aggregators and processors. These data are typically obtained by private firms from a number of sources and may include personal or sensitive data in its raw form. The data are aggregated, anonymised and processed to meet a particular public authority (or commercial) need for which governments (or firms) are willing to pay. Examples include processed location-based data built on data from mobile network operators or routing apps (ITF, 2015; 2016; 2021a).

Box 2. Public authority data acquisition models identified by the European Investment Bank

The European Investment Bank commissioned a technical note on *Data Sharing in Transport* to support and advise local authorities on data acquisition in the field of urban mobility. It aims to serve as a starting point for European cities and municipalities seeking practical information and to provide a basis for efficient and effective investment of resources. The note details seven data acquisition pathways for municipalities, cities and regional governments:

- Public procurement of data: a public procurement procedure is used to buy data. In the technical note on *Data Sharing in Transport*, a narrow definition is used to compare this model to others (especially the intermediaries model). This category refers to one-off or as-a-service procurement of a single source dataset without advanced pre-processing by the vendor.
- Intermediaries – integrators, aggregators and marketplaces: in this model, urban authorities call upon a third party that offers services to (pre)process data, extract information, merge data sources or interconnect systems. Marketplaces for mobility data are included in this category.
- Financially compensated partnerships between the public and private sectors: this model is an extension of straightforward public procurement, whereby government and market parties collaborate and exchange on a deeper level. Such co-operation can be contractual or based on innovative procurement procedures.
- In-kind partnerships between the public and private sectors: in this model, public authorities have a number of assets and exploit them in exchange for commercial data, or vice versa, e.g. a private company offers data in order to receive goodwill or data in return.
- Mandatory data sharing: in this model, urban authorities exercise their power to oblige service providers to share data in order for them to receive certain approvals or permissions to operate in a city.
- Collaborations between authorities: cities work together with other (urban) authorities to exchange data, jointly procure data or build platforms, services or data standards.
- Crowdsourcing: whereby urban authorities collaborate with the public to collect data and information, check or improve data quality or even outsource some of their tasks to residents.

Source: Adapted from EIB (2021).

Which data acquisition pathways are selected by public authorities depend on a number of factors. Foremost among these are the types of public mandates held by governments and the policy objectives they target. The relative costs of different data acquisition pathways (both to governments and concerning the burdens imposed on individuals and firms) are important, as are the proportionality and balance of these costs versus the expected benefits to be derived from the data collection. Finally, public authorities' technical and human resource capacity to materially collect, process and manage data are essential elements to consider when selecting a data acquisition pathway.

When mandates for public authorities are clear and well-articulated – for example, the need to ensure high levels of road safety or to manage public spaces for the public good – then direct or conditional compulsion of data is warranted if the burdens they impose on reporting parties are in line with expected benefits. In these cases, purchasing data produced by regulated activities in the public realm (e.g. traffic

on streets) may be a poor alternative unless public authorities lack the ability or capacity to collect, process or manage the data themselves. Even in these cases, it makes sense for public authorities to increase their technical and organisational capacity to process and manage data so that they may carry out those tasks directly or become more proficient at managing contracts with the third parties they delegate to carry out these functions. In either case, public authorities should carefully consider whether data should be reported to public authorities, irrespective of whether data processing is carried out by the public authority or a third party.

Creating trusted and actionable insight from data held by mobility stakeholders may involve more than just the transmission of data to public authorities. Behind direct or conditional requirements to report data to public authorities is their need to use these data to deliver actionable and policy-relevant insights. These insights may be obtained in other ways than the direct transmission of data from stakeholders to authorities. For example, authorities can either encourage or require stakeholders to accept that vetted algorithms run on their servers to extract key indicators of relevance for a wide range of potential users. This approach could, for example, return aggregate density of pick-up and drop-off events at street segment level from ride-service, taxi and public transport operators to authorities concerned about traffic congestion and safety, without ever revealing sensitive data from operators (ITF, 2018). In these cases, the obligation is not to *transmit* data *to* the public authority but to *accept* limited remote processing of that data by, or on behalf of, the public authority. The concept and potential legal standing for such virtual "secure processing environments" are under consideration in the forthcoming EU Data Governance Act (EU, 2020).

Mandates, costs, alternatives and capacity are at the heart of an effective public authority data-reporting framework. The purpose of collecting data must be clear and articulated, the benefits of its collection must be weighed against its costs and burdens, alternative data collection frameworks should be considered, and the capacity for public authorities to materially collect, process or manage data must be enhanced.

Data collection pathways

Upstream of the data reporting to authorities, data itself is collected via distinct pathways. Data may be volunteered by a data subject, observed based on the data subject's actions and behaviours or inferred from the data subject's actions or behaviours (WEF, 2011).

Volunteered data includes data provided by individuals in order to use a product or access a service. These might include name, birthdate, payment information, access credentials (e.g. driver's licence or public transport subscription). These may also include reviews and ratings offered by individuals in the context of a service they have used.

Observed data relate to behavioural data automatically triggered and logged by a user – or, more specifically, by a sensor platform or machine used by a user (e.g. car, smartphone, GPS tracker, etc.). These data may include GPS tracks and logs of vehicle usage characteristics (acceleration, deceleration, speed, web page clicks and navigation data). The collection of observed data may be purposively initiated by a user action (e.g. a search request) or passively recorded in the background (Wi-Fi networks queried for status updates or connections). These data constitute observations of individuals' behaviours but do not include inferences based on these observations.

Inferred data relate to data that result from the purposive processing and transformation of volunteered or observed data that may still relate to a specific person or the device they are using. These data may include profiles derived from grouping and processing volunteered or observed data. This processing may use proprietary algorithms or group different data together to create new data insights. Inferred data

derived from personal data following processing by a data processor (which may also be a data controller in the sense of the EU GDPR) is subject to disclosures by the data controller of both the purpose to which it will be put and to the consent given by the data subject for this processing and use.

These terms and distinctions do not constitute legal definitions – there are some grey areas – especially concerning what constitutes observed data (European Commission et al., 2019). These are nonetheless useful and important concepts to bear in mind when considering data-reporting mandates. Inferred data are at the core of competition in data and digital markets – it is in "mining" these insights that market actors create intellectual property and leverage it to gain an advantage. Public authorities should, by default, not require the reporting of inferred data.

Data heterogeneity and risks

Data are heterogeneous and some forms of data bear unique and significant risks that must be managed. There are many ways to evaluate the heterogeneity of data types and flows, but perhaps the most relevant consideration is the risk associated with the collection, processing, storage and release. The risk to individuals is related to whether data are *personal* data and if personal if those data are *sensitive*. The risk to firms and economic actors is related to the *commercial sensitivity* of that data. The following sections address these two types of data-related risks.

Data-related risks for individuals: Personal and sensitive data

Personal data can be linked directly or indirectly to natural persons and thus poses the greatest risk to privacy rights. According to the OECD, personal data refers to "any information relating to an identified or identifiable individual (data subject)" (OECD, 2013a). This definition is further extended in Article 4.1 of the EU's GDPR as:

> …any information relating to an identified or identifiable natural person ("data subject"); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, [e.g. social security number] location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity [e.g. name and first name, date of birth, biometrics data, fingerprints, DNA…] of that natural person. (EU, 2016a)

This definition serves as a state-of-the-art reference as to what may constitute personal data.

Sensitive data are a subclass of personal data which calls for enhanced protection. For example, in the context of the GDPR, sensitive data are comprised of "data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation" (Article 9.1) and can also be considered to include "data relating to criminal convictions and offences or related security measures" (Article 10) insofar as these data trigger enhanced protections or processing restrictions (EU, 2016a).

Data may only indirectly reveal details relating to a data subject. As outlined in the GDPR definition, the exact nature of personal data is not fixed in so far as many data that have no direct personal identifiers (e.g. name, identity number) and yet may, nonetheless, be used to re-identify an individual. A number of quasi-identifiers exist that, either on their own or in combination with other data, serve to identify a natural person (ITF, 2015; 2016). For example, the extent of potentially identity-revealing data in the context of connected cars is described by the European Data Protection Board (EDPB):

> Even if the data collected by a connected car are not directly linked to a name, but to technical aspects and features of the vehicle, it will concern the driver or the passengers of the car. As an illustration, data relating to the driving style or the distance covered, data relating to the wear and tear on vehicle parts, location data or data collected by cameras may concern driver behaviour as well as information about other people who could be inside or data subjects that pass by. Such technical data are produced by a natural person, and permit his/her direct or indirect identification, by the data controller or by another person. The vehicle can be considered as a terminal that can be used by different users. Therefore, as for a personal computer, this potential plurality of users does not affect the personal nature of the data. (EDPB, 2021)

The EDPB guidance notes that vehicles may be used by different users and yet may still produce data that can be directly or indirectly linked to each unique user – the vehicle in this respect functions as a terminal. This implies the types of data outlined by the EDPB guidelines conceivably extend to other vehicles as well – including shared micromobility devices or other forms of shared mobility – especially as these services are offered to single (or a very limited number of simultaneous) users.

Box 3. Evolving location data precision

The ubiquity of precise location data enables many new services but raises significant privacy concerns. As location-sensing technologies and protocols have progressed, the ability to consistently locate a device (and the person operating it) has evolved from a range of a few hundred metres (using mobile telecom cell tower triangulation) to less than a few metres (with different global navigation satellite systems [GNSS], including GPS, Galileo, GLONASS or BeiDou). The newest methods can locate a device to centimetre precision, even inside buildings or in complex urban environments with various hybrid protocols, 5G signals or other sensor inputs including barometric, Bluetooth, video and audio sensors.

The most advanced of these methods are based on 5G New Radio positioning and enable precise, reliable and extremely low latency localisation, both horizontally and vertically, such that it will soon be technically possible to consistently track devices in near real-time, even in multi-story buildings with no line of sight to geolocation satellites (ITF, 2016; Kanhere and Rappaport, 2021; Keating et al., 2021). The current (v.16) 5G protocol (by the 3rd Generation Partnership Project [3GPP], the group of organisations developing mobile telecommunications standards) specifies several enhanced hybrid positioning protocols combining 5G and other technologies such as Wi-Fi positioning or GNSS-based localisation. Planned future releases (v.17 and v.18) will integrate highly precise location awareness into the core 5G protocol for industrial and other applications (Keating et al., 2021). The technical ability to precisely track devices that can be reliably associated with natural persons is developing at a much faster pace than the regulatory protections that would ensure privacy and the ability for people to control geolocation data relating to themselves. This misalignment may impact the public acceptance of new geolocation technologies as well as the guarantee of fundamental privacy rights.

Location-based data are well-understood potential indirect identifiers of individuals and their behaviour and create unique privacy-related challenges. Location-based data are commonly collected for a number of ever-expanding purposes and are increasingly precise – down to an accuracy of just a few centimetres (see Box 3). These data may be comprised of specific co-ordinate or spatial location data (e.g. a parked dockless shared bicycle or a record of a mobile device's location based on Wi-Fi signal strength) or may be comprised of a string of such data recording the space-time trajectory of a connected device. Even if the logged record of these signals is not associated with direct personal identifiers, they may reveal unique patterns that can be used on their own, or in combination with other data sources, to re-identify unique

individuals and reveal sensitive details based on their travel behaviour (Khoshgozaran and Shahabi, 2009; ITF, 2015; 2016; 2021a; Pignatelli et al., 2020; Rajashekar and Sundaram, 2020; Yang et al., 2020; Cunha, Mendes and Vilela, 2021; Errounda and Liu, 2021; Kim et al., 2021). Re-identification is a fundamental and consequential risk that must be managed, especially in the context of location-based data.

In considering whether data containing potential indirect identifiers are considered personal, it is necessary to take into account the effort needed to link that data to a natural person. Identifying a natural person from data containing potential indirect identifiers (e.g. location-based data, video or audio data, gait data) requires analysing and processing. This processing can be relatively straightforward (matching recurrent starting and ending locations of trips with known work or home addresses) or may, alternatively, require greater efforts and access to other restricted data (matching trajectory data with credit card transactions, using facial or voice recognition algorithms to identify individuals or isolating gait characteristics and matching these to identified natural persons). The degree of difficulty required to carry out such re-identification and its cost must be considered in determining whether such data effectively constitutes personal data given the probability of it being linked to a unique natural person. The EU's GDPR notes that:

> To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments... (Recital 26: EU, 2016a)

The risks of re-identification are increasing over time, even for what have previously been considered robustly de-identified data. The technical ease with which data can be linked to natural persons is steadily progressing, alongside the development of new re-identification techniques, especially those leveraging advances in artificial intelligence (Wang, 2017; Rajashekar and Sundaram, 2020; Cunha, Mendes and Vilela, 2021; Kim et al., 2021). This trend implies that authorities should adopt a strong precautionary approach regarding what should be considered "personal data". In practical terms, this suggests that the "reasonable means" re-identification tests outlined in Recital 26 of the GDPR should be expansive rather than restrictive (accounting for "all the means likely to be used"). In circumstances where a potential but technically or materially onerous re-identification pathway is identified, a prudent approach would be to consider such data as personal and thus subject to enhanced privacy protection. Here, it is important to make the distinction between original data, which may be used to re-identify an individual, and data that has already undergone some form of de-identification processing (e.g. anonymisation, obfuscation, etc.), as discussed further below.

Personal data that is strongly or irreversibly de-identified should no longer be considered personal and should thus be subject to different handling and processing protocols. The risk of re-identification of original data has spurred the development of numerous privacy-preserving mechanisms (PPMs) and processing protocols (see Box 4). Some of these are weak and do not significantly reduce the risk of re-identification (e.g. pseudonymisation), whereas others are increasingly sophisticated and produce de-identified data that are robust to even sophisticated re-identification attacks. Effective de-identification strategies should be linked to data types (e.g. structured, semi-structured or unstructured data) and reflect best-practice methodologies with respect to anonymisation, obfuscation and cryptographic techniques. Cunha, Mendes and Vilela (2021) provide a useful taxonomy for selecting appropriate data de-

identification PPMs (Box 4). From a data-reporting perspective, public authorities should have a good understanding of the robustness of various PPMs with respect to personal data.

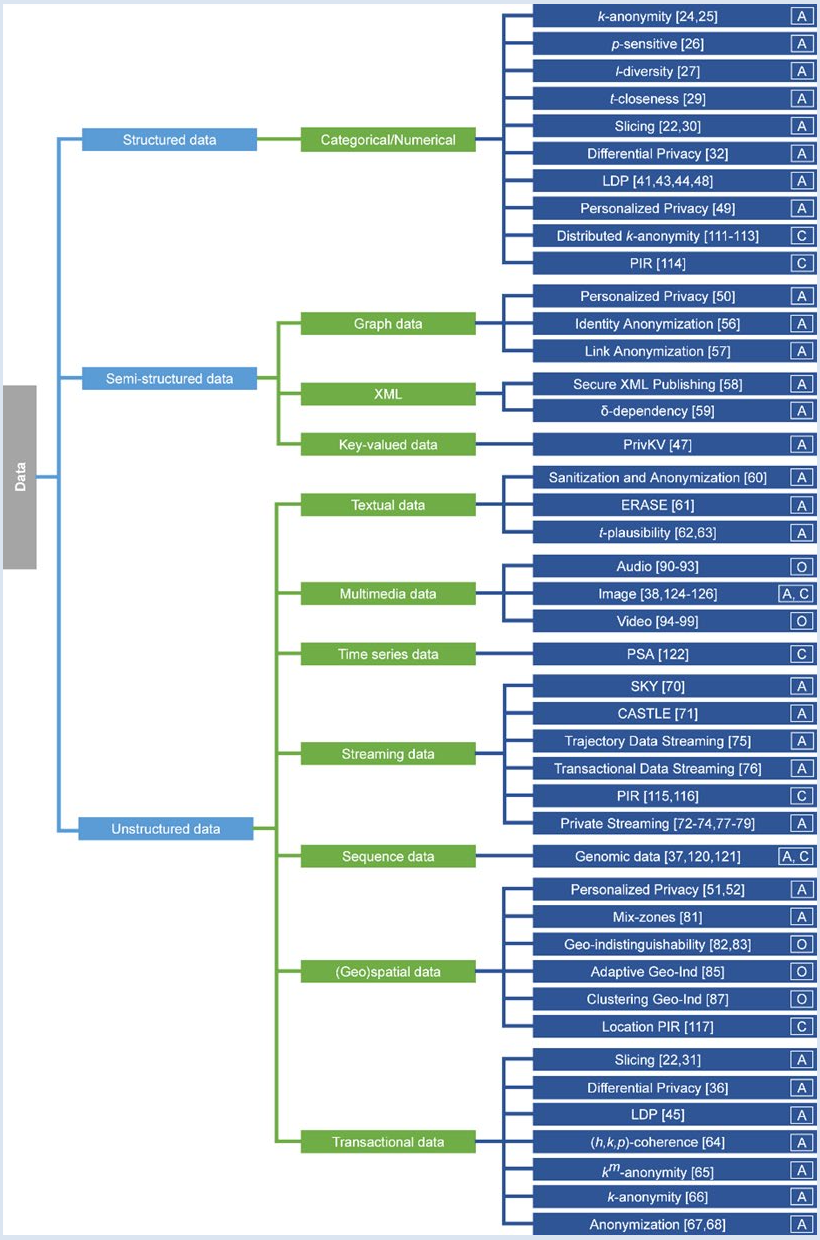
Box 4. Privacy-preserving mechanisms

The ability to associate data with the identity and behaviour of a specific person represents a serious challenge to privacy rights. Even data stripped of direct and even indirect identifiers may still be used to re-identify natural persons. Re-identification techniques are generally not trivial, but many are not difficult to apply successfully, especially as processing power increases and other sources of data that may be used to help re-identify individuals proliferate. In light of these risks, multiple forms of privacy-preserving mechanisms (PPMs) have been developed and applied. Just as data are heterogeneous, so too are PPMs – a privacy-preserving strategy for health data will likely not be effective or applicable for geospatial data. For this reason, privacy-preserving strategies must account not only for the data type (structured, semi-structured and unstructured) but also for data sub-type (text, numerical, streaming, geospatial, etc.). Generally, however, PPMs employ one, or a combination, of three broad approaches:

- anonymity-based approaches, which strip identifiers and group records into sufficiently indistinguishable sets
- obfuscation-based approaches, which introduce spurious records, synthetic records and data or generalise specific data (e.g. location data) to thwart re-identification
- cryptographic-based approaches, which employ computational cryptography and permissioned data record access to protect data.

Public authorities seeking to ensure personal data protection should assess if, when, and what type of PPM to apply to any personal data they receive or collect. Taxonomies of data types and appropriate PPMs – such as those illustrated below – are integral to the personal data-protection task and should be updated and re-assessed as PPM efficacy evolves and re-identification threats grow more sophisticated.

Figure 2. Personal data taxonomy and appropriate de-identification methodologies
(A = anonymisation-based, O = obfuscation-based, C = cryptographic-based)



Source: Adapted from Cunha, Mendes and Vilela (2021).

Personal data may be collected if warranted, but additional care is required in its processing and handling. The point of privacy laws is not to prevent the collection and processing of personal data but, rather, to set out how such data should be handled and processed so as to minimise privacy risks. Frameworks such as the EU's GDPR or California's Consumer Privacy Act and the California Privacy Rights Act expressly note that their purpose is to enable safe and privacy-preserving use of personal data. For instance, the GDPR sets out six lawful bases for collecting and processing personal data (EU, 2016a):

- when the data subject gives their **explicit consent**
- in order to meet a data subject's **contractual arrangements**
- in order to comply with a data controller's **legal obligations**
- to protect the **vital interests** of a data subject
- for purposes deemed to be in the **legitimate interests** of the data controller
- to allow tasks to be carried out in the **public interest**.

The legitimacy of the reporting of personal or otherwise sensitive data to public authorities is generally based on the need to carry out tasks in the public interest, to uphold contractual and legal obligations and, in certain circumstances, to protect the vital interests of data subjects (e.g. the protection of their life, health and livelihood). Nonetheless, these lawful bases for collecting, processing and reporting personal data should be subject to a standard of care in line with the potential risks to privacy.

Reporting of personal data should adhere to principles that ensure the highest level of privacy protection. For all of the reasons outlined above, a high standard of care must be followed when collecting, processing, reporting or transmitting personal data. This standard of care has evolved over time from the first set of principles outlined in the OECD's recommended Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data (issued in 1980 and most recently updated in 2013). Commonly accepted principles guiding the collection, storage, processing and reporting of personal data include the following (based on OECD, 2013; EU, 2016a; World Bank, 2019):

- **Lawfulness**: the collection, storage, processing and reporting of personal data should be conducted based on an established and known legal framework. The purposes and objectives underlying personal data collection should be defined and limited by law.
- **Purpose limitation**: the purpose and overall objectives of collecting personal data should be clearly identified, articulated and known to data subjects.
- **Personal data avoidance**: an assessment should be made as to whether personal data are necessary to carry out the purposes for which data is sought. If not, less privacy-sensitive forms of data should be used.
- **De-identification**: robust, effective and irreversible personal data de-identification techniques should be applied where the resulting data are still sufficient to achieve the lawful objectives.
- **Data minimisation and proportionality**: if personal data are considered necessary to carry out the stated objectives, only the minimum necessary data should be collected to carry out those tasks. Moreover, data collection should be proportionate to the stated purpose so as to minimise unnecessary data collection and "mission creep".
- **Consent**: data subjects must give explicit and meaningful consent to the collection and processing of their personal data. This consent must be linked to specific uses of that data and be renewed or re-stated for new data processing purposes – including processing by parties other than the original data controller.
- **Accuracy**: personal data should be accurate and kept up to date. Data subjects should have accessible pathways to identify inaccuracies and lapsed information. Data controllers and data processors should correct such information or otherwise exclude it from processing until corrected.

- Fairness and Transparency: the collection of personal data should be carried out fairly and transparently and should not give rise to representational biases.
- Storage, access and sharing limitations: personal data should be stored for a period explicitly linked to the purpose for which it is collected, and this period should be set out in law. Once that period has elapsed, personal data should be destroyed or de-identified in such a way that it no longer constitutes personal data (e.g. by aggregation). Access to personal data should be limited to only parties who have an identified and meaningful function in carrying out the objectives stated when the data was collected. The sharing and onward processing of personal data should be subject to the same constraints regarding lawfulness, purpose specification, data minimisation and proportionality, and consent as the original collection and processing.
- Accountability: the data subject and an independent oversight authority should monitor and assess the collection and processing of personal data in line with the preceding principles.

These fundamental guiding principles should be at the heart of public authorities' data-reporting policies.

Data-related risks: Commercially sensitive data

The second class of data risks are linked to the sharing or release of data deemed to be commercially confidential or sensitive. Several factors must be considered when reporting these types of data.

Personal data and commercially sensitive data pose different risk profiles and must be treated separately. The first thing to consider is that while all confidential data are sensitive, not all sensitive data are confidential. Thus, commercially sensitive data may be shared in certain circumstances with adequate safeguards. In addition, personal data collection, processing and reporting is typically set out in specific privacy laws (see Box 1); this is not the case for commercially sensitive data. Indeed, commercial information is not considered sensitive because it is protected by privacy laws but rather because it derives its value from its secrecy (Rosenblum and Maples, 2009).

There is no clear definition of what data are commercially sensitive – this determination depends on which industry or sector is being considered. Several countries' laws and codes describe commercial sensitivity by referring to "a trade secret" (Table 1), which has a technical definition. According to the World Intellectual Property Organization's definition (the same definition that the European Union uses), a trade secret qualifies: "what is commercially valuable, known to a limited group of persons, and subject to the use of confidentiality agreements" (EU, 2016b; World Intellectual Property Organization, 2021). The definitions in Table 1 help qualify the different dimensions of what constitutes commercially sensitive data.

Commercially sensitive data are valuable. Their value is positively correlated with the level of privacy. Such data contain information on the business, giving the holder a competitive advantage and including such things as operation-related data, algorithms, formulas, configuration, patterns, etc.

Commercially sensitive data often relate to parameters of competition such as innovation, price, quantity and quality. The publication of commercially sensitive data can affect the market by reducing the incentive to compete or by decreasing the independence of competitors when making decisions (OECD, 2011). As a result, the publication or disclosure of commercially sensitive data can also distort the market by promoting collusion or by prejudicing the competitive position of the initial data holder.

The commercial sensitivity of informational or operational data depends on the context and specific nature of collected and reported data. For example, the EU Commission's Delegated Regulation (EU) 2017/1926, concerning the provision of EU-wide multimodal travel information services, describes existing informational and operational data that should be shared for multimodal travel information services. The

EU Commission identifies two groups of data: static and dynamic. Each has three levels of services corresponding to their level of detail (European Commission, 2017).

Table 1. Trade secrets and commercially sensitive information according to different countries

Countries' acts referring to trade secrets	Excerpt referring to trade secrets
UK Freedom of Information Act 2000 Part II Exempt Information Section 43 – Commercial interests	"[…] Information is exempt information if it constitutes a **trade secret**. Information is exempt information if its disclosure under this Act would, or would **be likely to, prejudice the commercial interests of any person (including the public authority holding it).** The duty to confirm or deny does not arise if, or to the extent that, compliance with section 1 (1) (a) would, or would be likely to, prejudice the interests mentioned in subsection (2)."
Ireland's Freedom of Information Act 2014 Part 4 – Exempt Records, Section 36 – Commercially sensitive information	"[…] (1) Subject to subsection (2), a head shall refuse to grant an FOI request if the record concerned contains **trade secrets** of a person other than the requester concerned, **financial, commercial, scientific or technical or other information whose disclosure could reasonably be expected to result in a material financial loss or gain to the person to whom the information relates, or could prejudice the competitive position of that person in the conduct of his or her profession or business or otherwise in his or her occupation**, or **information whose disclosure could prejudice the conduct or outcome of contractual or other negotiations** of the person to whom the information relates."
US Freedom of Information Act, 5 U.S.C § 552	"[…] (b) This section does not apply to matters that are (1)(A) specifically authorized under criteria established by an Executive order to be kept secret in the interest of national defence or foreign policy and (B) are in fact properly classified pursuant to such Executive order; (2) related solely to the internal personnel rules and practices of an agency; (3) specifically exempted from disclosure by statute (other than section 552b of this title), if that statute • (A)(i) requires that the matters be withheld from the public in such a manner as to leave no discretion on the issue; or (ii) establishes particular criteria for withholding or refers to particular types of matters to be withheld; and • (B) if enacted after the date of enactment of the OPEN FOIA Act of 2009, specifically cites to this paragraph. **(4) trade secrets and commercial or financial information obtained from a person and privileged or confidential…**"

Not all technical or operational data will result in an anti-competitive situation if published. The OECD distinguishes commercially sensitive data from technical information (OECD, 2011). However, dynamic data, which relates to operational matters such as vehicle position and availability (National Transport

Commission, 2020), ridership (Johansen et al., 2019), and service disruptions (delays, cancellations, etc.) may, in certain circumstances, be considered commercially sensitive. Where this is the case, enhanced but adapted and proportionate protections should be in place in line with the framework outlined for the protection of personal data (e.g. in line with the principles outlined in the previous section).

Transactional information and other information regarding investments or service development are almost always commercially sensitive. Non-operational information related to customers, investments, operating cost structures, billing, and planned evolutions of services or prices is considered commercially sensitive since it relates to parameters of competition (European Commission, 2020a). If the reporting of these types of data is deemed necessary to achieve a lawful and clearly stated purpose, then adapted and proportionate data-reporting principles should mitigate potential competition risks.

As with personal data, new risks may emerge from the fusion of non-commercially sensitive data. For example, the profitability of a service, which is very commercially sensitive information, can be derived using service ridership and by estimating costs. Thus, data that is not commercially sensitive on its own can lead to the exposure of commercially sensitive attributes when joined or cross-referenced with other data. Several methods to reduce the risks posed by data fusion exist. As with personal data, data aggregation and the partial collection or obfuscation of data (e.g. space, time) may reduce the possibility of the inadvertent collection or release of commercially sensitive data (National Transport Commission, 2020).

Risks associated with commercially sensitive data collection and processing

The collection and processing of commercially sensitive data are associated with risks. These risks may lead actors to have objections to sharing or reporting that data – especially when the perceived costs of sharing or reporting that data outweigh the perceived benefits of doing so. According to the International Association of Public Transport (UITP), different types of risks are often associated with data sharing, although in some cases, these risks are equally relevant for reporting data to public authorities (UITP, 2020).

Data sharing or reporting comes with inherent risks relative to competitive position. Mobility operators may be reluctant to report data because they fear that they could provide competitors with insights into their organisation or business model (OECD, 2019; UITP, 2020). Operators, therefore, may view data reporting as potentially weakening their competitive position and this perception can negatively affect the operators' incentives to innovate (OECD, 2019).

Commercial operators fear that sharing or reporting sensitive data raises the risk of losing their ability to control how the data are used. Indeed, according to UITP, public transport authorities express concern that once data are made publicly available, it is almost impossible to control the way data are re-used (OECD, 2019; UITP, 2020). This risk is even more significant if the data are shared onwards among multiple actors and across different jurisdictions (OECD, 2019). This loss of control can have several consequences. One concern is that sharing or reporting data can lead to inaccurate uses or misuses. As a result, operators fear that while they provide high-quality and accurate data, this may be misused and misrepresented, leading to perceived inadequate quality of data which would reflect poorly on the company and cause reputational risk.

Operators can be reluctant to share or report commercially sensitive data because of security issues. Sharing data exposes an organisation to digital security threats (OECD, 2019; UITP, 2020) and this risk increases as data use grows. Furthermore, data mismanagement can also lead to accidental disclosure of commercially or privacy-sensitive information. Data breaches or leaks are associated with potentially substantial reputational and financial losses (OECD, 2019). In the case where laws or regulations protect

data from being publicly shared, the disclosure of commercially sensitive data could lead to fines due to privacy violations. This is the case with the disclosure of commercially sensitive data related to individuals (customer information, driver information, banking details) or to operations (path of trips) in many jurisdictions.

In 2015, the outsourcing deal made by *Transport Styrelsen*, the Swedish Transport Agency, with IBM Sweden led to the accidental leak of its entire database (Borg et al., 2018). Alongside classified and defence-related data, information related to registered vehicles, toll-paying motorists, train conductors or air traffic controllers were leaked (New York Times, 2017). Following the data leak, the Swedish government decided to restrict the outsourcing of sensitive data in 2017. This decision was followed by the implementation of a management model that makes background control systematic for personnel in administration in charge of data storing, network management, server administration and application management (Transport Styrelsen, 2017).

Data-reporting risk-assessment framework

Appropriate risk-assessment frameworks enable adapted and proportionate data-reporting requirements and policies. Risk is a significant obstacle to data sharing and data reporting. A survey among transport authorities undertaken by the UITP highlighted that 50% of respondents identified data privacy and liability risks as their main challenges, while 24% of surveyed organisations said competition-related issues were a key challenge when sharing data (UITP, 2020). These results illustrate that not all risks are equally perceived and that risk perception is subjective. Two essential components determine the way risk is perceived (ISO 31000:2018, 2018):

- the likelihood of an event occurring (from improbable to probable)
- the severity of consequences if the event occurs (from acceptable to intolerable).

These two components characterise risk levels. A situation will be risky when there is a high likelihood of the event occurring and when the consequences of that occurrence would be severe. Conversely, if an event is very improbable and would have a limited impact, the risk will be perceived as low (Figure 3). Thus, different data will not be associated with the same level of risk. In addition, this perception will vary among persons and organisations. Indeed, risk perception has two dimensions (Paek and Hove, 2017):

- the cognitive dimension, which refers to the understanding and knowledge of risks by a person
- the emotional dimension, which relates to how people feel about the risk.

Figure 3. Risk-assessment matrix

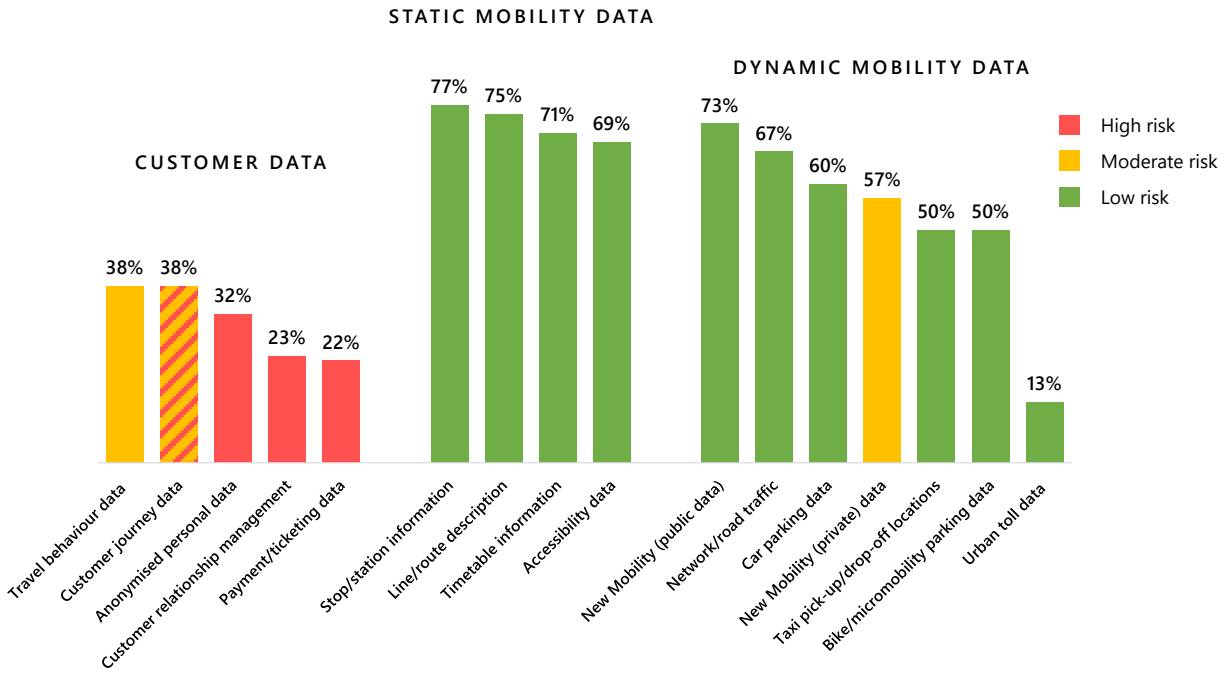
RISK RATING KEY	LOW	MEDIUM	HIGH	EXTREME
	ACCEPTABLE	ALARP – (as low as reasonably practicable)	GENERALLY UNACCEPTABLE	INTOLERABLE
	OK TO PROCEED	TAKE MITIGATION EFFORTS	SEEK SUPPORT	DO NOT PROCEED

LIKELIHOOD	SEVERITY: ACCEPTABLE LITTLE TO NO EFFECT ON EVENT	TOLERABLE EFFECTS ARE FELT, BUT NOT CRITICAL TO OUTCOME	UNDESIRABLE SERIOUS IMPACT TO THE COURSE OF ACTION AND OUTCOME	INTOLERABLE COULD RESULT IN DISASTER
IMPROBABLE RISK IS UNLIKELY TO OCCUR	LOW - 1 -	MEDIUM - 4 -	MEDIUM - 6 -	HIGH - 10 -
POSSIBLE RISK WILL LIKELY OCCUR	LOW - 2 -	MEDIUM - 5 -	HIGH - 8 -	EXTREME - 11 -
PROBABLE RISK WILL OCCUR	MEDIUM - 3 -	HIGH - 7 -	HIGH - 9 -	EXTREME - 12 -

Source: U3115299 (2017) (CC BY-SA 4.0).

In this regard, the perception of risk will vary depending on the type of actor, as well as their knowledge and understanding of risks, which is correlated to their experience. According to a UITP survey, transport organisations perceive sharing or reporting of data related to customers as risky but classify static mobility data, the most commonly shared data category, as “low risk”. Finally, UITP noted the heterogeneity of modes listed in the dynamic data category, which could explain why this category is seen as “low risk” (see Figure 4). Indeed, dynamic data related to public transport, which is commonly shared, is unlikely to have the same level of risk as car-related or other vehicle-specific data.

Figure 4. Perceived risks associated with data sharing and reporting as identified by members of the International Association of Public Transport



Note: The percentages represent the percentage of responding organisations to the UITP survey who identified with the indicator risk level. The risk level is depicted based on the risk level selected by most respondents. As not all indicators are relevant to every organisation, the base size varies depending on the number of respondents for each particular indicator. The customer journey data indicator is bi-coloured because there were two most chosen risk levels selected by respondents.

Source: UITP (2020).

Public authorities can play a substantive role in mitigating privacy and commercial risks associated with data sharing and reporting, as well as the perception of these risks. They can do this by clearly stating why they may incentivise or mandate data sharing and reporting. They can also do this by plausibly and meaningfully ensuring that stakeholders' data will only be used for these purposes. These strategies may better enable stakeholders to understand when the collective benefits of data reporting may outweigh the specific burdens imposed and thus build trust between public authorities and entities reporting data to them (UITP, 2020; OECD, 2019).

Public authorities can reduce data-reporting risks, and the perception of these risks, by adopting proportionate and transparent data-collection and handling protocols, which ensure the protection of personal and commercially sensitive data. These protocols should build on clear principles that ensure that risks are identified, avoided or otherwise mitigated by design and by default. These principles relate to purpose specification and limitation, avoidance of the collection of personal or commercially sensitive data where possible, data minimisation and robust de-identification and strict data access, processing and retention policies.

Data semantics, schemas and reporting syntaxes

Data reporting necessarily entails some burden on the part of reporting parties. These burdens stem from the need to collect, synthesise, format, ensure quality, transmit and document compliance with data-reporting requirements. These burdens should be minimised as much as possible without compromising the data-reporting objectives. Specifying and enabling the use of common data-access methods, semantics, schemas and syntaxes help mitigate regulatory compliance burdens.

Some form of data harmonisation can improve regulatory efficiency. Common or compatible data-reporting formats have emerged among established mobility services, such as public transport. More recently, however, the rapid expansion of digitally enabled mobility services and the proliferation of bespoke data syntaxes pose challenges to public authorities' ability to monitor and intervene where necessary. Moving towards a common understanding of terms (semantics), data structure (schemas) and data reporting with specific machine-encodable and readable formats (syntaxes) will improve the ability of public authorities to carry out their mandates.

Some data syntaxes encode personal or commercially sensitive data, and therefore reporting requirements that mandate the use of these syntaxes must be accompanied by appropriate risk avoidance or mitigation policies. Data semantics usefully structure how data are encoded and communicated. On their own, they are neutral constructs that allow their users to extract insights from data and communicate these to other parties. Which data are encoded, how they are processed and to whom they are communicated may entail privacy or commercial sensitivity risks. These risks can be avoided by preventing such data from being encoded by default (e.g. by not allowing individual trips to be expressed in the syntax) or minimised by design – for instance, by specifying and building-in risk management protocols (e.g. by defining protocols that aggregate personal data and only retain aggregates, destroying the original data).

Data semantics

Harmonised or compatible data semantics improve the ability of public authorities to make sense of the data they collect. Much as with human language, digital systems build on a shared understanding of the semantical building blocks of language – words and terms. However, unlike human language, which is often open to nuance and interpretation, machine language requires a clear, consistent and unequivocal definition of terms and meanings.

The first step in building standard mobility data reporting is to create a common understanding of this semantical lexicon. While clarity on the meaning of terms may be settled within each transport operator's own data architecture – for example, a public transport operator will have a consistent definition of a bus stop or what it means to say a passenger has commenced a trip – this may not be the case across other transport operators and with public authorities. In terms of new mobility operators, where little harmonisation of data terms has taken place, multiple definitions may exist for such basic information as "is an asset available" or "has a trip ended". Even across public administrations, multiple definitions may

exist for the same term (e.g. what comprises a parking violation). Improved regulatory oversight and enforcement require convergence on these terms.

Accepted semantic models exist for established mobility services but are still lacking for many new mobility services. Mode-specific semantical models exist for public transport and serve as the basis for the data syntaxes used to promote interoperability and reporting within those services. This is rarely the case for other services, and simply adapting the former to the latter may not prove an attractive option for new market entrants as this may not capture the specificities of their services.

Developing and incentivising or requiring the use of such a lexicon will improve interoperability and remove uncertainty as to whether public policy outcomes are being met. In order to achieve the broad uptake these semantic building blocks require, it seems appropriate that they are developed at the highest level, adopted by a wide number of actors and deployed widely within and amongst countries. This argues for voluntary development and incentivised deployment via traditional standard-setting processes. However, this will take time, and there is no well-defined broad initiative to do this. In the meantime, market actors and authorities can incentivise adherence to a set of emerging semantical models that at least provide some form of convergence around the meaning of terms.

Four examples of these are the OSLO-Mobility semantical model, SAE's Mobility Data Collaborative Data Sharing Glossary and Metrics for Shared Micromobility, the Mobility Data Specifications (MDS) Metrics application programming interface (API) and the semantical lexicon embedded in the EU's Transmodel data schema (described in the next section).

The Open Standards for Linked Organisations (OSLO) semantical model was developed in the Flanders region of Belgium to address the need for shared definitions and terms in support of the digital exchange of data in the domains of contact information management, localisation and public services (Van Roy, 2020; European Commission, 2021a; Flanders Department of Mobility and Public Works, 2021; OSLO, 2021). It seeks to facilitate semantic and technical interoperability through an open process amongst market actors and authorities, maintain these vocabularies and ensure that rules and governance principles are respected. The mobility component of the OSLO semantical model – OSLO-Mobility, released in its initial version in April 2020 – defines a common vocabulary to be used to exchange data about trips performed by people and the mobility services they have at their disposal. Specifically, OSLO-Mobility establishes a lexicon referring to traveller information, trip information, booking actions, network description, service supply on the part of operators and information relating to licences to operate.

SAE International (a US industry-based standard-setting body, formerly the Society of Automotive Engineers) convened a broad range of actors and authorities developing or seeking to regulate micromobility services in order to develop a shared understanding around terms used in the nascent industry. The Data Sharing Glossary and Metrics for Shared Micromobility (MDC, 2021) is composed of a standardised set of definitions and methodologies covering commonly used terms and indicators. These terms include "non-operational vehicle" or "maximum average number of vehicles available in a given territory". Disambiguation of these and other terms helps deliver more consistent reporting and monitoring of these services.

Another similar approach is being built into the current version of the Mobility Data Specification (MDS 1.1.0. – described below) and relates to standardised semantical models for reporting on MDS data via a proposed MDS Metrics Application Programming Interface (OMF, f2020). This API sets out standard definitions and parameters for calculating commonly used metrics based on MDS data. It builds on a standard set of semantical terms by defining a set of measurement outputs that enable consistent interpretation of data.

Data schemas

Data schemas can bridge the gap between completely bespoke data syntaxes and single harmonised data syntaxes. The strongest level of data harmonisation stems from the use of accepted (or imposed) standards and data syntaxes. These may be set by standard-setting bodies or *de facto* standards set by a firm or by consortia of dominant market actors. Public authorities may also set standards for data syntax for data-reporting purposes. Setting standards, even when these are evolutionary, may close out innovation and impose costs related to complying with the standard – especially for small- and medium-sized operators who have already built their data architecture around a specific standard. A less restrictive approach may be to ensure that there is broad functional alignment between the different standards used by actors in the market. This functional interoperability may be delivered by market actors agreeing to (or being required to) conform to a common data schema rather than a specific standard (ITF, 2021b).

The difference between a data standard or syntax and a data schema may be illustrated by analogy to the design of a house. A standard or syntax defines the function, size and disposition of every room in the house as well as their fittings – a schema defines only the functional attributes of the house (e.g. a house should have a bathroom, a kitchen, a living space, etc.). In the context of data reporting, these specific attribute "bins" relate to different data collection purposes (e.g. zonal origin-destination mapping, safety monitoring, speed monitoring, parking management, etc.). These can be seen as the basic functional outcomes that any mobility operator should track and report to reduce burdens and improve the ability of public authorities to carry out their mandates (ITF, 2021d).

The Public Transport Reference Data Model (EN 12896) – known as "Transmodel", is an example of a data schema (with an accompanying semantical lexicon) that encompasses common public transport concepts and data structures that can be used to build public transport-oriented and other mobility services in Europe (CEN, 2021a). These services include journey planning, routing, operations management and on-trip data services. The current version of Transmodel (v6) covers conventional public transport services (including demand-responsive services) as well as other mobility services (including taxis and vehicle rental, sharing and pooling). The model has been developed explicitly to facilitate interoperability amongst and between operators' and public authorities' information processing systems.

Transmodel is a reference model – its full or partial adoption is not compulsory – but it serves as a useful schematical reference for data-reporting standards whose use may be stipulated in data-reporting mandates (e.g. data reporting using the Network and Timetable Exchange [NeTex] syntax to assess the on-time performance of public transport services). In particular, Transmodel's management information and statistics sub-model (EN 112896-8) provides the basis for developing uniform data reporting for public transport and other mobility services. This sub-model provides "data semantics and structures of the raw data to provide indicators" (CEN, 2021b), some of which may potentially reveal personal data (e.g. origin-destination matrices for single-passenger services such as carsharing or bikesharing) or commercially sensitive information (e.g. load factors, commercial speeds) if embedded in data-reporting syntaxes.

Data syntaxes

If the semantical basis for mobility data are analogous to words in human language, the digital data syntaxes deployed to support those services are akin to the rules of grammar. They provide the structure in which the building blocks of language are organised to communicate meaning and trigger action. Again, there is little room for interpretation in machine language. Therefore, specifying a data syntax that enables

communication, or finding an efficient way to translate meaning from one syntax to another, is a core concern in the deployment of digital services.

At present, there is no universal mobility data syntax, either from an operational basis or from a data-reporting basis. Public Transport Authorities may require data reporting from public transport operators in one format; taxi and ridesourcing oversight bodies may require reporting from operators in a different format; parking authorities in yet another format; and agencies in charge of shared micromobility in yet another format. These different formats may include analogue and digital elements and may be only partially composed of machine-readable data – if at all. This hampers the uptake of mobility service integration, and it may also give rise to asymmetries in power within the market if those standards that are proposed, or imposed, favour certain operators over others, whether by design or in practice. For this reason, there has been a generalised call for the deployment of open and mode-agnostic data syntaxes.

The specification or adoption of a data syntax for data reporting may reduce burdens, but it may also create risks with respect to personal or commercially sensitive data. Enhanced data-protection protocols are called for if the specification of a data syntax includes or requires encoding data that, by its nature, is privacy or commercially sensitive. The first of these protocols is to determine if the collecting or reporting of such data is necessary to carry out the objective for which the public authority requires the data. When data syntaxes are designed expressly for reporting purposes, their structure, reach and internal logic should include personal data protection by default and by design. Data-handling protocols that minimise risks associated with collecting and transmitting commercially sensitive data should also be fully integrated into the syntax.

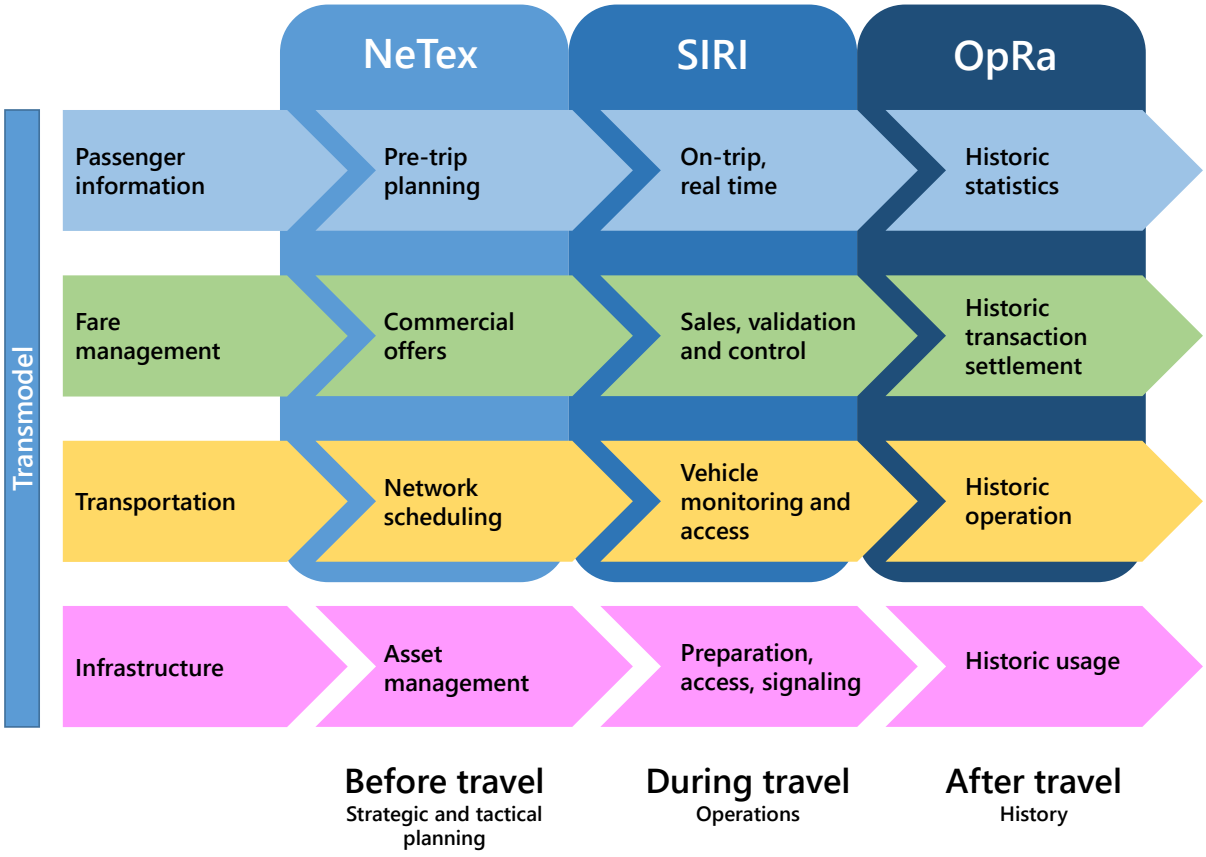
Public transport data syntaxes

Public transport data syntaxes generally do not encode personal data, but privacy data risks emerge as standards shift to cover more individualised services. The public transport industry has developed data syntaxes that enable the exchange of information in support of informational, operational and transactional integration amongst public transport operators within and between regions and countries. The development and use of these syntaxes and standards typically focus on collective transport services which are aggregate by their very nature. That is because, in the absence of specific data relating to individual passengers, public transport vehicle movements include many indistinguishable passengers. However, when syntaxes and standards start to encode more individualised trips and related trip information – or single-passenger vehicle movements (e.g. with shared micromobility or taxi/ridesourcing) – personal privacy risks emerge, which should trigger enhanced data-protection efforts.

Public transport data syntaxes include: NeTex, a standard for sharing public transport schedules and related data; Service Interface for Real-Time Information (SIRI), a syntax for exchanging data on planned and current (real-time) services; and Operating Raw Data and statistics exchange (OpRa), which focuses on raw data to being collected, exchanged and/or stored to support the study and control of public transport services (see Figure 5).

Another public transport-oriented standard, the General Transit Feed Specification (GTFS), is a syntax designed to enable the outward-facing sharing of data related to scheduled or real-time public transport operations. Unlike NeTex, SIRI and Transmodel, GTFS was designed solely to help share information about the state of services and not to support operational linkages among operators.

Figure 5. Scope of different European public transport standards



Note: NeTEx = Network and Timetable Exchange, SIRI = Service Interface for Real-Time Information, OpRa = Operating Raw Data and statistics exchange.
Source: Adapted from Knowles (2019).

Network and Timetable Exchange

Network and Timetable Exchange (NeTEx) is a European Committee for Standardization (CEN) standard for exchanging public transport-related information. It is intended to be used to exchange static data between public transport stakeholders – this means that the data primarily helps describe the offer of public transport services and the associated infrastructure, rather than the current running status (CEN, 2020). NeTEx is a multipart standard composed of information related to network topology (e.g. routes, stops, lines, networks, geographic elements, etc.), scheduled timetables (e.g. passing times, calendars, type of days, etc.), fare information (e.g. prices, fare products, access rights, etc.), and passenger information (Bourée et al., 2019; CEN, 2020).

In 2017, the European Commission made NeTEx the relevant standard to exchange static travel and traffic data (e.g. services' static location, plans, schedules, fares) related to public transport, long-distance bus, and maritime to be used in the context of National Access Points, which were established to underpin the digital market in transport services by centralising and facilitating access to interoperable data related to transport services (European Commission, 2017).

Current developments around NeTEx are looking to extend the standard to alternative modes of transport, as requested by the EU Commission Delegated Regulation (EU) 2017/1926 (CEN, 2020). This fifth part will be dedicated, but not limited to, the description of infrastructure related to alternative modes

(e.g. carsharing, carpooling, car rental, cycle-sharing systems, cycle rental, etc.). The extension of the standard will modify or add a few attributes, including the following:

- the type of mode operation (e.g. pooling, sharing, rental)
- location data regarding meeting points for new modes (e.g. floating or fixed pick-up and drop-off points, carpooling areas)
- extension of the notion of a single journey for unique trips that may be planned only a short time ahead (e.g. a ridesourcing trip)
- extension of the vehicle path (route) attribute to cover the operating profile of new modes
- parking locations for new modes.

The planned NeTex extension to alternative modes has not been validated, although there is potential for leakage of personally identifiable data for new modes whose transcription into NeTex would reveal unique trips. This could be the case for information about services that are not scheduled or fixed but generated in real-time. For instance, the description of a ridesourcing journey or taxi service could encompass a unique floating start/stop trip location associated with a single journey identifier and unique vehicle path data.

Service Interface for Real-Time Information

Service Interface for Real-Time Information (SIRI) is a CEN standard that allows the exchange of planned, real-time, and projected data related to public transport performance (e.g. location of vehicles, connections, and schedules). Although SIRI was developed as a European standard, it has since been implemented by many transport authorities and is not limited to European countries. However, it is less common in other continents than other syntaxes, such as General Transit Feed Specification Realtime (GTFS-RT) or NextBus (APTA, 2013).

In common with NeTEx, SIRI is an extensible standard in the process of incorporating new mobility services. In particular, work is underway to extend a component of SIRI's syntax – the Facility Monitoring Service (EN 15531-4). This extension will provide real-time information on available vehicles (micromobility, shared cars, ridesourcing/taxis or any other type of vehicle) at a specific location, real-time information about available spaces to return vehicles and updated information about the location of a "facility", which includes among other things, the location of a free-floating vehicle. The intention is not to follow a particular vehicle or service but to know where an available free-floating vehicle may be located (CEN, 2020). Personal data risks are low as long as these data only consist of vehicle or facility counts (e.g. the number of vehicles only and do not relate to specific direct or quasi-vehicle identifiers). However, if these data are comprised of persistent vehicle identifiers, then personal data risks exist that warrant enhanced data protection. For instance, if a unique vehicle is seen to switch to "unavailable" status and later appears at another location as "available", then a specific trip can be inferred and may potentially reveal personal data.

Operating Raw Data and statistics exchange

Operating Raw Data and statistics exchange (OpRa) is a CEN initiative to develop a technical specification or European norm to help support public transport data gathering, exchange or storage in order to facilitate the study and control of public transport services. OpRa's focus is on actual and measured data, which describes the logged reality of public transport operations (e.g. delays, passenger counts, etc.) built

from Transmodel's "Operations monitoring and control" and "Management information and statistics" domains (see above). OpRa is designed expressly for data reporting of individual measurements at specified sampling intervals or in aggregate (e.g. statistics) (CEN, 2018). While aggregated and other forms of statistical data bear few personal privacy risks, this is not the case for some forms of raw data representing individual measurements. The latter data types warrant enhanced data-protection protocols.

General Transit Feed Specification

The General Transit Feed Specification (GTFS) is a simple (Comma Separated Value – CSV – format) yet highly documented way to view, edit, and share static data on public transport services available at any location. GTFS became a globally used data format to describe public transport fixed-routes (Antrim and Barbeau, 2013). In 2012, the development of GTFS-realtime (GTFS-RT) complemented the initial specification and allowed public transport agencies to provide real-time updates on their networks.

GTFS, unlike NeTEx, SIRI and Transmodel (see above), was designed to outwardly share information about the state of public transport services. Thus, it does not support operational linkages among operators (ITF, 2021c), especially in a cross-border context. In addition, although the specification is evolving, it is not designed to describe many new and emerging mobility services (unlike the evolution of GBFS, MDS and the planned extension of NeTex, for example).

The open nature of GTFS fostered the development of new applications, such as multimodal trip planning, mobile apps and real-time information. Furthermore, it led to the development of analysis tools for planning, network visualisation, etc. (Antrim and Barbeau, 2013). In addition to new public transport-related services, GTFS provided public transport agencies with new possibilities to improve their efficiency (Catalá, Downing and Hayward, 2011; Fortin, Morency and Trépanier, 2016). Catalá, Downing and Hayward distinguish two levers for efficiency improvement: service evaluation and performance measures (Catalá, Downing and Hayward, 2011).

GTFS can help agencies evaluate their performance in terms of availability, frequency and coverage (Table 2), but this evaluation requires supplemental data not covered in the specification. Specifically, the original specification does not contain real-time data (location of vehicles, network alerts, etc.) or situational data (passenger loads, etc.). Combining GTFS with these other data can improve insights. For example, the Oregon Department of Transportation released a proof of concept showing how GTFS data helped the Department's Public Transit Division better understand how public transport networks in the state were structured and how the networks were correlated with the population (Porter, Kim and Ghanbartehrani, 2014). The development of GTFS-RT allows for low-latency performance tracking and improved service quality evaluation, but passenger and other non-GTFS data are still required for more comprehensive performance measurement.

The GTFS feed does not contain personally sensitive information since the data relates to collective transport vehicles and fixed stops. Nor does it contain commercially sensitive information since the data relates to publicly available service and route information (Antrim and Barbeau, 2013). In contrast, GTFS-RT relates to several observable, but not necessarily publicly available, data points such as service alerts (event affecting a route, station or the network, stop moved, etc.), vehicle positions (location, congestion level, etc.), and trip updates (cancellations, delays, changed routes, etc.) (Barbeau, 2018). These have the potential to reveal some data that may be construed as commercially sensitive (commercial operating speeds, for example) or privacy revealing in certain circumstances (e.g. pick-up and drop-off locations for on-demand transport in low-density settings). These risks are low but increase as GTFS-RT data describes increasingly personalised transport services.

Table 2. Opportunities to leverage General Transit Feed Specification datasets to evaluate transit system efficiency.

Service evaluation metric	Evaluation	Measure	Comments
Service availability	Service area characteristics	The measure of the extent a transit route or system serves the locations where potential customers live.	Requires the use of an application such as GIS or a spatially enabled database for evaluation.
	Service coverage	The extent that a route or a transit system serves the population.	GTFS does not contain population data. However, it contains data that can be used to evaluate the spatial coverage of transit services in combination with GIS.
	Time and distance calculations	The percentage of the population that can access a certain area within a set period.	Temporal information contained in GTFS can provide additional insights to measure service availability.
Route layout and design	Stop location and spacing	Distance between different stops on a route.	GTFS can be used to evaluate the distance between stops in relation to spacing standards.
	Route or service directness	The variation in distance between the route and the shortest path possible.	GTFS contain data on the location of the beginning and the end of a route. Using a GIS app can calculate the travel time between these two points.
Travel time and capacity	Service frequency	The number of times a bus stop is served within a set period.	By using time and trip information contained in GTFS, it is possible to measure the number of times a bus stop is visited within a period.
	Number of transfers	The number of transfers by route.	The most direct transit services have the least number of transfers and this can be evaluated by calculating the total number of transfers between two points through GTFS data.
	Span of service	The number of hours of service provided by route, system or period.	Service span can be evaluated by using GTFS and calculating the minimum and maximum stop times for a route or transit system.

Source: Adapted from Catalá, Downing and Hayward (2011).

Non-public transport oriented data-sharing and reporting syntaxes

For services other than public transport, alternative data-sharing and data-reporting syntaxes have emerged and are increasingly being adopted by mobility operators and specified in public authority data-reporting mandates. These include the General Bikeshare Feed Specification (GBFS), the MDS and the developing City Data Standard for Mobility (CDS-M).

General Bikeshare Feed Specification (GBFS)

The General Bikeshare Feed Specification (GBFS) is an open-data standard, initially designed for docked bikes, that was then extended to shared-mobility services – currently including docked and free-floating bicycles, e-push-scooters, mopeds and car-based services (Mobility Data, 2021). GBFS is a stand-alone data

syntax, although elements of GBFS are incorporated into other data syntaxes – most notably the MDS (see below). As with GTFS for public transport, the main purpose of GBFS is to provide shared-mobility related information to end-users by powering travel trip planning services. To do so, shared mobility operators use this specification to publicly share real-time or near real-time data publicly. The standard is explicitly designed to provide open and public access to data regarding shared services, and thus GBFS APIs should be "freely available on the open internet and require no API key, token, or other means of access or authentication" (Mobility Data, 2021). GBFS is not designed or intended to provide historical data (e.g. trip records). Nonetheless, GBFS data are not without privacy concerns.

The potential for personal data re-identification is greatest with floating-vehicle trip data, as opposed to station-based trips, since the floating-vehicle trip record will more closely fit the traveller's actual starting and stopping points. This risk emerges from the ability to infer unique vehicle and trip data (which, as noted earlier, are quasi-identifiers and should be considered personal data)(Gauquelin, 2020). This inference risk operates as follows: GBFS only provides information on devices that are currently available in the system, which means that data on in-use vehicles or disabled vehicles is not exposed via the GBFS API. However, several studies have raised the possibility of deriving trip information from raw GBFS datasets (Xu et al., 2020). McKenzie (2019) identified trips as the difference between a device's attributes (e.g. time, location) before it last appeared available and after the same device appears available again (McKenzie, 2019). This inference method required a persistent device identifier over time. In 2020, MobilityData, the technical steward for GBFS, released a new version of the specification to reduce the exposure of private data (MobilityData, 2020) and now recommends using unique vehicle identification (ID) for every trip. This means the same device is now required to have different identifiers in the GBFS feed before and after the rental is complete, thus making existing trip-inference methods less effective.

However, according to Xu et al. (2020), even if methods such as resetting the ID (e.g. changing ID after a trip is completed) or dynamic vehicle ID (e.g. changing ID according to a defined time interval) mitigate privacy concerns, it is still possible to accurately infer information on trips thanks to specific algorithms. When trip inference is possible using reasonably available means, the resulting trip data should be considered personal data and thus trigger enhanced protection measures. Increasing the refresh rate for an ID modification can make information inference harder, thus improving privacy. However, this would make it harder for cities to gain information on shared-mobility use. Therefore, there is a balance between GBFS's utility and privacy concerns.

Mobility Data Specification (MDS)

The Mobility Data Specification (MDS) is a data standard and API specification currently configured for shared micromobility services (OMF, 2021a). MDS was originally developed for the City of Los Angeles in order to manage and regulate the deployment of shared micromobility services. However, the longer-term vision is that MDS is extended to all mobility services (shared micromobility and other forms of shared mobility, taxis, ridesourcing, urban aerial systems, freight and logistics, and autonomous vehicle-based services) (OMF, 2021b). Exploratory work is underway already to extend and develop MDS capabilities to incorporate taxi and sidewalk robot-based services (OMF, 2021c; Henry, 2021).

MDS is managed by the Open Mobility Foundation (OMF) – a not-for-profit, open-source software foundation that provides the governance framework for the specification (OMF, 2021b). MDS was developed to facilitate two-way communication from regulated entities to a regulator and from the regulator to regulated entities. The core motivation was to develop "a set of digital tools for public entities

… to manage the public right of way through data, APIs, and the dynamic application of policy" (OMF, 2021b).

The specification is a way to implement data sharing, monitoring, and communication of regulatory intent for public authorities and mobility service providers. It is explicitly designed to support many public governance tasks, including policy development and strategic planning, asset and infrastructure planning, transport operations and enforcement actions. As such, MDS is intended to be a toolkit that provides a digital equivalence to other standards and specifications for physical assets, such as stop signals and other traffic control devices, road and street signs and other physical means of conveying how the public right of way may or may not be used (OMF, 2021b). Public authorities increasingly require MDS to be adopted by mobility service providers in return for receiving operating approval for shared micromobility services – especially in, but not limited to, North America.

At present, MDS comprises three core components (see Table 3):

- Provider API is implemented by mobility service operators. It allows public authorities to access recent historical data or a snapshot of vehicle status held by the mobility service operator. These data allow the public authority to monitor compliance, adjust licensing terms, or plan based on past activity.
- Agency API is implemented by regulatory agencies. It is a gateway for mobility service operators to submit information to authorities regarding real-time operations and service delivery and enables agencies to dynamically manage public rights-of-way.
- Policy API is implemented by regulatory agencies and allows rules and regulations to be communicated in machine-readable formats that are directly ingested into mobility service providers' back-office systems.

Table 3. Mobility Data Specification core application programming interfaces

	Agency API	Provider API	Policy API
Purpose	Real-time management of public space	Planning based on asset and infrastructure use, operations planning and compliance monitoring	Publication of machine-readable and possibly dynamic regulations regarding the use of public rights of way
Data Flow	Provider pushes data to the public authority ("agency") Hosted by the agency	Agency pulls data from a mobility service provider ("provider") Hosted by the provider	Provider pulls policy information from an agency Hosted by the agency
Key benefits/features	Designed for real-time data collection Agency maintains an authoritative database of information reported by all providers Designed to support real-time analysis and adaptive regulation Available reference implementation and data auditing and verification tools	Designed to provide recent historical data and a snapshot of vehicle status Easier to use Lower IT complexity Some historical data are available on request, which may reduce the need for agencies to store data Many available commercial and open-source analysis tools	Allows agency to publish geography-based regulations (e.g. restricted riding/parking areas, vehicle caps, etc.) Providers can automatically adjust their services/apps as policies change Removes the need to manually communicate policy changes to providers Works via a static webpage or as a dynamic API

Drawbacks	More complex to implement Agency must operate or procure IT systems capable of handling real-time API calls Data only published once, not available for re-ingestion Agencies must store any data needed for future analysis or reporting Fewer software vendors currently providing processing solutions	Harder to scale as datasets get larger Agency needs to query each provider individually Availability of all necessary historical data from providers is not guaranteed Not intended to provide real-time data about events in the right-of-way	Some complex policies or rules may not be supported by API Newer MDS API has limited software tools available to implement Need to set provider expectations for how and when policies will change
Comprises personal or commercially sensitive data	Yes Data relating to uniquely and persistently* identified vehicles can be considered to be personal data, as are data relating to single trips (including start and end points and trajectory data) Data relating to service performance and asset use may be considered as commercially sensitive data * Persistent identifiers are long-lasting unique references to digital objects (e.g. an identity number) (CERN, 2020)	Potentially If *ex-post* queries of provider data return data relating to uniquely and persistently identified vehicles or to individual trip characteristics, then such data can be considered to be personal data If aggregate data relating to provider-specific performance characteristics are collected by agencies, then this could comprise commercially sensitive data	No
Targeted user	Agencies focused on dynamic or real-time management of rights-of-way Agencies with technical capacity to run a more complex system and store sensitive data	Agencies focused on using historical data for planning or compliance Agencies with more limited technical capacity or a desire to minimise technical complexity	Agencies focused on dynamic or real-time management of rights-of-way Agencies with technical capacity to host and run this API

Source: Adapted from OMF (2021d).

Three additional APIs support the core APIs:

- Geography API looks at spatial coverage and boundaries that define where the rules set out in the Policy API apply or that trigger data logging and reporting in the Agency, Provider and Metrics APIs (e.g. a geo-fenced zone where vehicle speeds are reduced or a boundary whose crossing initiates logging of vehicle trip characteristics) (OMF, 2021d).
- Jurisdiction API is used to define and communicate hierarchical or overlapping administrative or operational areas associated with specific data access rights. This API also enables co-ordination among different agencies (OMF, 2021d).
- Metrics API establishes common methodologies for creating indicators from MDS data. This API defines common indicator semantics and ensures that indicators are consistently being

calculated. It thus provides the basis for consistent processing and the production of comparable outputs that can be incorporated into data-reporting mandates. Importantly, it can be used to process indicators from raw data on the mobility operator side, which reduces the risk of compromising personal or commercially sensitive data.

One of the most innovative and compelling functionalities of MDS is the formalisation of a legal and machine-readable bi-directional regulatory framework for mobility services. This helps both public authorities and service providers achieve their objectives for better regulation and more innovation. Nonetheless, there are strong concerns with the specific formulation of some MDS APIs. These relate especially to the detail and granularity of data collected (particularly the frequent reporting of individual vehicle-trip telemetry data) and associated risks for individual privacy and commercial sensitivity. In California, these concerns have led to legal challenges over the collection of these data by public authorities via MDS (Carey, 2021; Henry, 2021). While the principal challenge initiated by the American Civil Liberties Union was dismissed in February 2021, this case and another lawsuit initiated by Uber (Carey, 2021) underscore tensions related to the expansive collection of some forms of personal data. In particular, the current implementation of MDS, most notably the real-time management aspects of the Agency API, pose privacy risks that warrant enhanced personal data-protection efforts and data-handling protocols.

Implementation of the Agency API (or implementations of the Provider API that return historical data containing unique vehicle movements or individual trips) will result in personal data being delivered to public authorities. In the context where strong privacy protections exist, as in the case of the General Data Protection Regulation 2016 (GDPR) in Europe, rules define if and how that data should be collected, processed, handled, potentially shared and retained. In other contexts, without such explicit and robust rules, strong privacy-preserving policies should be enacted before personal data from MDS is collected, reported and processed. These tensions are indicative of the greater challenge to ensuring that privacy risks are not exacerbated by the design of regulatory frameworks for smart mobility systems.

The inherent risks to privacy that stem from the implementation of some parts of MDS are not insurmountable but do require some thought and specific actions on the part of public authorities to minimise these risks (Febvre, 2021; OMF, 2021e). OMF has issued a privacy guide to help authorities implement MDS, although it is largely focused on the US context (OMF, 2020). In December 2021, OMF also released detailed guidance on using MDS in the context of the GDPR (OMF, 2021e) (See Box 5). Further, the most recent release of MDS (1.2.0 – 4 November 2021) includes a beta feature that enables public authorities to specify which exact APIs, data endpoints and data fields are needed to carry out their mandates. The “digital expression of agency data requirements” feature is included as part of the MDS policy API and allows public authorities to specify in digital form which elements of MDS or GBFS they require mobility operators to use in their data-reporting streams. This enables public authorities to match data-reporting requirements to specific-use cases (e.g. inter-zonal flows versus parking enforcement) and helps minimise the potential for over-broad data reporting by allowing public authorities to only ask for the data they need (OMF, 2021f).

The challenges outlined here in the discussion around MDS also hold more generally for the public deployment and use of algorithmic governance frameworks (such as MDS), as well as for private operators of mobility services who are governed by those same frameworks (ITF, 2019a).

Box 5. Open Mobility Foundation guidance on using the Mobility Data Specification under the European Union's General Data Protection Regulation

Legal analysis undertaken on behalf of the Open Mobility Foundation (OMF) concludes that it is generally lawful to collect and process Mobility Data Specification (MDS) data, even in a non-aggregated form, including data which would be considered personal under the EU's General Data Protection Regulation 2016 (GDPR), provided that the requirements of the GDPR are respected. Such personal data may include: native vehicle IDs, vehicle location data, trip data, and any data associated with such vehicle IDs, location or trip data. MDS users may sometimes combine MDS data with other data, including directly identifying information. OMF cautions that such a combination may significantly increase privacy risks and should be subject to a rigorous impact assessment.

A key GDPR requirement is to have a legitimate purpose (i.e. at a minimum, a purpose that is not against the law) to collect personal data. Many use cases for MDS are legitimate, and some are even encouraged by EU law. Some examples of purposes for which MDS data may be used include:

- building useful statistics for urban planning and development
- data-driven policy making
- monitoring fleets of vehicles (either by mobility operators or by cities)
- enforcement of regulations imposed on mobility operators by cities.

The purpose will be key to determining how much MDS data should be collected, whether it should be in aggregated form or not, how long it can be retained, and with whom it can lawfully be shared.

While obtaining riders' consent is generally not required, the principle of transparency requires MDS users to inform riders of how data are collected or shared. There are limited exceptions under which informing riders may not be necessary, although OMF suggests that MDS users utilise these exceptions with caution and only after a proper impact assessment. As MDS data does not identify riders, MDS users will generally not be able or required to process data subjects' access requests.

The GDPR distinguishes between the different roles of data users: data controllers and data processors. Each qualification comes with its own obligations and liabilities, which must be fulfilled. In some cases, an organisation may be both a data controller and data processor or may hold these roles jointly with another organisation. Specific language may be required in contracts and other agreements to ensure all parties fulfil their requirements under the GDPR. In all cases, access to data should be limited to those persons and entities who need the data to fulfil legitimate use cases, and accordingly, the data should be protected by access controls and other security measures.

As complying with the GDPR often requires MDS users to make considered decisions based on a proper legal and technical assessment, they should work in conjunction with a Data Protection Officer.

OMF's *Using MDS under GDPR* guide (OMF, 2021e) provides a detailed review of the GDPR's applicability to MDS (as well as other components of the EU data-protection regulatory framework, including the Law Enforcement Directive) and further clarification as to how MDS users can ensure they are following legal requirements.

Source: Adapted from OMF (2021e).

City Data Standard for Mobility (CDS-M)

The kind of privacy concerns that have been raised with regard to certain components of the MDS toolkit has spurred on the development of an alternative new mobility data specification that integrates strong GDPR-compliance in Europe. The City Data Standard for Mobility (CDS-M) has been pioneered by Amsterdam and four other Dutch cities in the context of a national programme of Mobility as a Service (MaaS) pilots in the Netherlands (CDS-M, 2021). It is targeted at data exchanges between mobility service providers and public authorities.

CDS-M is designed to be both a standard and to codify agreed practices on data processing and storage. This standard seeks to deploy a similar set of functionalities as MDS but extending beyond micromobility and accounting for the European data-protection framework. In particular, the standard seeks to enable a similar, two-way communication stream of machine-readable information on services and operations between service providers and cities regarding rules, regulation and incident management. As part of their CDS-M development work, the consortium is developing a set of required data necessary for carrying out public authority actions and oversight, following the principles of purpose specification and data minimisation. CDS-M addresses three public authority requirements (CDS-M, 2021):

- planning – enabling cities to better manage public spaces for the adoption and use of multi-modal transport
- policy – enabling transport operators to have a clear understanding of policies that a city has for the use of their infrastructure
- enforcement – enabling cities to ensure a high level of service by transport operators within a city's boundaries through policy enforcement.

The CDS-M code architecture combines elements of MDS, other syntaxes (e.g. GBFS), bespoke code and API that ensures compliance with the GDPR and meets the needs of European cities. The working group piloting the design of CDS-M is composed of public authorities, MaaS stakeholders, transport service operators and experts. It liaises and co-operates with other standard-setting organisations, including CEN and the OMF. Use of the initial CDS-M specification is being piloted in Amsterdam, Utrecht and Eindhoven.

Mobility data-reporting principles and framework

Coherently framing public authority access to data in support of public service delivery and other societal outcomes is essential for enhancing trust by the public and market actors that such data is being put to good use. As noted earlier, this, in turn, helps create the right conditions to support data-reporting mandates where they are seen as beneficial and necessary. So what is the right way to frame mobility data-reporting policies?

Data-sharing frameworks and principles

A number of general data-governance frameworks and principles have been proposed, including those set out in various data-protection laws (described in Box 1). This section looks at some of the most recent of these, including the high-level framework set out by the OECD in its *Recommendations on Enhancing Access to and Sharing of Data (EADS)* (OECD, 2021c), the more mobility-specific frameworks set out by the World Business Council for Sustainable Development (WBCSD) in its report *Enabling data sharing: Emerging principles for transforming urban mobility* (WBCSD, 2020), the Sustainable Mobility for All (SuM4All) initiative in its report *Sustainable Mobility: Policy Making for Data Sharing* (SuM4All, 2021) and the New Urban Mobility alliance's (NUMO) *Privacy Principles for Mobility* (NUMO, 2021). This section then sets out recommendations for a mobility data-reporting framework that can help guide if, when and how public authorities should create obligatory or conditional data-reporting requirements.

OECD Recommendations on Enhancing Access to and Sharing of Data (EADS)

In October 2021, The OECD Council issued recommendations encompassing the first internationally agreed principles and policy guidance encouraging governments to develop coherent data-governance frameworks and policies (OECD, 2021c).

The recommendations build on prior work within the OECD, including but not limited to, the *OECD Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data* (1980, updated 2013 [OECD, 2013b]), the report *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-Use Across Societies* (OECD, 2019) and the *Good Practice Principles for Data Ethics in the Public Sector* (OECD, 2020). These documents all address the general question of how to maximise the benefits that flow from access and sharing of data while simultaneously protecting the fundamental rights of individuals and organisations (OECD, 2021b).

The more specific question of how to frame data-reporting requirements is woven into the broader set of OECD recommendations – for example, the framing text for the recommendations note that access to some forms of data improve public service delivery and may help identify emerging governmental or societal needs (OECD, 2021b).

The recommendations address three broad sets of measures that constitute an effective data-governance framework. These are measures that reinforce trust across the data ecosystem, stimulate investment in

data and incentivise data access and sharing, and measures that foster effective and responsible data access, sharing and use across society.

Reinforcing trust across the data ecosystem

Lack of trust amongst stakeholders is a key barrier to data sharing and can weaken specific mandates for data reporting from mobility operators and other stakeholders towards public authorities. In order to maximise trust in data-sharing practices and data-reporting mandates, OECD (2021c) recommends seeking inclusive representation of all relevant stakeholders in the data ecosystem and encouraging data-sharing arrangements that are competition-neutral.

Transparency with regards to data collection or reporting mandates and data-sharing arrangements can further improve confidence that data will contribute to beneficial outcomes. In particular, the recommendations stress that data-sharing and data-reporting initiatives should meet "applicable, recognised, and widely accepted technical, organisational, and legal standards and obligations, including codes of conduct, ethical principles and privacy and data protection regulation" (OECD, 2021c). These form the fundamental legal or policy basis for data-sharing and reporting initiatives. Also important is the agency and meaningful control individuals and other data ecosystem stakeholders have over data relating to them or that they have contributed.

A coherent whole-of-government approach to data sharing and reporting also helps improve trust in data collection. Foremost is the need to clearly and consistently state the objectives to which data sharing and reporting can contribute and the specific purpose for which data are being collected – especially when the reporting of personal or commercially sensitive data are compelled or made a condition of licensure or legal operation. More generally, the recommendations note that conditioned access to data is an especially effective data-sharing option when paired with data access control mechanisms (e.g. APIs) or technologies that preserve privacy by default.

From a broader perspective, public authorities should regularly review and adapt their data-governance frameworks and ensure that these are technology neutral (e.g. that they avoid "lock-in") and are able to account for changes in technology and practice (e.g. that they are "future-proof"). Delivering on a coherent whole-of-government approach requires strong leadership and co-ordination from the highest level (OECD, 2021b).

Finally, the recommendations on trust-building underscore the importance of balancing potential benefits from the sharing and reporting of data with necessary safeguards to preserve privacy rights and other interests, including those related to "national security, law enforcement, […] and intellectual property rights as well as ethical values and norms such as fairness, human dignity, autonomy, self-determination, and the protection against undue bias and discrimination between individuals or social groups" (OECD, 2021c). Authorities should ensure that all actors in the data ecosystem are informed of their data rights and responsibilities, as well as potential liabilities if these interests are violated. Robust accountability must be built into the data ecosystem.

Stimulating investment in data and incentivising data access and sharing

The second set of measures in the OECD recommendations refers to incentives that support and framework conditions that enable data sharing. These include measures to support the development of competitive markets for data where these do not erode personal privacy or damage competition. In some cases, where it is appropriate, some forms of self- or co-regulation may enable new societal and commercial value to be derived from data sharing, but these initiatives must be framed in law and accompanied with safeguards that guarantee privacy rights.

Creating supportive conditions for emerging data-enabled business models and applications can also create demand for shared data and further the need for a coherent data-governance framework. Finally, there is a need to support long-term investments in data sharing – this means fostering durable financing and revenue models where necessary and appropriate.

Fostering effective and responsible data access, sharing, and use across society

The OECD recommendations highlight that trust in data ecosystems must extend beyond regional or national borders, and thus efforts must be made to ensure that cross-border data access and sharing are not impeded where such sharing does not harm fundamental rights or other interests. Fostering effective data access and sharing also implies tracking and sharing meta-data, documentation, common and preferably open APIs and, where relevant, algorithms. From a capacity perspective, the OECD notes the importance of improving data-related skills and competencies, especially for public authority staff. Finally, the recommendations highlight the need to ensure wide and equitable access to "sustainable, open, scalable, safe, and secure foundational infrastructure" (OECD, 2021b).

The OECD recommendations serve to outline the broad basis for economy-wide data-sharing policies and practices – in essence, outlining what it is that stakeholders must think about when they think about data sharing. They serve to outline the fundamental elements of data-reporting mandates adopted by public authorities across all sectors and equally help guide how to structure sectoral data-reporting mandates – including in the transport sector.

World Business Council for Sustainable Development High-level Principles for Data Sharing

In January 2020, the World Business Council for Sustainable Development (WBCSD) identified five fundamental principles that should guide data-sharing initiatives (WBCSD, 2020). This section summarises the main elements of those principles:

- Data sharing should enable all stakeholders to create and capture value. This focus on value creation necessarily entails balancing the perceived benefits of retaining privileged access to data versus the real benefits generated by sharing that data. One of the more difficult arbitrations will be between the value of commercially held data and the societal benefits unlocked through its sharing and aggregation (WBCSD, 2020).
- Data sharing must be ethical, inclusive and unbiased. There are real risks inherent in sharing data, including the erosion of privacy rights or inequitable outcomes based on biased data collection and processing. Transparent and community-supported data-sharing models and protocols mitigate these risks.
- Data sharing should incorporate privacy by design. Privacy is a central concern with respect to the collection and sharing of personal data. Addressing privacy by design and ensuring high levels of privacy protection by default can improve trust in data-sharing and data-reporting initiatives.
- Data sharing should embrace cyber-security by design. A whole-of-system approach that embeds robust cybersecurity by both design and by default can also go a long way to improving trust in data sharing and reporting. Cybersecurity strategies should be designed to deliver outcomes even in the context of convergence, interoperability and interconnection of mobility systems (WBCSD, 2020).
- Data-sharing frameworks should be adaptive and iterative. The data ecosystem is evolving rapidly, and so too should the data-governance framework. This suggests regular re-assessment

of existing data-governance policies alongside flexibility in their application so as to avoid lock-in and ensure the ongoing relevance of data-sharing and reporting frameworks.

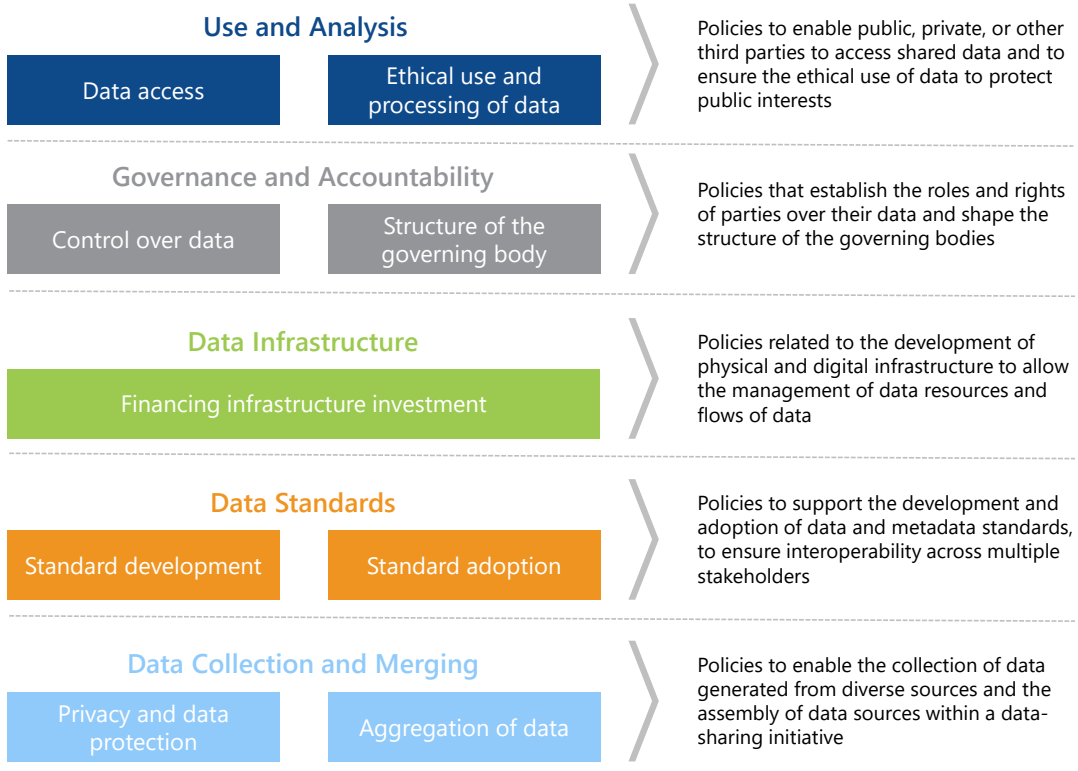
As with the data-sharing guidelines proposed by WBCSD, these five principles are also helpful in developing and framing public authority data-reporting initiatives.

Sustainable Mobility for All policy framework for mobility data sharing

In March 2021, the WBCSD, the International Road Federation and the Sustainable Mobility for All Initiative (SuM4All) released a report building on the guidance outlined in *Enabling data sharing: Emerging principles for transforming urban mobility* (WBCSD, 2020). The report *Sustainable Mobility: Policy Making for Data Sharing* (SuM4All, 2021) comprises part of SuM4All guidance on operationalising its Global Roadmap of Action towards Sustainable Mobility. It explores policies and measures that have been deployed in support of sustainable mobility with respect to five thematic areas: (i) data-sharing programmes and platforms; (ii) establishing data-protection regulations; (iii) requirements for service providers to report standardised data; (iv) developing data repositories and data collection guidelines; and (v) using data to support decision making.

SuM4All (2021) describes five interdependent and complementary layers of an effective data-sharing framework comprised of nine fundamental building blocks (see Figure 6). Each of the five layers describes an area for policy focus in the data-sharing ecosystem: (i) use and analysis; (ii) governance and accountability; (iii) data infrastructure; (iv) data standards; and (v) data collection and merging. These five layers are relevant not only for the general data-governance framework but are also important in helping structure and guide any public authority data-reporting initiatives.

Figure 6. Sustainable mobility for all data-sharing policy frameworks



Source: Adapted from SuM4All (2021).

In the context of the framework, SuM4All (2021) outlines six recommendations for policy authorities at municipal, regional and national levels. While not expressly identified, these recommendations seem equally suited for supra-national bodies with mandates to address data sharing (e.g. the European Commission). The six recommendations are:

- Adopt a collaborative approach for data sharing between diverse stakeholders. SuM4All highlights the need for a shared and stated vision for mobility and how data sharing (including data reporting) fits into and supports that vision. This pre-supposes that mechanisms for developing that vision are in place (such as the Sustainable Urban Mobility Framework in Europe). Sum4All underscores the need to align data sharing – and in particular, public sector data-reporting mandates – with that vision (purposive data reporting) and to ensure that only the minimum necessary data are collected in support of achieving identified tasks and with meaningful consent on the part of data subjects (data minimisation and consent).
- Commit to shared value across stakeholders to enable and accelerate benefits. Aligning the governance of data sharing with a common vision helps to channel and maximise overall public and commercial value creation. Public authorities should consider and account for the interests and capacities of stakeholders when creating a fair and competitive data-sharing ecosystem.
- Prioritise skill development and capacity building to increase competitiveness. Enabling a fair and competitive data-sharing framework that protects peoples' fundamental rights requires a set of skills (e.g. relating to data science, machine learning, artificial intelligence, etc.) that are generally not present, nor sought out, within public administrations – in transport and elsewhere. This puts public authorities at a disadvantage when it comes to assessing, guiding and regulating data-sharing initiatives.
- Seek harmonisation across jurisdictions while allowing for customisation based on the local context. A balance must be struck between tailoring data-sharing initiatives to the local context and creating data-sharing frameworks and policies that are scalable and replicable across jurisdictions and countries. Adopting common, high-level data-sharing architectures and mechanisms can deliver such balance and also help align approaches within governments themselves.
- Establish trust frameworks as a foundation for the implementation of multi-stakeholder data sharing. SuM4All underscores the importance of creating "trust frameworks – standardised legal and contractual agreements" (SuM4All, 2021) to ensure that data-sharing initiatives contribute to the shared vision outlined above. Part of that trust framework entails addressing and directly accounting for biases inherent to different data types and the results of data processing.
- Embrace iterative, incremental, and adaptive policy-making processes. Finally, SuM4All's recommendations underscore the need to adopt an iterative and continuous learning-based approach to trial different elements of the data-sharing ecosystem. This allows for the continual improvement in the effectiveness and skill with which the regulatory framework delivers on public policy objectives in a rapidly evolving data ecosystem.

New Urban Mobility Alliance's (NUMO) Privacy Principles for Mobility Data

In November 2021, a group of public and private actors convened by the New Urban Mobility Alliance (NUMO) released a set of seven framing principles to guide the collection, processing and use of mobility-related data (NUMO, 2021) (see Box 6). The principles are statements of values and priorities which provide a framework for public, private and non-profit organisations to collect and manage data in line

with the imperative to protect personal privacy. Their focus is on personal data in general and location data in particular. Importantly, they note that while not all mobility data present privacy risks, all data ecosystem stakeholders should, by default, consider them as personal data and manage them accordingly, unless they can be demonstrably shown not to pose a privacy risk to individuals (NUMO, 2021).

Box 6. New Urban Mobility Alliance Privacy Principles for Mobility Data

The New Urban Mobility Alliance (NUMO) Privacy Principles for Mobility Data are:

a set of values and priorities intended to guide the mobility ecosystem in the responsible use of data and the protection of individual privacy. Developed by a collaborative of cities, mobility service providers, technology companies, privacy advocates and academics, these Principles are meant to serve as a guiding "North Star" to assess technical and policy decisions that have implications for privacy when handling mobility data.

1. We will uphold the rights of individuals to privacy in their movements.

In practice:

a) Protect the privacy of users of shared mobility services, going above and beyond what is strictly required by law.

b) Be accountable for our privacy policies and practices, and encourage others in the mobility services industry to do the same.

c) Approach privacy of mobility data as an interdisciplinary effort, drawing on technical, operational, policy, economic and legal expertise.

d) Seek perspectives from marginalized communities and civil society.

e) Seek external input on the approach to privacy for mobility data. Treat all contributors with respect, and ensure their input is considered and handled fairly.

2. We will ensure community engagement and input, especially from those that have been historically marginalized, as we define our purposes, practices and policies related to mobility data.

In practice:

a) Build power with individuals and communities to influence decisions about the use of mobility data that they generate and about the ways in which their privacy is protected, and remain accountable to these individuals and communities.

b) Make engagement methods accessible to those without technical backgrounds or specialized knowledge.

c) Explore benefits, harms and risk mitigation strategies with the community.

d) Partner with other organizations, clients and vendors to understand and incorporate community needs and feedback, and embrace the shared responsibility of protecting privacy.

3. We will clearly and specifically define our purposes for working with mobility data.

In practice:

a) Clearly articulate and publicly document the specific purposes for which we collect, process, store and share mobility data.

b) Ensure purposes are lawful, relevant, narrowly tailored, specific, reasonable and fair.

c) Seek independent perspectives as we define purposes.

d) Be honest about purposes and the interests they serve.

e) Revisit and revise purposes and provide notice as an organization's goals and practices change and evolve.

4. We will communicate our purposes, practices and policies around mobility data to the people and communities we serve.

In practice:

a) Communicate in ways that are public, accessible, clear, specific and up-to-date.

b) Use best practices for communicating privacy policies such as a privacy statement clearly linked on the website, use of plain language, a layered and/or contextual approach, icons or physical cues.

c) Communicate with the goal of informing and educating, not simply to achieve compliance, and discuss both the benefits and risks of mobility data use.

d) Review communication with the intended audience in mind to ensure it is effective and useful.

5. We will collect and retain the minimum amount of mobility data that is necessary to fulfil our purposes.

In practice:

a) Consider whether purposes can reasonably be fulfilled without collecting mobility data, or by collecting less data.

b) Prior to collection, design policies and technology systems with the minimum amount of data necessary to achieve the purposes.

c) Choose the least granular data that is needed for the purposes, and use aggregate data rather than individual data where adequate for the purposes.

d) Discard, aggregate or obfuscate data which is no longer needed to fulfil the purposes or satisfy other legal requirements.

6. We will establish policies and practices that protect mobility data privacy.

In practice:

a) Consider the privacy risks against the benefits of the use of mobility data.

b) Establish policies that hold organizations, staff, contractors, vendors and other partners accountable to apply these Principles.

c) Keep anonymized data anonymous and prohibit re-identification.

d) Implement strong data security practices and procedures.

e) Establish transparency and response procedures for data breaches.

f) Regularly review policies to ensure that they are up-to-date and reflect any changes to purposes or practices.

g) Data held by public entities may be subject to disclosure under public records laws. Take steps to prevent any disclosure that could create privacy risk, especially as it pertains to individual trip records and geolocation data.

7. We will protect privacy when sharing mobility data.

In practice:

a) Establish clear policies and processes for sharing of mobility data, whether it be with internal teams, business partners, government or researchers.

b) Only share individual user or trip data when it is compatible with defined purposes or required by law.

c) Share the minimum amount of mobility data necessary for its intended purpose.

d) Establish data-sharing arrangements that preserve the protections afforded by these Principles even after mobility data is shared.

e) Limit any sharing of mobility data with law enforcement to instances where it is legally required or reasonably necessary.

Source: NUMO (2021).

The NUMO Privacy Principles are based on the following foundational values (NUMO, 2021):

- The collection and use of mobility data have the potential to benefit society, for example, by enabling shared mobility services and helping public agencies with city planning and management.
- Highly personal and sensitive information can be derived from mobility data.
- Individual trip records and geolocation data are sensitive and can be personally-identifying information.
- People should not have to choose between using essential mobility services and maintaining their privacy, especially those from marginalised communities and others who may have limited mobility options.
- Organisations that collect, process, retain, store, share or sell mobility data have a unique responsibility to act ethically and with accountability in their handling of mobility data.

Framework guidelines for mobility data reporting

Public authorities must legally and purposively collect data that enables them to carry out their mandates while imposing the least burden on data-reporting parties and posing the least risk to individuals' privacy and other desired outcomes. Doing so in a transparent, fair and effective way requires a coherent mobility data-reporting framework. This section recommends adopting a five-part data-reporting framework built on guidance provided by OECD (2013), EU (2016), European Commission (2018), WBCSD (2020), OECD (2021a), SuM4ALL (2021) and NUMO (2021), ICO (2021).

These guidelines target public authorities who require data reporting or make data reporting a condition of licensure or obtaining rights, including the operation of commercial services.

Establish and document the fundamental basis for data reporting

Public authorities should establish and document the fundamental basis for their data collection initiatives.

Addressing the following four questions will help accomplish this outcome.

i. Is there a legal framework or an explicit policy addressing the collection of personal or commercially sensitive data?

At the outset, public authority data-reporting mandates – especially those that are mandatory or are made conditional to licensure or for obtaining an operations permit – should be made on the basis of a clearly defined comprehensive legal framework. This framework exists in many countries and in some sub-national jurisdictions (e.g. throughout the European Union, California, etc. – see Box 1).

Where this legal framework is missing, or where it only partially addresses core issues such as the protection of fundamental human rights, including the right to privacy, open competition and other fundamental values, then the basis for collecting personal or commercially sensitive data should be explicitly stated in policy and with reference to applicable legal frameworks.

ii. Is there a coherent understanding across government, and within each agency, of all current data-reporting initiatives?

Public authorities may impose data-reporting requirements on a number of different stakeholders, for various purposes and across different government departments. Part of an effective data-reporting framework comprises inventorying all of these mandates and documenting them consistently and comparably. Carrying out data-protection impact assessments facilitates this task, as does having a single entity responsible for compiling, monitoring and managing data-reporting requests (see below).

iii. Is responsibility for data-reporting management clearly and meaningfully attributed to an individual or a team?

Clearly establishing roles and responsibilities for overseeing and managing data-reporting processes enables all the other elements of the data-reporting framework to function effectively. These roles entail two key functions – *data stewardship* and *data custodianship*.

Data stewards have oversight over all institutional data requirements, quality and fitness for the purpose of data assets. They are primarily concerned with data *content* and *context*. Data stewards are responsible for collecting, merging, logging meta-data, and addressing issues and problems with data. They ensure compliance with all applicable legal and policy obligations and ensure that data collected and processed can meet the defined purposes for which they were collected and achieve desired outcomes (Plotkin, 2021).

Data custodianship functions relate to the handling of data. Data custodians manage how data are stored, processed and transmitted internally and externally. They also are responsible for deploying physical and technical safeguards to protect the integrity, confidentiality, security and availability of data necessary for carrying out public authority mandates (CMU, 2021).

Both data stewardship and data custodianship roles should be clearly defined within public agencies that impose data-reporting requirements.

iv. *Is there an effective data quality control process in place?*

An effective data-reporting framework must ensure that data provided to public authorities is accurate and useable. This implies that authorities should carry out periodic data audits on reporting parties, especially when the latter provide aggregated or anonymised data. If third parties carry out data aggregation and anonymisation, then public authorities should similarly audit these intermediaries. This means that a formal audit process should be in place, outlining data access conditions, frequency and audit triggering mechanisms.

Purposive data collection

Public authority data-reporting mandates should be linked to explicit, identified and lawful purposes.

Addressing the following six questions will help ensure this outcome.

i. What is the purpose of collecting these data?

The specific purpose for which a public authority collects data – especially when this is compelled or required for licensure or service operation – should be clearly stated, documented and publicly available. The purpose should be linked both to high-level policy goals (e.g. improved safety, public space management, equitable access) and specific policy measures and tools that serve to achieve those high-level goals (e.g. speed control, parking management, assessing service coverage). A good example of this kind of mapping is NUMO's "Micromobility and Your City" digital tool (see Box 7).

ii. Is this purpose lawful?

All data reporting to public authorities should be for a lawful purpose. Explicitly referencing the legal basis for data collection will help communicate this and build trust.

iii. Are the data accurate and relevant?

The ability for data collected by public authorities to deliver on public policy outcomes is linked to its accuracy and representativeness. *Ex-ante* and *ex-post* assessments of data quality and potential data biases should be regularly conducted. If possible, poor-quality data should be improved, and inherent biases should be assessed and accounted for; otherwise, the data should be discarded. Publicly documenting data quality issues, biases and any remedial actions will improve confidence that data-reporting mandates are leading to their desired outcomes.

iv. Are the data necessary to carry out the stated purpose of their collection?

Data collected by public authorities should demonstrably be necessary and able to achieve the public policy objective for which they are collected. Describing and demonstrating how this will happen ensures that appropriate data are collected and over-broad data requests are avoided. The adequacy of data collection should regularly be reassessed given how well public policy objectives have been attained in the past and to account for new data sources and data-processing methods.

v. Do these data comprise personal or commercially sensitive data?

The collection of personal data or other sensitive data should trigger specific and robust collection, processing, retention, transmission and destruction protocols. In some cases, these are outlined by law. Where this is not the case, rules relating to collecting and using personal and other sensitive data should be set out in clear, meaningful and publicly communicated terms. Given the privacy risks inherent in many forms of data, especially location-based data, public authorities should err on the side of caution and treat any potentially personal data as personal data unless it can clearly be shown not to be the case.

When personal data are collected, the three rules of minimisation should apply: only the minimum amount of personal data necessary to achieve its stated purpose should be collected; these data should be transmitted to the minimum number of parties necessary to achieve its goals; and, the data should be retained for the minimum amount of time necessary to attain its stated objective.

vi. Is there an alternative to collecting personal or commercially sensitive data?

Alternatives to the mandatory or conditional reporting of personal data to public authorities should be evaluated before a final decision is made to require its collection. This assessment should be open and provide opportunities for stakeholders to identify and provide evidence on the ability and efficacy of non-personal data alternatives to achieve stated objectives.

Box 7. Setting purposive data collection: The New Urban Mobility Alliance's "Micromobility & Your City" tool

In August 2020, the New Urban Mobility Alliance (NUMO) launched a mobility data platform for public authorities to evaluate micromobility services against policy goals that foster safe, sustainable and equitable communities for all. The "Micromobility & Your City" platform allows public authorities to understand how micromobility policies contribute to broader social and sustainability objectives, rather than focusing only on how shared micromobility services comply with existing regulations. In particular, the platform allows authorities to match specific data-reporting elements based on common micromobility-oriented reporting syntaxes (e.g. Mobility Data Specifications [MDS] and General Bikeshare Feed Specification [GBFS]) with equity, sustainability and safety goals.

NUMO convened a coalition of over 50 experts from city and regional governments, research organisations, mobility service operators and data aggregation platforms to discuss and reach a consensus on how to use micromobility data to achieve public objectives. The discussions were complemented by a survey of 16 local regulations governing micromobility services and evaluations of seven pilot programmes in Canada, Mexico and the United States to gather the underlying data and cases that comprise the basis of Micromobility & Your City.

Source: Adapted from Numo (2020).

Transparent and relevant data processing

The results of data processing should be aligned with the purposes for which it was designed.

i. Were data subjects notified of public authority processing when their consent was obtained?

A cornerstone of most data-privacy frameworks is the meaningful consent given by data subjects for the collection and processing of data concerning them. Data subjects should also give their consent to the onward collection and processing of data by public authorities. In law enforcement actions, the rules

relating to consent are often different, but in all cases, the notion of consent should be clearly addressed and enacted in data-reporting frameworks. As much as legally possible, data subjects should be made aware of onward processing by public authorities in clear and easy to understand terms and express their consent to this via simple and easily actionable consent mechanisms.

ii. Is the outcome of processing suited for the purposes for which it was collected?

As with the data itself, an assessment should be made as to whether the outcomes of data processing are suited to satisfactorily achieve the purpose for which the processing was targeted – especially if this processing combines separate data streams.

iii. Are there biases in the outcome of the data processing?

In common with bias detection and management for data, an assessment should be made regarding potential biases that result from data processing – especially when multiple data sources are co-processed.

Limited data sharing

Public authorities' sharing of personal and sensitive data should be limited to only the extent and the parties necessary to achieve the purpose for its collection.

i. Are the conditions under which reported data would be shared with other agencies or third parties clearly stated?

Public authorities should clearly identify to whom reported data will be transmitted (if at all). These other parties may include other government departments or agencies but may also include third parties. In all cases, the transmission of data should be clearly and demonstrably linked to achieving the purpose for which it was collected. Onward transmission of personal data should be avoided by default and, if deemed necessary to achieve the stated purpose for its collection, should be limited to the minimal number of parties required. Consent for this transmission should be obtained at the time the data was originally collected from data subjects.

ii. Are sufficiently strong conditional access controls enacted for personal or commercially sensitive data?

Data custodians should enact strong and conditional access controls for personal and other sensitive data. These controls should minimise security risks and prevent unwanted access to data and its processing.

Appropriate data retention and destruction

Clear data retention, transformation and destruction policies build confidence that sensitive data will only be retained as long as strictly necessary.

i. Are data retained and stored in a secure manner in line with their sensitivity?

Data custodians should ensure that data are protected and secure throughout their lifetime. This is especially the case for personal and other sensitive data.

ii. Are retention periods specified in line with the purposes for which data are collected?

Data should only be retained for as long as necessary to fulfil the purpose of their collection. Data retention may be permanent for aggregated and anonymised data collected for planning purposes, as this provides valuable time-series data. However, retention periods for personal or other sensitive data should be strictly minimised to the time necessary to carry out the purpose for their collection.

iii. Are protocols for safe deprecation or irreversible de-identification of personal data specified?

If personal data are collected, they should be processed and transformed once the purpose for their collection is attained so that they no longer represent a privacy risk. Specific and documented protocols for irreversibly de-identifying personal data should be adopted, communicated and applied by public authorities.

iv. Are protocols for data destruction and meta-data archiving specified?

Once de-identified, original data should be irreversibly destroyed. Prior to this destruction, public authorities may wish to log and archive meta-data for onward use in planning and evaluation.

Specific actions to support data-reporting initiatives

The above framework outlines specific actions that public authorities should take to build trust in their data-reporting mandates – especially regarding the potential impact these would have on privacy rights and the protection of other unwanted outcomes (e.g. competition, security, etc.).

The framework has three immediate implications for how public authorities should organise their data-reporting processes. The first two are prerequisites to establish the *conditions* of a data-reporting framework, and the third relates to how public authorities should *organise* their data-reporting processes.

The first is that public authorities should clearly identify data stewardship and data custodianship roles and responsibilities. These roles may already exist in-house for institutional data-governance tasks. However, these roles and responsibilities should be explicitly extended and adapted to manage and oversee data-reporting mandates and the flow of information they generate. This may be a challenge for smaller or resource-constrained public authorities. Developing model guidance on these roles and responsibilities may help in these instances, as would guidance on managing contracts where some of these roles are outsourced.

The second is that an inventory, initiated and guided by a data steward, should be taken of all data-reporting initiatives – especially those that are mandatory, conditional to licensure or affect the ability to operate a service.

The third is that the entire data-reporting cycle should be documented and publicly available. Data reporting is part of a greater data-governance framework that is still developing across different levels of government. People should have visibility and access to those governance mechanisms so that they can meaningfully participate in their development and implementation (Waag, 2021). Accordingly, public authorities should develop a visible and auditable record of their data-reporting governance processes.

Data-protection impact assessments for data reporting

The EU's General Data Protection Regulation 2016 (GDPR) makes the provision for a data-protection impact assessment (DPIA) which enables organisations to record and demonstrate their compliance with European data-protection rules. DPIAs are a legal obligation under the GDPR when any of the following conditions are met (European Commission, 2021b):

- a systematic and extensive evaluation of the personal aspects of an individual, including profiling
- processing of sensitive data on a large scale
- systematic monitoring of public areas on a large scale.

However, the provisions of a DPIA are useful to consider when thinking more generally about how authorities can document and communicate about their data-reporting initiatives, especially for jurisdictions not covered by the GDPR. They can serve (indeed do serve, within the EU Member States) as a basis for framing potentially impactful data-reporting initiatives. More generally, DPIA-inspired approaches can serve as a way of structuring and enacting all data-reporting activities initiated by public authorities.

Public authorities should use the DPIA process to frame their data-reporting mandates and note how risks have been mitigated. Below are the main elements of a DPIA as outlined by (ICO, 2021).

What does a DPIA do?

A DPIA helps authorities identify and minimise the data-protection risks of an initiative. DPIAs are necessary under the GDPR for data collection and processing that is likely to be deemed high risk. Risk assessments should account for both the likelihood and severity of any impact on individuals – a high risk could result from a high probability of some harm or from a low probability of consequential harm. More generally, producing a DPIA is good practice for any data-reporting initiative involving personal or sensitive data. A DPIA (ICO, 2021):

- describes the nature, scope, context and purposes of the processing
- assesses necessity, proportionality and compliance measures
- identifies and assesses risks to individuals
- identifies any additional measures to mitigate those risks.

When should a DPIA be undertaken?

ICO (2021) notes that a DPIA should be conducted for any major project involving the use of personal data, as well as for any project which would:

- use systematic and extensive profiling or automated decision making to make significant decisions about people
- process special-category data or criminal-offence data on a large scale
- systematically monitor a publicly accessible place on a large scale
- use innovative technology
- use profiling, automated decision making or “special category” data to help make decisions on an individual’s access to a service, opportunity or benefit
- carry out profiling on a large scale
- process biometric or genetic data
- combine, compare or match data from multiple sources
- process personal data without providing a privacy notice directly to the individual (e.g. in the case of certain law enforcement actions)
- process personal data in a way that involves tracking individuals’ online or offline location or behaviour, in combination with any of the criteria in the GDPR
- process children’s personal data for profiling or automated decision making or for marketing purposes, or offer online services directly to them

- process personal data that could result in a risk of physical harm in the event of a security breach.

A DPIA should also be considered if a data-reporting initiative will contribute to:

- evaluation or scoring
- automated decision making with significant effects
- systematic monitoring
- processing of sensitive data or data of a highly personal nature
- processing on a large scale
- processing of data concerning vulnerable data subjects
- innovative technological or organisational solutions
- processing that involves preventing data subjects from exercising a right or using a service or contract.

How should a DPIA be undertaken, documented and communicated?

An effective DPIA helps document the way in which data collection is tied to lawful and publicly stated policy objectives. The DPIA should (ICO, 2021):

- describe the nature, scope, context and purposes of the data processing
- check that the processing is necessary for and proportionate to its stated purposes, and describe how compliance will be ensured with explicit data-protection principles
- undertake an objective assessment of the likelihood and severity of any risks to individuals' rights and interests
- identify measures that can be put in place to eliminate or reduce high risks
- document implementation of the measures identified
- consider how best to consult individuals (or their representatives) and other relevant stakeholders
- record decision making on the outcome of the DPIA, including any difference of opinion with internal or external stakeholders
- if third-party data processors are involved, they should document their processing activities and identify any associated risks.

How to assess the quality and efficacy of the DPIA?

Public authorities will need to ensure that the DPIA can deliver on its objectives and be as relevant and useful for framing data-reporting initiatives as possible. An effective DPIA will demonstrate that data-reporting initiatives are managed in line with public policy objectives and that data and data processing risks have been clearly identified and addressed. An effective DPIA will (ICO, 2021):

- explain why the DPIA was undertaken, detailing the types of intended processing that made it a requirement
- be structured clearly, systematically and logically

- written in plain language, with a non-specialist audience in mind, explaining any technical terms and acronyms used
- set out clearly the relationships between controllers, processors, data subjects and systems, using both text and data-flow diagrams where appropriate
- ensure that the specifics of any flows of personal data between people, systems, organisations and countries have been clearly explained and presented
- explicitly state the lawful basis for data collection and processing (and special category conditions if relevant)
- explain how relevant information rights of data subjects will be supported
- identify all relevant risks to individuals' rights and freedoms, including the likelihood and severity of such risks, and detail all relevant mitigations
- explain how any proposed mitigation reduces the identified risk in question
- provide evidence of the consideration of any less risky alternatives for achieving the purposes of the processing and why these were not chosen
- give details of stakeholder consultation (e.g. data subjects, representative bodies) and include summaries of the findings
- attach any relevant additional documents referenced in the DPIA, e.g. privacy notices, consent documents
- ensure the appropriate people sign off the DPIA
- agree and document a schedule for reviewing the DPIA regularly, or when the nature, scope, context or purposes of the processing are changed.

Establishing data protection by design and by default

Current practices in data governance, including in the governance of data sharing and reporting, are largely inadequate for delivering on desired policy outcomes – including the protection of the fundamental rights of individuals, welfare-improving competition, equity and sustainability. This is largely because current data-governance frameworks were developed outside the sphere of public governance and were not designed to deliver on these outcomes. This is true in transport as it is in many other sectors.

One of the most significant challenges going forward regarding the collective governance of data will be establishing a framework able to align personal, public and commercial value. Addressing this challenge will require deliberately (re)designing data-governance frameworks to deliver on these values. Such a framework should not necessarily be seen as a time-bound fixed objective (regulation is constantly evolving) but rather as conforming to a trajectory more aligned with the outcomes sought.

This section looks at how that trajectory may be shifted from its current direction to one that leads to the convergence of private, public and commercial value for data governance more broadly, and for data-sharing and data-reporting frameworks more specifically. It discusses the broader issue of the governance of digital space (or more precisely of hybrid digital-physical domains), the role of legal frameworks (with example initiatives from within the European Union), examples of fundamental building blocks designed to deliver public value, and how purposively designed digital infrastructure (e.g. the "public stack") could enable public-value data sharing and reporting by default.

Governance of digital space

Unlike public spaces, digital spaces are largely unregulated – or rather, they are only regulated partially, imperfectly and with no clear consensus on how and for whom they should be regulated. Such regulation is a fundamental challenge, not only for transport – where more and more activity is digitally mediated – but for all sectors of human activity. As in "real life", there is a need to set common, accepted and transparent rules for the digital domain that ensure that individuals' fundamental rights and interests are unimpeded, except where doing so would impede the fundamental rights and interests of others.

The balance between what is good for an individual and what is good for society is a constant and evolving challenge but one that leads to accepted "rules of the road" – quite literally in the case of transport. Such balances and rules are still very much in development for digital spaces, yet these spaces are increasingly at the centre of what happens on streets and in public spaces. Transport and public space governance must now address hybrid contexts, where physical spaces and activities are publicly regulated for one set of outcomes and yet where the digital domains that weave their way through everyday life are organised to deliver on a much more constrained set of private, often commercial, outcomes (van der Waal et al., 2020). This situation is detrimental to people individually and collectively. However, it is also detrimental to commercial interests in the longer term as a lack of trust in the entire data ecosystem is harmful to data-led innovation and the value it could generate if channelled appropriately.

The hybrid digital spaces emerging in transport are not designed to enable simple, straightforward, actionable and robust protection of personal and sensitive data by default. This challenge has motivated many data-governance principles, frameworks and measures described in the previous sections. However, these measures are more or less designed to adjust existing data practices without fundamentally changing the material, technical and foundational basis for the generation, collection and transmission of data. These approaches change how data could be *managed* – but do not change what data are or how frameworks could be *structured* to enable the kind of agency and control by data subjects that would enable data protection by design.

At the heart of this challenge are the affordances that data currently offer and how these may be ultimately restructured to deliver public value outcomes by default. There are many ways to understand "affordances" (Diver, 2018), but in the present case, "affordance" should be seen as "a relationship between the properties of an object and the capabilities of the agent that determine just how the object could possibly be used" (Norman, 1990). Affordances are not properties of things, but they are what those properties allow when those things are used. For example, the *physical properties* of a particular street may be that it is flat and narrow, with multiple entries and exits and alternating planters and parking spaces placed on either the left or right side. Among the affordances of that street is that a motorised vehicle may travel on that street (the street is drivable) but that it may not travel fast (the properties of the street preclude fast driving).

Digital spaces and objects also have affordances – things that they may or may not allow based on their properties. For example, one of the technical *properties* of mobile and web interfaces is that their code enables the installation and interaction with trackers that log the browsing, navigation or other behaviours of the users of those services – e.g. "cookies". In the European Union, following the adoption of the ePrivacy Directive (EU, 2002), digital or web interfaces provide EU users with the affordance to refuse or otherwise manage tracking cookies.

Nowhere in the current emerging data ecosystem for transport (or elsewhere) are users provided the affordance to control, assign or remove access and usage rights or otherwise manage data relating to themselves. For instance, none of the data syntaxes presented earlier provides data subjects with meaningful affordances relating to the control of their data. In some cases (but not universally), individuals are provided with the affordance to consent, or not, to the onward use of data concerning them, but this embodies a minimal level of agency. This shortcoming may be because there is a specific intent to prevent individuals or other actors from accessing these affordances: that there is an intent to create "disaffordances" for personal control of data. Negative affordances emerge when systems or objects are designed with functionalities specifically removed "or with the functionality deliberately hidden or obscured to reduce users' ability to use the (system or) product in certain ways, or a combination of the two" (Lockton, 2006). Perhaps more realistically, these affordances are simply not present because there are no legal or other incentives to ensure that the design of the data ecosystem inherently builds them in. One of the principal challenges in governing digital spaces is creating a legal and regulatory framework that incentivises or requires data systems to allow affordances to protect personal or sensitive data (Hildebrandt, 2015).

Updating legal frameworks to establish coherent public value for data governance

Adapted and new governance frameworks will need to emerge in order to govern digital spaces effectively. These, in turn, should enable public-value data sharing and reporting by *default* rather than as an

afterthought. Efforts to govern rapidly evolving data ecosystems with decades-old regulatory frameworks designed for analogue services are bound to result in suboptimal, inefficient and possibly privacy-damaging outcomes. In particular, data-sharing and data-reporting frameworks built on outdated regulations inherit deficiencies stemming from the poor fit between those regulatory frameworks and the needs of digital regulation.

The European Union has undertaken the most extensive and ambitious effort to update regulation for a data-rich society. It has outlined a vision for the digital transformation of European society and markets, outlined in the flagship plan *Shaping Europe's Digital Future* (European Commission, 2020b). This strategy articulates three framing principles to guide the digital transformation of Europe from 2020 to 2025:

> Technology that works for people: Development, deployment and uptake of technology that makes a real difference to people's daily lives. A strong and competitive economy that masters and shapes technology in a way that respects European values.
>
> A fair and competitive economy: A frictionless single market, where companies of all sizes and in any sector can compete on equal terms, and can develop, market and use digital technologies, products and services at a scale that boosts their productivity and global competitiveness, and consumers can be confident that their rights are respected.
>
> An open, democratic and sustainable society: A trustworthy environment in which citizens are empowered in how they act and interact, and of the data they provide both online and offline. A European way to digital transformation which enhances our democratic values, respects our fundamental rights, and contributes to a sustainable, climate-neutral and resource-efficient economy. (European Commission, 2020b)

The EU Communication *A European strategy for data* outlines how the various data-related facets of that digital transformation will be handled. This communication expressly notes the need to address the "use of privately-held data by government authorities (business-to-government – B2G) data sharing (e.g. data reporting)" (European Commission, 2020c). In line with the visions outlined in both documents, the European Union continues to establish new data-governance frameworks with successive iterations of the ePrivacy Directive (and its forthcoming latest iteration – the ePrivacy Regulation), the General Data Protection Regulation 2016 (GDPR) and now the soon to be enacted Data Governance Act (EU, 2002, 2016, 2020; Kayali and Manancourt, 2021).

The most recent of these initiatives – the proposed Data Governance Act (DGA) (EU, 2020) – seeks to establish a common legal basis for the sharing (and reporting) of data while respecting personal privacy and proprietary data rights. For public authorities, the DGA calls explicitly for new data-sharing pathways for data held by authorities but which is subject to the rights of others as defined in the European Union – e.g. personal data or proprietary data. In transport, authorities often collect these data via data-reporting mandates or as a condition for licensure or to operate a service. The DGA seeks to enable new opportunities to share these data in a way that creates new value while upholding privacy rights and fair competition. Doing so will require authorities to be technically able to provide privacy and confidentiality protection – or otherwise be able to avail themselves of the services of data intermediaries that are so equipped.

Beyond addressing the ability for public authorities to share personal or sensitive data in safe ways, the DGA comprises three other principal components: the first relates to establishing a legal framework for ensuring strict neutrality for data intermediaries (data intermediaries should not use the data they exchange for any other purpose); the second relates to the governance of data exchanges and data sharing

undertaken for altruistic (non-commercial) purposes; and the third relates to the establishment of a European Data Innovation Board which will help ensure data interoperability and oversight (EU, 2020).

The formalisation and specification of the data intermediary role contained within the DGA are aligned with the concept of a "data trust" – for example, "legal mechanisms to promote multi-party data sharing with some form [of] independent stewardship of data" (Rauer and Cameron, 2020).

Another EU initiative – a proposed "Data Act" – has passed the consultation phase and is currently in preparation (European Commission, 2021c). Whereas the DGA establishes the basis and framework for the governance of data sharing and data access, the Data Act seeks to establish specific rights, responsibilities and roles within that framework. In particular, it may seek to establish stronger individual rights and control mechanisms for individuals' data – for example, by operationalising the ambition set out in *A European Strategy for Data* that "individuals should be further supported in enforcing their rights with regard to the use of the data they generate. They can be empowered to be in control of their data through tools and means to decide at a granular level about what is done with their data ("personal data spaces") and establish a framework for this" (European Commission, 2020c).

These two elements – data trusts and personal data spaces – are examples of new and adapted building blocks that could help support new data governance, sharing and reporting frameworks. These and other building blocks could enable the systematic emergence of privacy-protective affordances within the data ecosystem. This would enable data protection for data sharing and reporting by design, not by retro-fit.

Building blocks of public value data governance

Ensuring that the data ecosystem preserves privacy and protects commercial interests by default requires new mechanisms and building blocks. These will build on new concepts – such as data trusts and personal data spaces – but will require tangible and actionable tools. These are only just emerging – for example, the Minimum Interoperability Mechanisms (MIMs) and MyData described below – but they are indicative of the new mechanisms that may underpin public value data governance in the future.

Minimum Interoperability Mechanisms (MIMs)

One of the difficulties in establishing a coherent and effective data-governance framework – including for data reporting – is the multiplicity of data schemas and syntaxes. As noted earlier, these are often created for bespoke applications and are rarely designed for strong data protection. This hinders the capabilities for actors in the data ecosystem – including commercial operators, public authorities and individuals – to share, use or re-use data effectively. Enhancing data interoperability is a key challenge and mechanisms for enhanced data interoperability could enable or improve more consistent and trustworthy data-protection functionalities. The MIMs proposed by the Open and Agile Smart Cities network seeks to address this challenge (OASC, 2021).

MIMs "are the minimal but sufficient capabilities needed to achieve interoperability of data, systems, and services between buyers, suppliers and regulators across governance levels around the world" (LI.EU, 2021). The most current version (v 4.0 – also referred to MIMs Plus) sets out a comprehensive vision for an open, technology-agnostic and vendor-neutral data schema which, if implemented, would enable all stakeholders in the data ecosystem to achieve interoperability building on existing and bespoke data syntaxes and standards already in place.

MIMs Plus are organised into three functional layers, each encompassing a set of technical mechanisms which enable meaningful interoperability (LI.EU, 2021):

- Interaction Layer: including "knowledge and context information exchange, rules of access and use for data and services, and management of location data".
- Integrity Layer: encompassing "protection of rights (personal data, privacy, dignity, equality…); transparency in automated decision making (societal governance of all technology use and deployment), and security (systems and society)".
- Impact Layer: which is "driven by societal objectives with measurable outcomes towards those objectives, taking into account existing indicators, analytics, and resource management frameworks".

The MIMs Plus schema seeks to ensure that broader societal goals regarding data collection and use are built into the technical systems that collect and process data. MIMs Plus compliance, for example, would ensure that data-reporting processes are privacy-preserving by default. The schema also seeks to provide a flexible and scalable framework that enables data-governance frameworks to account for new forms of data and processing. Although recent in their conception and uptake, they are already referenced in various initiatives and pilot projects – as in the case of the SynchroniCity pilots, which sought to validate the MIMs in various city contexts (see Box 8).

Box 8. SynchroniCity: Implementing Minimum Interoperability Mechanisms

SynchroniCity – a three-year EU 2020 research and innovation project (2017-19) funded primarily by the European Commission as well as by Korea and Switzerland – defined and validated a framework for a new digital single market, where local authorities and technology providers of all sizes can easily exchange data and digital goods and services in a fair data economy. The framework was piloted at scale in multiple cities around the world, where a variety of Internet of Things (IoT) services were deployed – demonstrating that a multi-vendor ecosystem is achievable.

Rather than creating new technology, SynchroniCity employed the Minimal Interoperability Mechanisms (MIMs), developed and supported by the Open and Agile Smart Cities (OASC) community. These outline the minimum technical requirements needed for technology providers to interface their IoT solutions with local authorities' digital systems. The interoperability of these mechanisms enables impactful IoT solutions to be easily deployed and replicated in any local authority experiencing similar challenges.

Source: Adapted from SynchroniCity (2020).

The MIMs Plus is built around a core set of architectural design principles, including the following (LI.EU, 2021):

- adopting a layered and capability-based common architectural model which can be used across different cities and domains
- building on open international standards (where these are available)
- working with existing technical solutions and focusing on interoperable interfaces between these
- enabling modular, flexible and scalable solutions to allow for bespoke adoption by all kinds of cities and in different contexts
- ensuring privacy and security by design

- adopting global, standard-based open application programming interfaces (APIs) to enable broad interoperability
- ensuring data harmonisation and global standards based on semantic interoperability through the adoption of common, linked data models.

There are currently ten mechanisms identified in the MIMS Plus schema (as of June 2021) – five are already implemented and five are currently work items to be implemented. Organised by MIMs Plus layer, these are:

Overall architectural capabilities

Shared data models (MIM 2 – implemented): sets guidelines for representation and catalogues various data models used across cities and sectors. It outlines harmonised formats and semantics to be used when utilising and publishing data across stakeholders. Clearly defining and mapping data models enables improved interoperability.

Interaction layer

Context information management (MIM 1 – implemented): harmonised context representations (e.g. data about data) enable diverse systems and stakeholders to seek and find data useful for an application. This improves access, use, sharing and management of data.

Ecosystem transactions management (MIM 3 – implemented): enables the standardised exposure of data and datasets in a privacy- and security-maximising manner in order to enable the sharing of data for commercial applications (and, indirectly, for data-reporting purposes). This mechanism could be facilitated by the uptake of standardised and open application programming interface (API) suites (e.g. the Business API Ecosystem [FIWARE-TMForum, 2021]).

Geospatial information management (MIM 7 – work item): this mechanism indicates how to share spatial and spatio-temporal data in such a way so as to minimise privacy risks and maximise public and commercial value. This mechanism could be facilitated by open API-based building blocks that address spatial data referencing and sharing (e.g. the Open Geospatial Consortium's SensorThings API [OGC, 2021]).

Integrity Layer

Personal data management (MIM 4 – implemented): this mechanism provides individuals with clear, easy to use and meaningful ways to control which of their data or attributes they want to share and specify the terms and conditions under which service or solution providers use that data. It seeks to create full agency for individuals with respect to the control and sharing of their data. This mechanism references the MyData model (discussed below).

Fair artificial intelligence (MIM 5 – implemented): sets out six minimal requirements for the suppliers of algorithmic systems to ensure that they are transparent, trustworthy and fair. These cover procedural transparency, technical transparency, technical explainability, fairness, context-setting and accountability.

Security management (MIM 6 – work item): will establish interoperability in security risk assessment exercises and provide a common basis for establishing counter-measures.

Impact layer

Ecosystem indicator management (MIM 8 – work item): will provide consistent and comparable benchmarking frameworks across various cities, sectors and stakeholders.

Data analytics management (MIM 9 – work item): complex data models (e.g. agent-based simulations) are increasingly used to address complex problems. This mechanism will enhance their interoperability.

Resource impact management (MIM 10 – work item): this mechanism will develop interoperability for assessments of scarcity and resource requirements in relation to nature, people and investments.

MIMS focuses on interoperability, referencing and building on existing systems. They provide a framework to adapt existing data practices to the kinds of goals and outcomes that citizens and public authorities require for data ecosystems. They are a practical means of ensuring that the trajectory of current data practices can be shifted to one more in line with public value outcomes. More to the point, the uptake ensures that data sharing and data reporting contribute to these outcomes even without completely re-configuring the architecture of the data ecosystem.

MyData

MyData is an open standards platform that provides people with meaningful access and control over information pertaining to and produced by them. It provides practical ways to operationalise the concepts of individual agency for data sharing and data portability, allowing people to set the terms and conditions for access to and use of their data. It is the reference standard for MIMS Plus No. 4: *Personal Data Management.* Established in 2014 with the Finnish Ministry of Transport and Communications, MyData has since evolved into a global standard supported by multiple national and regional hubs and managed through a not for profit organisation (Poikola et al., 2020).

The MyData framework is organised around three guiding principles (1001 Lakes Oy et al., 2021):

- from formal to actionable rights
- from data protection to data empowerment
- from closed to open ecosystems.

In order to operationalise these principles, MyData outlines specific roles and responsibilities relating to four data ecosystem actors: individuals to whom the personal data pertains; data sources who may want to collect and share individuals' data; data-using services who may wish to receive and use individuals' data; and (data-sharing) operators who act as agents for individuals in the data-sharing framework and host and implement individuals' permissions (see Figure 7).

At the heart of the MyData framework are *natural persons* who manage and control the use of data relating to them. These data could be transactional data (things they bought), service-related data (when they accessed different mobility services), spatial data (where and when they travelled), or any other form of personal data. In the MyData framework, natural persons set the conditions and terms of access to and use of their data.

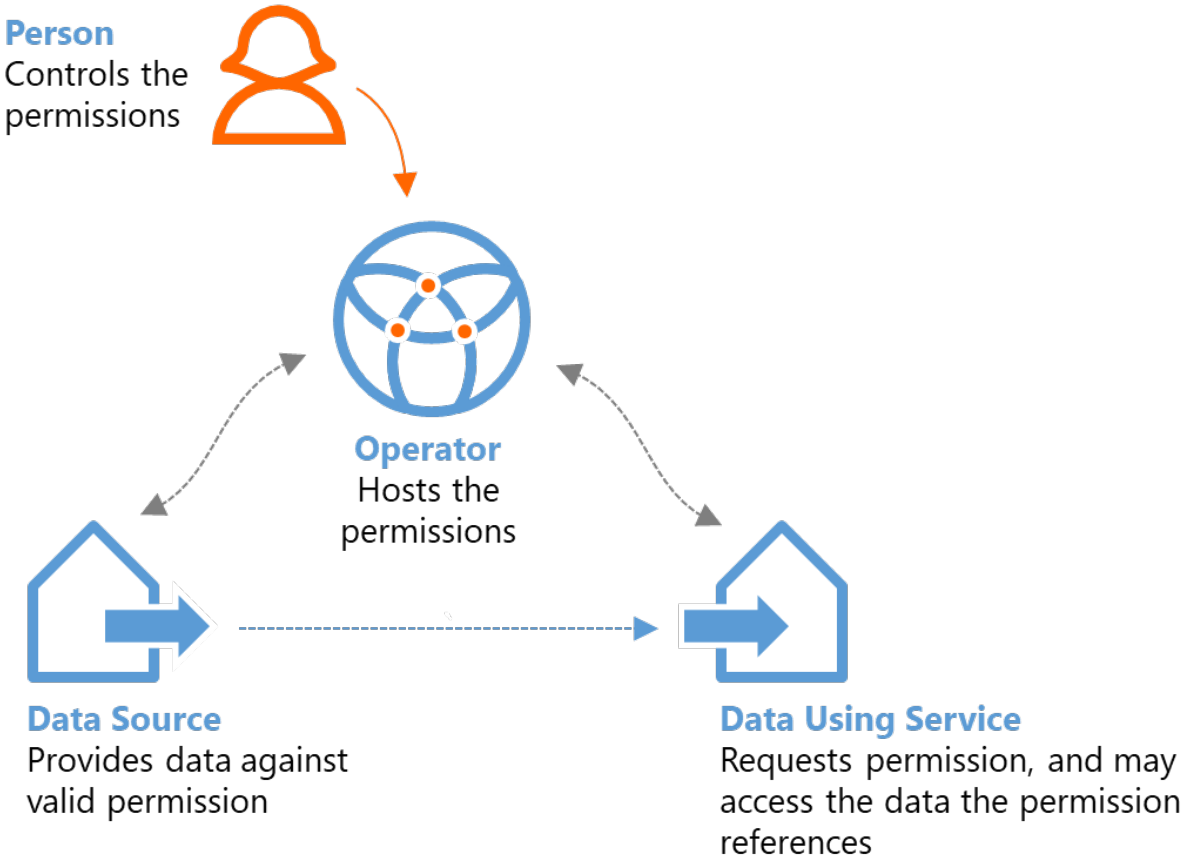
Personal data, however, is not collected by individuals themselves but rather by entities that sense, track or otherwise collect data about individuals for commercial or public purposes. In the MyData framework, these are "data sources" or "data providers". Data sources must receive permission from individuals to share their personal data.

There are also entities that wish to have access to or use personal data collected and held by data sources – these are "data-using services" or "data consumers" in the MyData framework. Data-using services must seek permission to access personal data held by data sources.

The *data operator* is the entity that manages permissions to share or access personal data on behalf of natural persons. Individuals set the terms and conditions of access and use of their data and provide these to data operators who, in turn, mediate data requests and validate only those that conform to the

permissions set by individuals. The data operator(s) provides infrastructure and tools for people to manage and control their data in a meaningful way.

Figure 7. Actors and roles in MyData-based personal data management



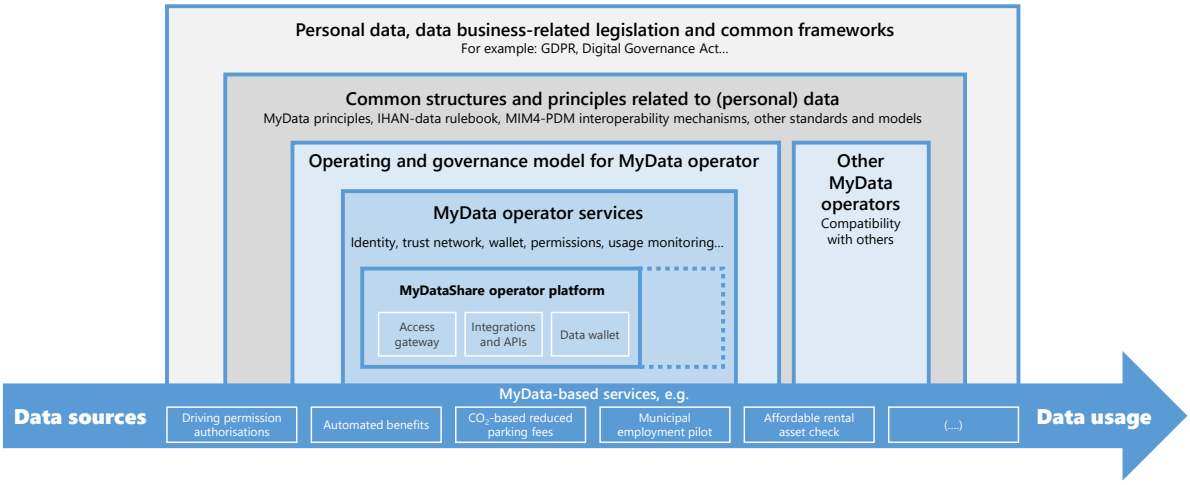
Source: Adapted from 1001 Lakes Oy et al. (2021).

The MyData model can serve as the basis for all data-sharing and reporting interactions involving personal data. It incentivises actors, including public authorities, to only seek to use personal data for those cases where it is most useful and where the costs of compliance with the MyData framework generate clear commercial or public value outcomes. It also provides the basis for a coherent data ecosystem with respect to the material protection of fundamental values, value creation and transparency.

Figure 8 illustrates how the MyData framework fits into a broader data-governance framework for a city. At the highest level, an over-arching regulatory framework sets out common principles and rules relating to the use of data and of personal data in particular. As described earlier, this is the kind of regulatory architecture that the European Union is putting in place. Below this level are the common data structures, schemas and syntaxes that enable the operationalisation of the top-level regulatory objectives. This is where common or interoperable standards and mechanisms come into play – such as those being developed in the context of the MIMs framework and, to a lesser and more constrained degree, by various data syntaxes such as the Network and Timetable Exchange (NeTex), Mobility Data Specification (MDS), General Bikeshare Feed Specification (GBFS) and City Data Standard for Mobility (CDS-M). Next, there is the level at which individuals interact with designated MyData operators to specify which data access and use permissions should be granted between data sources and data-using services. Based on these

permissions, individuals gain access to services (and only those services and outcomes) for which they are willing to share personal data. This is a quite different system than what exists today and one that, by its very design, provides individuals with the affordance to control when, how and by whom their personal data are used.

Figure 8. MyData in a coherent and privacy-preserving data-governance framework for cities



Source: Adapted from 1001 Lakes Oy et al. (2021).

Designing data infrastructure for public value: The public stack

Digital and data-governance outcomes are directly linked to data architectures and infrastructures – how the latter are conceived and designed can enable or block various governance affordances. Aligning digital architectures and infrastructures to deliver public-value outcomes by default will require rethinking and reframing the "stacks" at the heart of the digital ecosystems. This will, in turn, ensure that data-sharing and data-reporting initiatives contribute to the outcomes people want for the use of their data (van der Waal et al., 2020).

A "stack" is comprised of the different digital and physical layers which, each with their own complexity and function, create a service or activate an outcome. van der Waal et al. (2020) specify that "the stack of any given technological object or service is the entire range of components that make that object or service what it is".

Think of the stack that enables on-demand ridesourcing. The technological layer is embodied in a handheld mobile device that enables the user to access the service via an interface and allows the service to locate the user via the location sensing chip. The communication module allows the handheld device to be connected to a remote server, itself comprised of multiple interlocked systems. The vehicle is driven by the driver (and the mobile device they use to know where and whom to pick up). There are technical modules that handle payment processing and security functions. Together, all of these work to activate access rights for users and make the ridesourcing service function.

Throughout the stack, data (including personal data) are created, logged, transmitted and processed in order to facilitate achieving the stack's primary function. It is these data that public authorities sometimes wish to access in order to better carry out their governance mandates. However, this stack is not created to support public policy outcomes – these and other such stacks' "constituent parts have been developed

by mostly private companies, and the stack is thus tailored to ensuring profit for those companies" (van der Waal et al., 2020). That companies derive profit from systems they deploy is normal. However, when the entire governance of digital public spaces – and increasingly significant parts of real spaces – are governed for the benefit of commercial actors at the expense of all others, then fundamental rights, including those of agency and privacy, are threatened. In fact, the entire discussion around setting up effective data-governance frameworks is linked to this tension between how technical stacks instantiate action and outcomes and for whom and by whom these stacks are built.

Not all stacks are "private" stacks – some stacks are initiated by or organised to deliver state functions – such as state digital identity services (e.g. digital identification systems like those deployed in Estonia) or online tax systems. Some state stacks integrate with private stacks in order to carry out functions entrusted to the state, but that carry fundamental risks, which is why they are often limited in real public spaces – e.g. surveillance of citizens. Risks to fundamental rights and other public values emerge when state functions are carried out via stacks that were not designed for these purposes and over which public authorities exert little meaningful control. This is one of the risks incurred with over-broad reporting mandates, which include personal data.

For these reasons, it has been proposed that digital and data architectures should be built from the ground up to ensure that public values are incorporated from the outset and guide the functions and services that these stacks support. Such "public stacks" would enable public value creation from the outset by incorporating many of the elements described in this section – an overarching legal and regulatory framework, supportive data schemas and syntaxes, and specific building blocks and modules – to deliver an open democratic and sustainable society (van der Waal et al., 2020).

The public stack is comprised of three principal layers (see Figure 9). At the base is the *foundation layer*. This layer underpins all the other layers and helps to define the entire architecture of the stack. It clarifies starting points and basic assumptions, establishes and states fundamental rights and values, outlines governance and oversight mechanisms and accounts for socio-economic considerations where these are relevant to the delivery of fundamental rights and values.

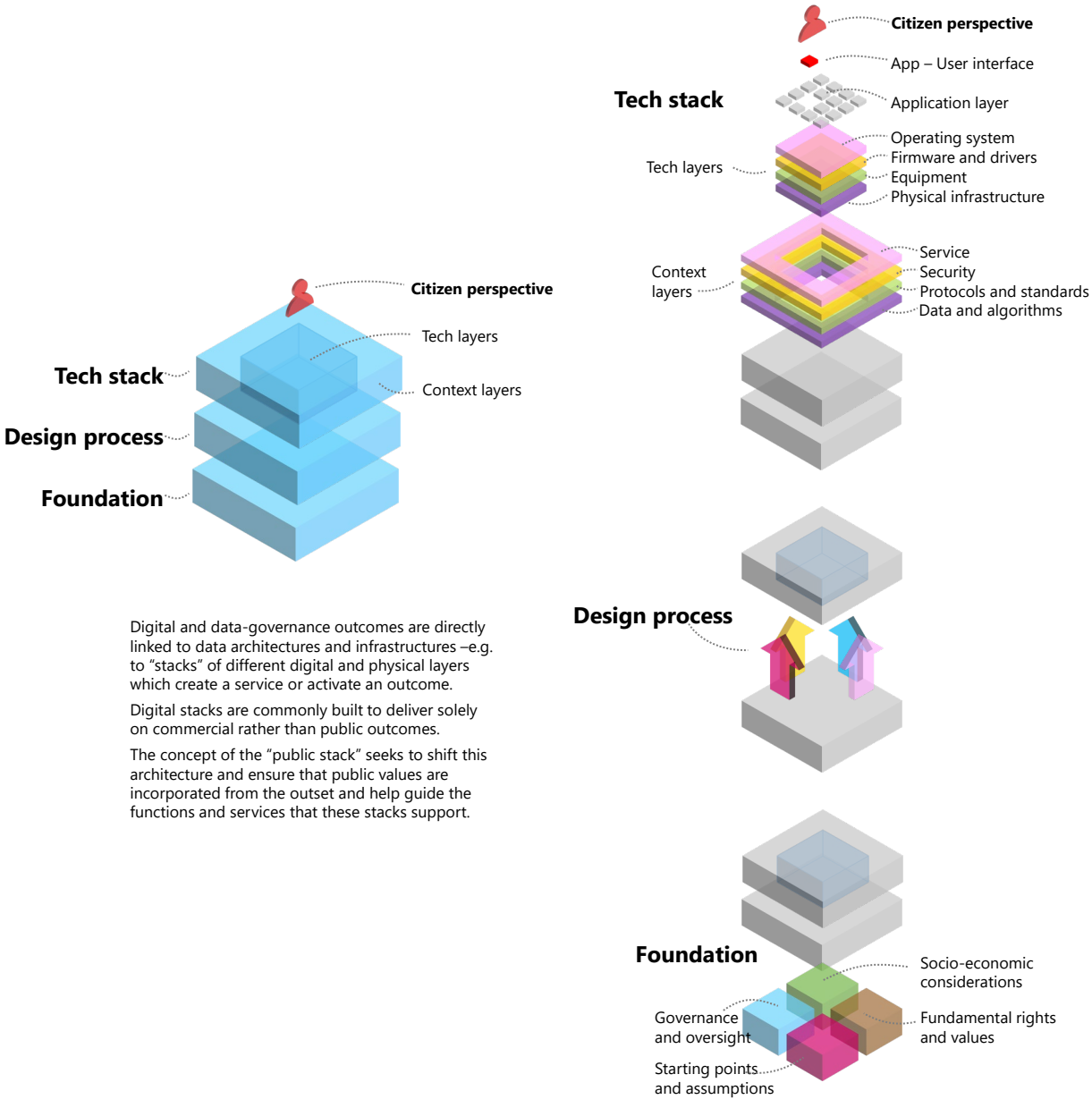
The second layer is comprised of the *design processes* that contribute to the technology stack. Systems, syntaxes, protocols, devices and the ways in which they are integrated are each the result of a design process, and the way in which they all join together is also the result of a design process. These design processes are often closed off and optimised to very specific outcomes. Ensuring that they are all aligned to a common foundational base will ensure their orientation towards desired public value outcomes, but there are other ways to support that result. From a public authority perspective, investment in public research and public participation help ensure that the design and public funding of parts of the technology stack account for a broad and representative range of voices.

The *technology stack* is a combination of two sub-layers: the tech layer, which is comprised of material infrastructure (global positioning satellites, servers, fibre optic cables, etc.), firmware and operating systems, and the user interface layer or application layer, which may be a browser, an app or another interface with human users. Serving as the "glue" that brings together the components of the tech layer is the context layer. This layer is comprised of protocols, syntaxes and standards that enable communication and data exchange. Data and algorithms manifest themselves at each layer in the stack and, together, create the services for which the stack is designed. As discussed earlier, these context layers are often hidden but may have consequential impacts on data privacy and security. Throughout the tech stack is a need to ensure security against potential misuse.

At the top of the public stack concept are natural persons endowed with fundamental rights and whose privacy and agency must be preserved. Building the public stack on foundations that recognise and seek

to guarantee these rights, using design processes that transparently incorporate these foundational principles, and by ensuring that technology and context layers work to deliver these outcomes as they create commercial value, enables the entire stack to function for people. The public stack is then to digital spaces what the street and other public spaces represent in most of the real world – democratic interaction spaces that guarantee public value.

Figure 9. The public stack: Embedding public values into data and technology architecture and infrastructure



Source: Adapted from van der Waal et al. (2020) and Waag (2021).

References

1001 Lakes Oy et al. (2021), *Personal Data Management Minimum Interoperability Mechanism (MIM) 4 Introduction and Specifications*, www.vastuugroup.fi/hubfs/MyData/2021-05-31%20MIM4%20-%20Personal%20Data%20Management%20Specification%20-%20Final%20draft.pdf.

Antrim, A. and S.J. Barbeau (2013), "The many uses of GTFS data: Opening the door to transit and multimodal applications", ITS America's 23rd Annual Meeting and Exposition, pp. 1-24, https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.391.5421&rep=rep1&type=pdf.

APTA (2013), "APTA Surveys Transit Agencies on Providing Information and Real-Time Arrivals to Customers", www.apta.com/wp-content/uploads/Resources/resources/reportsandpublications/Documents/APTA-Real-Time-Data-Survey.pdf.

Atmaca, U.I. et al. (2021), "A privacy-preserving route planning scheme for the Internet of Vehicles", *Ad Hoc Networks*, Vol. 123, p. 102680, https://doi.org/10.1016/j.adhoc.2021.102680.

Barbeau, S. J. (2018), Quality Control – Lessons Learned From the Deployment and Evaluation of GTFS-Realtime Feeds.

Borg, M. et al. (2018), "Digitalization of Swedish Government Agencies: A perspective through the lens of a software development census", Proceedings of the 40th International Conference on Software Engineering, pp. 37-46, https://doi.org/10.1145/3183428.3183434.

Bourée, K. et al. (2019), *INSPIRE-MMTIS Overlap in standards related to the Delegated Regulation (EU) 2017/1926: Final report with recommendations to Member States and to the EC*, https://doi.org/10.2760/404745.

Carey, C. (2021), "US court dismisses lawsuit against LA's mobility data sharing requirement", *Cities Today,* https://cities-today.com/us-court-dismisses-lawsuit-against-las-mobility-data-sharing-requirement/ (accessed 12 January 2022).

Catalá, M., Downing, S. and Hayward, D. (2011), "Expanding the Google Transit Feed Specification to Support Operations and Planning", *Florida Department of Transportation Research*, https://rosap.ntl.bts.gov/view/dot/39780 (accessed 14 February 2022).

Cavoukian, A. (2011), *Privacy by Design – The 7 Foundational Principles: Implementation and Mapping of Fair Information Practices*, Ontario, CA, https://iapp.org/media/pdf/resource_center/pbd_implement_7found_principles.pdf.

CDS-M (2021), "City Data Standard – Mobility", Github, https://github.com/CDSM-WG/CDS-M (accessed 27 February 2021).

CEN (2021a), *Transmodel at a glance*, www.transmodel-cen.eu/transmodel-at-a-glance/ (accessed 21 January 2022).

CEN (2021b), *Transmodel Tutorial- Management Information and Statistics EN12896 – 8*, www.transmodel-cen.eu/overview/ (accessed 15 Febraury 2022).

CEN (2020), *NeTEx extension for New Modes – Scope and overview*, http://netex-cen.eu/wp-content/uploads/2021/03/NeTEx-extension-for-New-Modes-Detailed-Scope-v04.pdf.

CEN (2018), *OpRa – Overview*, www.opra-cen.eu/overview/ (accessed 12 January 2022).

CERN (2020), *What are persistent identifiers?,* https://scientific-info.cern/submit-and-publish/persistent-identifiers/what-are-pids (accessed 2 November 2021).

Chatzigiannakis, I., A. Vitaletti and A. Pyrgelis (2016), "A privacy-preserving smart parking system using an IoT elliptic curve based security platform", *Computer Communications*, Vol. 89-90, pp. 165-177, https://doi.org/10.1016/j.comcom.2016.03.014.

CMU (2021), "Roles and responsibilities: Data Custodian", *Computing Services*, Carnegie Mellon University, www.cmu.edu/iso/governance/roles/data-custodian.html (accessed 21 January 2022).

Cunha, M., R. Mendes and J.P. Vilela (2021), "A survey of privacy-preserving mechanisms for heterogeneous data types", *Computer Science Review*, Vol. 41, pp. 100403, https://doi.org/10.1016/J.COSREV.2021.100403.

Custer, B. (2019), "Reuse of data in smart cities legal and ethical frameworks for big data in the public arena" in *Appropriate Use of Data in Public Space*, NL DIGITAAL, www.nldigitalgovernment.nl/document/appropriate-use-of-data-in-public-space/ (accessed 21 January 2022).

Diver, L. (2018), "Law as a User: Design, Affordance, and the Technological Mediation of Norms", *SCRIPTed: A Journal of Law, Technology and Society*, Vol. 15(1), https://script-ed.org/article/law-as-a-user-design-affordance-and-the-technological-mediation-of-norms/ (accessed 12 January 2022).

DLA Piper (2021), *Data protection laws of the world*, www.dlapiperdataprotection.com/index.html?t=world-map&c=FR (accessed 21 January 2022).

EDPB (2021), *Guidelines 01/2020 on processing personal data in the context of connected vehicles and mobility related applications – Version 2.0*, https://edpb.europa.eu/system/files/2021-03/edpb_guidelines_202001_connected_vehicles_v2.0_adopted_en.pdf.

EIB (2021), *EIB Technical Note on Data Sharing in Transport", European Investment Bank*, www.eib.org/attachments/publications/technical_note_on_data_sharing_in_transport_en.pdf.

Errounda, F.Z. and Y. Liu (2021), "Collective location statistics release with local differential privacy", *Future Generation Computer Systems*, Vol. 124, pp. 174-186, https://doi.org/10.1016/J.FUTURE.2021.05.020.

EU (2020), *Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act),* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767 (accessed 22 January 2022).

EU (2016a), *Regulation of the European Parliament and of the European Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC – (EU) 2016/679 (General Data Protection Regulation),* https://eur-lex.europa.eu/eli/reg/2016/679/oj (accessed 22 January 2022).

EU (2016b), *Directive on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure – (EU) 2016/943*, https://eur-lex.europa.eu/eli/dir/2016/943/oj (accessed 8 February 2022).

EU (2002*), Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic*

communications sector (Directive on privacy and electronic communications), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32002L0058&from=EN (accessed 22 January 2022).

European Commission (2021a), "Open Standards for Linked Organisations – OSLO", Joinup – European Commission, https://joinup.ec.europa.eu/collection/oslo-open-standards-linked-organisations-0 (accessed 21 January 2022).

European Commission (2021b), "When is a Data Protection Impact Assessment (DPIA) required?", https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/when-data-protection-impact-assessment-dpia-required_en (accessed 22 January 2022).

European Commission (2021c), "Data Act (including the review of the Directive 96/9/EC on the legal protection of databases) – Inception Impact Assessment", https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13045-Data-Act-&-amended-rules-on-the-legal-protection-of-databases_en (accessed 22 January 2022).

European Commission (2020a), "Proposal for a Regulation of the European Parliament and of the Council on European data governance", COM/2020/767 final, *Data Governance Act*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767 (accessed 22 January 2022).

European Commission (2020b), *Shaping Europe"s Digital Future*, https://ec.europa.eu/info/sites/default/files/communication-shaping-europes-digital-future-feb2020_en_4.pdf.

European Commission (2020c), *A European strategy for data*, https://ec.europa.eu/info/sites/default/files/communication-european-strategy-data-19feb2020_en.pdf.

European Commission (2018), "Staff Working Document – Guidance on sharing private sector data in the European data economy", https://digital-strategy.ec.europa.eu/en/news/staff-working-document-guidance-sharing-private-sector-data-european-data-economy (accessed 22 January 2022).

European Commission (2017), *Commission Delegated Regulation supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of EU-wide multimodal travel information services*, (EU) 2017/1926, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R1926 (accessed 14 February 2022).

European Commission et al. (2019), *Competition Policy for the Digital Era*, https://data.europa.eu/doi/10.2763/407537 .

Febvre, T. (2021), "Practical Guide to Mobility Data Sharing and Personal Privacy under GDPR ruling", VIANOVA blog, www.vianova.io/blog/practical-guide-to-mobility-data-sharing-and-personal-privacy-under-gdpr-ruling (accessed 5 July 2021).

FIWARE-TMForum (2021), "Business-API-Ecosystem", GitHub, https://github.com/FIWARE-TMForum/Business-API-Ecosystem (accessed 22 January 2022).

Flanders Department of Mobility and Public Works (2020), Application Profile Mobility: Trips and offers, https://data.vlaanderen.be/standaarden/erkende-standaard/applicatieprofiel-mobiliteit-trips-en-aanbod.html (accessed 27 February 2021).

Fortin, P., C. Morency and M. Trépanier (2016), "Innovative GTFS Data Application for Transit Network Analysis Using a Graph-Oriented Method", *Journal of Public Transportation,* Vol. 19(4).

Gauquelin, A. (2020), "The evolution of GBFS", Shared Micromobility, https://shared-micromobility.com/the-evolution-of-gbfs/ (accessed 14 February 2022).

Geist, S., B. Klievink and B. Steunenberg (2019), "Policy challenges relating to data use" in *Appropriate Use of Data in Public Space,* NL DIGITAAL, www.nldigitalgovernment.nl/document/appropriate-use-of-data-in-public-space/ (accessed 22 January 2022).

Green, B. (2019), *The Smart Enough City: Putting technology in Its place to reclaim our urban future*, Cambridge, Massachusetts, The MIT Press, https://doi.org/10.7551/mitpress/11555.001.0001.

Henry, J. (2021), "Expansion of LA tracking system to taxis, ride-hailing services stirs privacy concerns", *Los Angeles Daily News*, www.dailynews.com/2021/02/28/expansion-of-la-tracking-system-to-taxis-ride-hailing-services-stirs-privacy-concerns/ (accessed 22 January 2022).

Hildebrandt, M. (2015), *Smart technologies and the end(s) of law: novel entanglements of law and technology*, Cheltenham, United Kingdom, Edward Elgar Publishing.

ICO (2021), *Data protection impact assessments*, UK Information Commissioner's Office, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/ (accessed 22 January 2022).

ISO 31000:2018 (2018*), ISO 31000:2018, Risk management – Guidelines*, International Organization for Standardization.

ITF (2021a), "Big Data for Travel Demand Modelling: Summary and Conclusions", *ITF Roundtable Reports*, No. 186, OECD Publishing, Paris, https://doi.org/10.1787/08378837-en.

ITF (2021b), *Reversing Car Dependency: Summary and Conclusions*, ITF Roundtable Reports, No. 181, OECD Publishing, Paris, https://doi.org/10.1787/bebe3b6e-en.

ITF (2021c), *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.

ITF (2021d), "Developing Innovative Mobility Solutions in the Brussels-Capital Region", *International Transport Forum Policy Papers,* No. 97, OECD Publishing, Paris, www.itf-oecd.org/sites/default/files/docs/innovative-mobility-brussels-captial-region.pdf.

ITF (2021e), "The Innovative Mobility Landscape: The Case of Mobility as a Service", *International Transport Forum Policy Papers*, No. 92, OECD Publishing, Paris, www.itf-oecd.org/sites/default/files/docs/innovative-mobility-landscape-maas.pdf.

ITF (2019a), "Governing Transport in the Algorithmic Age", *International Transport Forum Policy Papers*, No. 82, OECD Publishing, Paris, https://doi.org/10.1787/eec0b9aa-en.

ITF (2019b), *Regulating App-Based Mobility Services: Summary and Conclusions*, ITF Roundtable Reports, No. 175, OECD Publishing, Paris, https://doi.org/10.1787/94d27a3a-en.

ITF (2018), "Blockchain and Beyond: Encoding 21st Century Transport", *International Transport Forum Policy Papers*, No. 52, OECD Publishing, Paris, https://doi.org/10.1787/bf31443f-en.

ITF (2016), "Data-Driven Transport Policy", *International Transport Forum Policy Papers*, No. 20, OECD Publishing, Paris, https://doi.org/10.1787/5jlwvz8g4vbs-en.

ITF (2015), "Big Data and Transport: Understanding and Assessing Options", *International Transport Forum Policy Papers*, No. 8, OECD Publishing, Paris, https://doi.org/10.1787/5jlwvzdb6r47-en.

Johansen, A. et al. (2019), *Project Risk and Opportunity Management: The Owner's Perspective*, Routledge, https://books.google.fr/books?id=nHGPDwAAQBAJ&printsec=frontcover#v=onepage&q&f=false (accessed 14 February 2022).

Kanhere, O. and T.S. Rappaport (2021), "Position Location for Futuristic Cellular Communications – 5G and Beyond", *IEEE Communications Magazine*, Vol. 59(1), https://arxiv.org/pdf/2102.12074.pdf.

Kayali, L. and V. Manancourt, V. (2021), "How Europe's new privacy rules survived years of negotiations, lobbying and drama", *Politicio-EU*, www.politico.eu/article/europe-privacy-rules-survived-years-of-negotiations-lobbying/ (accessed 23 January 2022).

Keating, R. et al. (2021), "The evolution of 5G New Radio positioning technologies", *White Paper*, Nokia – Bell Labs, https://d1p0gxnqcu0lvz.cloudfront.net/documents/Nokia_The_Evolution_of_5G_New_Radio_Positioning_Technologies_White_Paper_EN.pdf.

Khoshgozaran, A. and C. Shahabi (2009), "Private Information Retrieval Techniques for Enabling Location Privacy in Location-Based Services", *Lecture Notes in Computer Science*, Vol. 5 599, pp. 59-83, https://doi.org/10.1007/978-3-642-03511-1_3.

Kim, J. W. et al. (2021), "A Survey of differential privacy-based techniques and their applicability to location-Based services", *Computers and Security*, Vol. 111, pp. 102464, https://doi.org/10.1016/J.COSE.2021.102464.

Knowles, N. J. (2019). Transmodel Standards Harmonisation: Report on Standards Harmonisation Including a GTFS to Transmodel / NeTEx mapping, https://www.transmodel-cen.eu/wp-content/uploads/2019/10/StandardsHarmonisation-2019-njsk-v1.0-1.pdf.

Laroui, M. et al. (2021), "Edge and fog computing for IoT: A survey on current research activities and future directions", *Computer Communications*, Vol. 180, pp. 210-231, https://doi.org/10.1016/J.COMCOM.2021.09.003.

Lehrer, E. (2021), "Public Spaces in the Digital Age", *National Affairs*, Fall 2021, Vol. 49, www.nationalaffairs.com/publications/detail/public-spaces-in-the-digital-age (accessed 23 January 2022).

LI.EU (2021), "MIMs Plus: Living-in EU Technical Specifications – version 4.0 FINAL", Living-in.EU, https://living-in.eu/sites/default/files/files/MIMs-Plus-v4-0-Final.pdf.

Lockton, D. (2006), "Disaffordances and engineering obedience", *Architectures*, https://web.archive.org/web/20210411044727/http://architectures.danlockton.co.uk/2006/10/22/disaffordances-and-engineering-obedience/ (accessed 23 January 2022).

McKenzie, G. (2019), "Spatiotemporal comparative analysis of scooter-share and bike-share usage patterns in Washington D.C.", *Journal of Transport Geography*, Vol. 78, pp. 19-28, https://doi.org/10.1016/J.JTRANGEO.2019.05.007.

MDC (2021), Mobility Data Collaborative, https://mdc.sae-itc.com/#work (accessed 27 February 2021).

MobilityData (2021), "GBFS and Shared Mobility Data Policy for European Cities", https://mobilitydata.org/gbfs-and-shared-mobility-data-policy-in-europe/ (accessed 23 January 2022).

MobilityData (2020), "What's new in GBFS v2.0.", https://mobilitydata.medium.com/whats-new-in-gbfs-v2-0-63eb46e6bdc4 (accessed 14 February 2022).

Mundhe, P., S. Verma and S. Venkatesan (2021), "A comprehensive survey on authentication and privacy-preserving schemes in VANETs", *Computer Science Review*, Vol. 41, pp. 100411, https://doi.org/10.1016/j.cosrev.2021.100411.

National Transport Commission (2020), "Government access to vehicle-generated data", *Policy Paper November 2020*, www.ntc.gov.au/sites/default/files/assets/files/NTC-Policy-Paper-Vehicle-generated-data.pdf.

New York Times (2017), "Swedish Government Scrambles to Contain Damage From Data Breach", nytimes.com, www.nytimes.com/2017/07/25/world/europe/ibm-sweden-data-outsourcing.html (accessed 15 February 2022).

Norman, D.A. (1990), *The Design of Everyday Things*, Doubleday, https://books.google.fr/books?id=b09jQgAACAAJ (accessed 22 January 2022).

NUMO (2021), "About the Privacy Principles for Mobility Data", www.mobilitydataprivacyprinciples.org/about (accessed 22 January 2022).

NUMO (2020), "Leveraging Data to Achieve Policy Outcomes", https://policydata.numo.global (accessed 25 January 2022).

OASC (2021), "Minimal Interoperability Mechanisms – MIMs", *Open and Agile Smart Cities network*, https://oascities.org/minimal-interoperability-mechanisms/ (accessed 22 January 2022).

OECD (2021a), "Report on the Implementation of the Recommendation of the Council Concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data", *Note by the Secretary-General*, https://one.oecd.org/document/C(2021)42/en/pdf.

OECD (2021b), "Data governance: Enhancing access to and sharing of data", High-level virtual launch event, 10 December 2021, www.oecd.org/sti/ieconomy/enhanced-data-access.htm (accessed 22 January 2022).

OECD (2021c), *Recommendation of the Council on Enhancing Access to and Sharing of Data*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463 (accessed 22 January 2022).

OECD (2020), *Good Practice Principles for Data Ethics in the Public Sector*, www.oecd.org/digital/digital-government/good-practice-principles-for-data-ethics-in-the-public-sector.htm (accessed 22 January 2022).

OECD (2019), *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*, OECD Publishing, Paris, https://doi.org/10.1787/276aaca8-en.

OECD (2013), *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0188 (accessed 22 January 2022).

OECD (2011), "Price Transparency", in *Information Exchanges Between Competitors under Competition Law*, Competition Policy Roundtables, p35, www.oecd.org/competition/cartels/48379006.pdf

OGC (2021), "OGC SensorThings API", Open Geospatial Consortium, www.ogc.org/standards/sensorthings (accessed 22 January 2022).

OMF (2021a), "About MDS", Open Mobility Foundation, www.openmobilityfoundation.org/about-mds/ (accessed 27 February 2021).

OMF (2021b), "Mobility Data Specification – Architectural Landscape", Open Mobility Foundation, https://github.com/openmobilityfoundation/governance/blob/main/documents/OMF-MDS-Architectural-Landscape.pdf.

OMF (2021c), "MDS – Community Projects", Open Mobility Foundation, www.openmobilityfoundation.org/community-projects/ (accessed 2 November 2021).

OMF (2021d), "Understanding MDS APIs", Open Mobility Foundation, https://github.com/openmobilityfoundation/governance/blob/main/technical/Understanding-MDS-APIs.md (accessed 12 September 2021).

OMF (2021e), "Using MDS under GDPR", Open Mobility Foundation, www.openmobilityfoundation.org/using-mds-under-gdpr/ (accessed 24 January 2022).

OMF (2021f), "Announcing MDS 1.2.0", Open Mobility Foundation, www.openmobilityfoundation.org/mds-version-1-2-0/ (accessed 22 January 2022).

OMF (2020), *Privacy Guide for Cities – Mobility data specification*, https://github.com/openmobilityfoundation/governance/raw/main/documents/OMF-MDS-Privacy-Guide-for-Cities.pdf.

OSLO (2021), Open Standards for Linked Organisations, https://joinup.ec.europa.eu/collection/oslo-open-standards-linked-organisations-0 (accessed 24 January 2022).

Paek, H.-J. and T. Hove (2017), "Risk Perceptions and Risk Characteristics", *Oxford Research Encyclopedia of Communication*, https://doi.org/10.1093/ACREFORE/9780190228613.013.283.

Pignatelli, F. et al. (2020), "Guidelines for public administrations on location privacy", European Commission DG Joint Research Centre, https://publications.jrc.ec.europa.eu/repository/handle/JRC103110 (accessed 24 January 2022).

Plotkin, D. (2021), "Chapter 3 – Stewardship Roles and Responsibilities", in Plotkin, D. (ed.), *Data Stewardship (Second Edition)*, 2nd edition, Academic Press, pp. 39-64, https://doi.org/10.1016/B978-0-12-822132-7.00003-6.

Poikola, A. et al. (2020), "MyData – An introduction to human-centric use of personal data", in Lähteenoja, V. (ed.), *MyData,* Finnish Ministry of Transport and Communications.

Porter, D.J., D.S. Kim and S. Ghanbartehrani (2014), "Proof of Concept: GTFS Data as a basis for optimization of Oregon's regional and statewide transit networks", *Final Report SPR 752*, www.oregon.gov/ODOT/Programs/ResearchDocuments/SPR752_GTFS_Data.pdf.

Qi, Y. et al. (2021), "Privacy-preserving blockchain-based federated learning for traffic flow prediction", *Future Generation Computer Systems*, Vol. 117, pp. 328-337, https://doi.org/10.1016/j.future.2020.12.003.

Rajashekar, M.B. and S.M. Sundaram (2020), "A Survey on User's Location Detail Privacy-Preserving Models", *SN Computer Science*, Vol. 1(3), pp. 174, https://doi.org/10.1007/s42979-020-00178-z.

Rauer, N. and S. Cameron (2020), "The EU's Data Governance Act just part of data sharing puzzle", Pinsent Masons, www.pinsentmasons.com/out-law/analysis/the-eus-data-governance-act-just-part-data-sharing-puzzle (accessed 24 January 2022).

Ren, W. and S. Tang (2020), "EGeoIndis: An effective and efficient location privacy protection framework in traffic density detection", *Vehicular Communications*, Vol. 21, pp. 100187, https://doi.org/10.1016/j.vehcom.2019.100187.

Rosenblum, P. and S. Maples (2009), *Contracts Confidential: Ending Secret Deals in the Extractive Industries*, Revenue Watch Institute.

Safi, Q.G.K. et al. (2017), "PIaaS: Cloud-oriented secure and privacy-conscious parking information as a service using VANETs", *Computer Networks*, Vol. 124, pp. 33-45, https://doi.org/10.1016/j.comnet.2017.06.001.

Schmandt-Besserat, D. (2015), "The Evolution of Writing", in Wright, J. (ed.), *International Encyclopedia of Social and Behavioral Sciences*, 2nd edition, Elsevier B.V., https://sites.utexas.edu/dsb/tokens/the-evolution-of-writing/.

Seo, S.A. (2019), *Policing the open road: How cars transformed American freedom*, Cambridge, Massachusetts: Harvard University Press.

SuM4All (2021), *Sustainable Mobility: Policy Making for Data Sharing*, wbcsd, www.wbcsd.org/Programs/Cities-and-Mobility/Transforming-Mobility/Digitalization-and-Data-in-Urban-Mobility/Policy-to-Enable-Data-Sharing/Resources/Sustainable-mobility-Policy-making-for-data-sharing (accessed 24 January 2022).

SynchroniCity (2020), "A guide to SynchroniCity: A universal approach to developing, procuring and deploying IoT- and AI-enabled services", https://synchronicity-iot.eu/ (accessed 24 January 2022).

Thales (2021), "Beyond GDPR: Data protection around the world", www.thalesgroup.com/en/markets/digital-identity-and-security/government/magazine/beyond-gdpr-data-protection-around-world (accessed 24 January 2022).

Transport Styrelsen (2017), *Statement about the information in media regarding our IT public procurement*, https://transportstyrelsen.se/sv/Om-transportstyrelsen/statement-about-the-information-in-media-regarding-our-it-public-procurement/ (accessed 22 January 2022).

U3115299 (2017), Wikimedia Commons - File:IC-Risk-Assessment-Matrix-Template.jpg, https://commons.wikimedia.org/wiki/File:IC-Risk-Assessment-Matrix-Template.jpg

UITP (2020), *Sharing of Data in Public Transport: Value, Governance and Sustainability*, www.uitp.org/publications/sharing-of-data-in-public-transport-value-governance-and-sustainability/ (accessed 26 January 2022).

UN-OHCHR (2021), "Special Rapporteur on the right to privacy", United Nations Office of the High Commissioner on Human Rights, www.ohchr.org/en/issues/privacy/sr/pages/srprivacyindex.aspx (accessed 24 January 2022).

United Nations (1948), *Universal Declaration of Human Rights*, www.un.org/en/about-us/universal-declaration-of-human-rights (accessed 24 January 2022).

van der Waal, S. et al. (2020), "Digital European Public Spaces", Waag – Technology and Society. https://culturalfoundation.eu/wp-content/uploads/2021/05/Waag-Report-on-European-Digital-Public-Spaces.pdf.

Van Roy, J. (2020), "Flanders develops data standard for smart mobility", newmobility.news, https://newmobility.news/2020/04/30/flanders-develops-data-standard-for-smart-mobility/ (accessed 27 February 2021).

Ville de Paris (2021), "Signaleurs « Dans ma rue » : agir sur l'espace public"[Flaggers 'In my street': acting on the public space], www.paris.fr/pages/signaleurs-dans-ma-rue-agir-sur-l-espace-public-6682 (accessed 2 September 2021).

Waag (2021), "Governance of Digital Public Spaces", Public Stack, Waag – Technology and Society, https://publicstack.net/digital-public-spaces/ (accessed 24 January 2022).

Wang, L. (2017), "Heterogeneous Data and Big Data Analytics", *Automatic Control and Information Sciences*, Vol. 3(1), pp. 8-15, https://doi.org/10.12691/ACIS-3-1-3.

WBCSD (2020), *Enabling data sharing: Emerging principles for transforming urban mobility*, www.wbcsd.org/Programs/Cities-and-Mobility/Transforming-Urban-Mobility/Digitalization-and-Data-in-Urban-Mobility/Resources/Enabling-data-sharing-Emerging-principles-for-transforming-urban-mobility (accessed 24 January 2022).

Webb, K. (2020a), "Local power in the age of digital policing", Triangulator, https://triangulator.org/blog/local-power-digital-policing/ (accessed 24 January 2022).

Webb, K. (2020b), "What 'data' is for: Designing digital systems for transport planning, oversight and enforcement", presentation made to ITF project workshop "Rules and Principles for Public Value Data Sharing" on 23 November, 2020.

WEF (2011), "Personal Data: The Emergence of a New Asset Class", World Economic Forum, www.weforum.org/reports/personal-data-emergence-new-asset-class (accessed 24 January 2022).

Whitelaw, M. (2010), "This is data? Arguing with Data Baby", Teeming Void, http://teemingvoid.blogspot.com/2010/05/this-is-data-arguing-with-data-baby.html (accessed 2 November 2021).

Williams, R. (2021), *Whose Streets? Our Streets! (Tech Edition) 2020-21 "Smart City" Cautionary Trends and 10 Calls to Action to Promote and Protect Democracy*, Cambridge, Massachusetts: Harvard University Press, www.belfercenter.org/publication/whose-streets-our-streets-tech-edition (accessed 24 January 2022).

World Bank (2019), ID4D Practitioner's Guide: Version 1.0., https://id4d.worldbank.org/guide (accessed 24 January 2022).

World Intellectual Property Organization (2021), "Trade Secrets: What is a trade secret?", WIPO, www.wipo.int/tradesecrets/en/ (accessed 14 February 2022).

Xu, V.X., J. Leibold and D. Impiombato (2021), "The architecture of repression", ASPI_ICPC, www.aspi.org.au/report/architecture-repression?__cf_chl_jschl_tk__=pmd_wJ8BYf8.AXda0RC8Q9CIbYKXhrx0uqMn9lNXGnzAuNU-1634636784-0-gqNtZGzNAhCjcnBszQoR (accessed 24 January 2022).

Xu, Y. et al. (2020), "Micromobility Trip Origin and Destination Inference Using General Bikeshare Feed Specification (GBFS) Data", http://arxiv.org/abs/2010.12006 (accessed 14 February 2022).

Yang, M. et al. (2020), "Local Differential Privacy and Its Applications: A Comprehensive Survey", http://arxiv.org/abs/2008.03686 (accessed 30 October 2021).

Zhang, L., Z. Yan and R. Kantola (2017), "Privacy-preserving trust management for unwanted traffic control", *Future Generation Computer Systems*, Vol. 72, pp. 305-318, https://doi.org/10.1016/j.future.2016.06.036.

Zuboff, S. (2019), *The age of surveillance capitalism: the fight for a human future at the new frontier of power*, London: Profile books.

International Transport Forum

Reporting Mobility Data

Good Governance Principles and Practices

This report explores the issues public authorities must address when establishing data-reporting mandates and policies. Transport systems and the people using them generate an ever-increasing amount of data which can help improve transport system performance. The risks from missing or overly broad reporting policies can be mitigated by following reporting guidelines and principles that focus on the public value of data.

International Transport Forum
2 rue André Pascal
F-75775 Paris Cedex 16
+33 (0)1 73 31 25 00
contact@itf-oecd.org
www.itf-oecd.org

OECD

Published: 03/2022 | Photo credit: Philippe Crist